Danica Stefanović
Maher Kayal

ACSP
Analog Circuits And Signal Processing

Structured Analog CMOS Design

Springer

STRUCTURED ANALOG CMOS DESIGN

ANALOG CIRCUITS AND SIGNAL PROCESSING SERIES

Consulting Editor: Mohammed Ismail. Ohio State University

Titles in Series:

LOWPOWER UWB CMOS RADAR SENSORS
Paulino, Nuno, Goes, Jo„o, Steiger GarÁ„o, Adolfo
ISBN: 978-1-4020-8409-6

THE GM/ID DESIGN METHODOLOGY FOR CMOS ANALOG LOWPOWER INTEGRATED CIRCUITS
Jespers, Paul G.A.
ISBN-10: 0-387-47100-6

SUBSTRATE NOISE COUPLING IN RFICS
Helmy, Ahmed, Ismail, Mohammed
ISBN: 978-1-4020-8165-1

CIRCUIT AND INTERCONNECT DESIGN FOR HIGH BIT-RATE APPLICATIONS
Veenstra, Hugo, Long, John R.
ISBN: 978-1-4020-6882-9

HIGH-RESOLUTION IF-TO-BASEBAND SIGMADELTA ADC FOR CAR RADIOS
Silva, Paulo G.R., Huijsing, Johan H.
ISBN: 978-1-4020-8163-7

MULTI-BAND RF FRONT-ENDS WITH ADAPTIVE IMAGE REJECTION
A DECT/BLUETOOTH CASE STUDY
Vidojkovic, V., van der Tang, J., Leeuwenburgh, A., van Roermund, A.H.M.
ISBN: 978-1-4020-6533-0

SILICON-BASED RF FRONT-ENDS FOR ULTRA WIDEBAND RADIOS
Safarian, Aminghasem, Heydari, Payam
ISBN: 978-1-4020-6721-1

DESIGN OF HIGH VOLTAGE XDSL LINE DRIVERS IN STANDARD CMOS
Serneels, Bert, Steyaert, Michiel
ISBN: 978-1-4020-6789-1

BASEBAND ANALOG CIRCUITS FOR SOFTWARE DEFINED RADIO
Giannini, Vito, Craninckx, Jan, Baschirotto, Andrea
ISBN: 978-1-4020-6537-8

HIGH-LEVEL MODELING AND SYNTHESIS OF ANALOG INTEGRATED SYSTEMS
Martens, Ewout S.J., Gielen, Georges
ISBN: 978-1-4020-6801-0

CMOS MULTI-CHANNEL SINGLE-CHIP RECEIVERS FOR MULTI-GIGABIT OPT...
Muller, P., Leblebici, Y.
ISBN 978-1-4020-5911-7

ANALOG-BASEBAND ARCHITECTURES AND CIRCUITS
FOR MULTISTANDARD AND LOW-VOLTAGE WIRELESS TRANSCEIVERS
Mak, Pui In, U, Seng-Pan, Martins, Rui Paulo
ISBN: 978-1-4020-6432-6

FULL-CHIP NANOMETER ROUTING TECHNIQUES
Ho, Tsung-Yi, Chang, Yao-Wen, Chen, Sao-Jie
ISBN: 978-1-4020-6194-3

ANALOG CIRCUIT DESIGN TECHNIQUES AT 0.5V
Chatterjee, S., Kinget, P., Tsividis, Y., Pun, K.P.
ISBN-10: 0-387-69953-8

SWITCHED-CAPACITOR TECHNIQUES FOR HIGH-ACCURACY FILTER AND ADC...
Quinn, P.J., Roermund, A.H.M.v.
ISBN 978-1-4020-6257-5

LOW-FREQUENCY NOISE IN ADVANCED MOS DEVICES
von Haartman, M., ÷stling, M.
ISBN 978-1-4020-5909-4

ULTRA LOWPOWER CAPACITIVE SENSOR INTERFACES
Bracke, W., Puers, R. (et al.)
ISBN 978-1-4020-6231-5

BROADBAND OPTO-ELECTRICAL RECEIVERS IN STANDARD CMOS
Hermans, C., Steyaert, M.
ISBN 978-1-4020-6221-6

CMOS SINGLE CHIP FAST FREQUENCY HOPPING SYNTHESIZERS FOR WIRELESS
MULTI-GIGAHERTZ APPLICATIONS
Bourdi, Taoufik, Kale, Izzet
ISBN: 978-1-4020-5927-8

For other titles published in this series, go to
www.springer.com/series/7381

STRUCTURED ANALOG CMOS DESIGN

DANICA STEFANOVIĆ

and

MAHER KAYAL
Ecole Polytechnique Fédérale de Lausanne, Switzerland

Springer

Danica Stefanović
Ecole Polytechnique
Fédérale de Lausanne,
Switzerland

Maher Kayal
Ecole Polytechnique
Fédérale de Lausanne,
Switzerland

ISBN 978-1-4020-8572-7 e-ISBN 978-1-4020-8573-4

Library of Congress Control Number: 2008928848

Printed on acid-free paper.

9 8 7 6 5 4 3 2 1

springer.com

Preface

Standard analog design procedure is usually based on a large number of simulations, strongly depends on the type of analog circuit that has to be implemented and requires a lot of manipulation at the transistor level. Simulators offer accurate modeling and precise calculations, but on the other hand ascertaining circuit parameter dependences is difficult and depends on analog designer expertise and knowledge. Moreover, with CMOS technology improvements, the complexity of analog design tasks further increases, since the design specifications become more severe in terms of gain and speed, requiring at the same time minimization of the circuit surface.

This book describes a structured analog design approach that makes it possible to simplify complex analog design problems and develop a global design strategy that can be used for the design of different analog cells. The basic concept consists in analog cell partitioning into the basic analog structures and sizing of these basic analog structures in a predefined procedural design sequence. The basic analog structure specifications are derived from circuit-level requirements, and thus its sizing in the environment imposed by the circuit demands less effort. Furthermore, the procedural design sequence ensures the correct propagation of design specifications, the verification of parameter limits and the local optimization loops.

The proposed transistor-level design procedure is based on the continuous MOS modeling approach and relies on the device inversion level as a fundamental design variable. Since all important design parameters can be expressed as continuous functions of the device inversion level, the design optimum as well as the technology limits can be easily found.

Finally, the developed design procedure is implemented as a CAD tool that guides and assists the user during analog design tasks and provides an interactive interface that allows instantaneous visualization of design trade-offs.

All material presented in this book is the result of the Ph.D. project that was performed at EPFL (Ecole Polytechnique Fédérale de Lausanne), Lausanne, Switzerland and supervised by Prof. Maher Kayal. The developed CAD tool can be downloaded from the Electronics Labs web site: http://analog.epfl.ch.

Danica Stefanović
Lausanne, March 2008

Table of Contents

1 Introduction

2 Transistor level design

3 BSIM to EKV conversion

4 Basic analog structures

5 Procedural design scenarios

6 PAD tool

7 Topology variants

8 Practical example: the design of analog amplifiers in the Delta-Sigma modulator system

CHAPTER 1

Introduction

Even though it becomes possible to realize more and more functionalities using digital circuits, the design of analog circuits still has its important role in the design of numerous systems on chip that are produced nowadays. Analog cells are extensively used in analog-to-digital and digital-to-analog interfaces in mixed-mode systems, as well as for the implementation of different analog functions such as amplifiers or integrators.

With the advances in technology development and the use of submicron and sub-100 nm CMOS processes, digital circuits have become faster and more precise, occupying at the same time smaller surfaces. This imposes more severe specifications in terms of gain, speed and acceptable surface for the analog counterparts. Hence, the design of analog cells becomes a very sophisticated task, since modern technologies introduce two major difficulties: larger parasitic capacitances (since oxide thickness decreases) and smaller output resistances (due to short channel-effects and drain induced barrier lowering DIBL). In other words, a common analog design task nowadays is to design a circuit that will achieve the demanding specifications, using a smaller voltage headroom (as the supply voltage headroom is also reduced) and transistors that have more parasitic effects and smaller intrinsic gain. Thus, the eternal question "How to size analog circuits to achieve the required performances?" arises with even more importance than before.

The state-of-the-art in analog design today can be summarized in the few following comments:

- There is no general analog design methodology - the analog design is still based on previous experience and knowledge, and in some cases even on "trial and error" simulation loops.
- There is no general approach of how to deal with the tasks on different levels of complexity and different levels of abstraction.
- There exist a large number of **C**omputer **A**ided **D**esign (CAD) tools that are primarily used for simulation, layout generation and postlayout verification of analog cells.
- There exist a large number of analog design automation tools that are based on different approaches and methods, but these are used in only a limited number of design cases, and always in the university domain.

- There are few analog design automaton tools, compared to the digital domain, that are widely used and largely accepted by the analog designer communities in industry.

Consequently, the questions that need answers are the following:

- How to deal with the increasing complexity of the analog circuits, and the very demanding specification sets?
- How to estimate the technology limits?
- How to bridge the gap between hand-calculations, simulations and measurement results?
- Is it possible to understand the device physics and apply it successfully for the analog design purposes instead of blindly believing in simulators and extremely complicated mathematical models?
- How to optimize the analog circuits and find the best design trade-offs?
- Is it possible to develop CAD tools for analog design assistance rather than for analog design automation?
- Is it possible to encapsulate expertise and knowledge at different abstraction levels - transistor level, analog cell level, system level?

1.1 The objectives of this work

This work proposes the approach that favors the simplification of complex problems and global design strategy. It intentionally avoids treating the analog design as a mathematical problem, developing a design procedure based on the understanding of device physics and approximations that give insight into parameter interdependences. Therefore, a structured design approach that can be used for the design of a large number of analog cells is proposed. The basic concept is the following: since each analog circuit can be seen as a set of basic analog blocks, and since each block has a different role in the circuit and affects some specific circuit-level parameters, the design procedure consists in sizing the basic analog structures to achieve the required specifications. The basic analog structures are sized in a specific design sequence, denoted here as a procedural design sequence.

On the other hand, each basic analog structure can be seen as a set of transistors working in a certain design environment. Thus, there are a finite number of transistor-level design situations that can be easily resolved using the device inversion level as a design variable. To make this possible a continuous model based on device physics and dedicated to design of analog circuits is used.

Finally, the proposed design concept is implemented as a CAD tool that can be used interactively with the simulator. It provides the graphical interface that

enables the user to explore the design space and find the design trade-offs without limiting his decisions.

1.2 Structured analog design

The structured design approach is demonstrated using the example of an analog amplifier. Nevertheless, it can be used for the design of any other analog cell. Figure 1.1 depicts the design flow from system-level to transistor-level based on the structured design approach. The amplifier is replaced in the early design stage by its behavioral model. The circuit-level specifications are then derived from both system-level requirements and system-level simulations. Next, the chosen amplifier topology is partitioned into basic analog structures and the design specifications are derived for each basic analog structure from the circuit-level requirements. To facilitate this task a **basic analog structure library** is defined, and for each basic structure the important design cases and design parameters are specified.

Procedural design scenario - After the circuit has been partitioned, and the specifications translated to the basic analog structures level, it is demonstrated how to develop procedural design sequence in order to achieve the required specifications.

Local optimization - The optimization is also translated at the level of basic analog structures. Hence, it is proposed how to estimate the technology limits, and determine the parameter bounds. Since there are always several ways to find a solution, and it strongly depends on user decisions, the local optimization guidelines and comments are given in critical points in procedural design scenarios.

Topology variants - Finally, for the cases where the specifications are very demanding and cannot be achieved, the proposed solution is to replace only the basic analog structure affecting the parameter in question by its more advanced version.

When the circuit is completed at the transistor-level, its performance must be confirmed by repeating the same system-level simulations used at the beginning of the design. If some system-level specifications are not achieved, then the parasitic effects can be extracted, added to the behavioral model and the design sequence repeated.

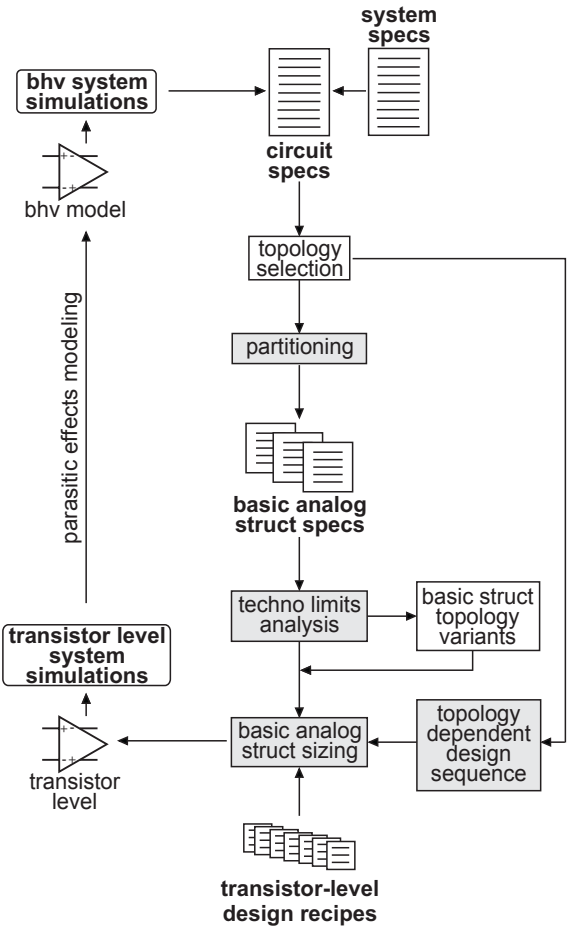


Figure 1.1 Design flow based on the structured design approach

1.3 Transistor-level design based on the device inversion level

The problem of transistor sizing is posed differently in this book. The classical design variables: I_D (transistor bias current), W (transistor width) and L (transistor length) are replaced by two design variables: IF (device inversion level) and L (transistor length). According to the structured design approach, and the specification derivation guidelines, it can be shown that the transistor bias current is not an independent variable at the transistor-level since it is imposed by the circuit-level specifications. Thus, it can be changed only at

circuit-level, and if so, the redesign loop on the level of basic analog structures is required.

The transistor-level design is based on the EKV modeling approach where the transistor design parameters are given as functions of device inversion level. Therefore, it is more convenient to use the device inversion level as a design variable. The main advantage of this is that there is no need for weak and strong inversion asymptotes (or approximations) since each parameter is given as a continuous and single expression of the inversion level. Furthermore, this makes it possible to easily estimate parameter bounds and define the optimization strategies at the transistor-level. On this basis, the possible design cases are identified and the transistor-level design recipes are proposed.

1.4 CAD tools for analog design assistance

The proposed procedural design scenarios and the basic analog structures library are encapsulated as a **Procedural Analog Design (PAD) tool**. This tool is dedicated to analog design assistance, and it is intended to be used interactively with a simulator. For each analog topology a step-by-step design sequence is proposed, and guidelines and advice from the experienced designer are provided at each step. Moreover, its interactive interface provides instantaneous visualization of the design trade-offs and design space exploration.

To make it possible to use the PAD tool with both the current industrial standard MOS model (BSIM model) and the EKV MOS model library files, the **BSIM2EKV converter** for the automatic conversion of BSIM to EKV model parameters is developed as a separate module.

1.5 Practical design example

The proposed design approach is validated on the design of three analog amplifiers forming part of multi-bit hybrid Delta-Sigma modulator system, and using 0.18 μm CMOS technology.

1.6 Book organization

Each chapter in this book is dedicated to one aspect of the previously presented concept. Since there are different levels of explanation, and a large number of facts and information at each level, a global picture of the design procedure can be created only after all chapters have been read. The

experienced designer may find a lot of familiar information in each topic. However, the main contribution of this work is not in discovering new circuits, but in providing a general vision and a way of how to progress and where to look in the immense forest of analog design knowledge.

Chapter 2 - Firstly, an overview of the MOS models available on the market today is given. Then, it is explained why the EKV MOS model is chosen as a basis of transistor-level design. Secondly, an overview of the transistor design parameter definitions from the EKV model is provided in order to determine the design variables. Finally, transistor-level design recipes are proposed.

Chapter 3 - This chapter describes the program for the automatic conversion of the parameters of the MOS model that represents the present industrial standard (BSIM model) and the chosen EKV MOS model. It provides details on the conversion algorithm, and presents the results of the conversion for the 0.18 μm CMOS technology example. At the end, comments on the interpretation of results for analog design purposes are given.

Chapter 4 - After the library of basic analog structures mainly used in linear electronics has been defined, the guidelines on circuit partitioning, specification derivation, and behavioral modeling of the analog amplifier are given. All basic analog structures are classified as transconductance, load and bias structures. For each class, the important design parameters and the design cases (that point to transistor-level design recipes) are specified.

Chapter 5 - The material presented in this chapter is focused on procedural design sequence definition. It is presented in the form of complete procedural design scenarios, including circuit-level parameters definition, frequency analysis, circuit partitioning, specification derivation and design sequence presentation, for three different amplifier topologies (folded cascode OTA, fully-differential folded cascode OTA and Miller operational amplifier).

Chapter 6 - This chapter gives a brief presentation of the PAD tool and its basic features as a practical implementation of all previously discussed concepts.

Chapter 7 - The concept of topology variants is demonstrated on the example of a fully-differential folded cascode OTA. For each topology variant the design guidelines, as well as the design drawbacks, are given.

Chapter 8 - As a practical example, the design of analog amplifiers in a Delta-Sigma modulator system is described from system level to transistor level. The following amplifier topologies are presented: high-gain fully-differential folded cascode OTA, high-gain difference differential amplifier and two-stage amplifier with Miller cascode compensation.

CHAPTER 2

Transistor level design

Different MOS transistor models and their basic characteristics are discussed in the beginning of this chapter. The requirements that a MOS model must fulfill to be suitable for the design at the transistor level are identified. Then, an overview of transistor parameters needed for design of analog structures is provided. Finally, a transistor-level design approach based on the choice of the inversion factor and the transistor length is proposed. It includes the design recipes for hand-calculations, and the analyses of parameter limits and parameter optima.

2.1 MOS transistor model

The fundamental step of any analog design procedure is the design at the transistor level, that means sizing of each transistor to achieve the required specifications. To accomplish this task successfully, it is essential to understand the transistor behavior and to calculate the important parameters. Since the transistor behavior is described by its model, a "good analog" model is prerequisite. Several studies are presented in [1]-[3] with the aim of defining what is a good MOS transistor model. The most important features are analyzed from the point of view of device modeling for analog design purposes. The general conclusions are resumed hereafter:

- A good transistor model is based on physical behavior covering significant physical effects such as non-uniform doping effects, mobility effects, velocity saturation, short/narrow channel effects, substrate current, thermal/flicker noise, and temperature effects.
- It should be global and compact (preferably without binning), accurate and with as small as possible number of fitting parameters.
- The parameter extraction procedure should be neither complicated nor time-consuming.
- The model has to cover all geometry ranges of interest and ensure good fitting for the device geometries that are not used during parameter extraction. The accurate modeling means not only the correct I-V

characteristics, but also the correct current derivatives, i.e. transconductances which are important parameters for the design of analog circuits. In addition, the correct best/worst case modeling must be guaranteed for design robustness.
- The requirement for high speed and high frequency operation also impose an accurate modeling of the intrinsic capacitances. On the other hand for low voltage or small current operation, accurate weak and moderate inversion modeling is obligatory.
- A good model is easily implemented, and does not present convergency problems when used with different simulators.

In the next subsection, the good models available today are presented. Then, the features of models that are appropriate for analog design at the transistor level are discussed.

2.1.1 Overview of MOS models

Spice models [4] - The early Spice models (levels 1, 2, 3) were developed at the University of California, Berkeley to be used with their circuit simulator. All model parameters have physical meaning, but some effects are too simplified or are not modeled at all. These models are not used nowadays, since they are inaccurate for submicron technologies and present discontinuities in current derivatives. However, they are mentioned here because of their historical importance.

BSIM(1,2,3,4) models [5]**,** [6] - To overcome modeling problems, a new family of models dedicated to submicron technologies was developed at the same university. The BSIM1 and its improved version BSIM2 are threshold voltage-based empirical models, having good accuracy and continuous derivatives. However, most of the parameters lose their physical meaning, and a large number of fitting parameters is introduced. For example, the BSIM1 model requires about 60 DC parameters, whereas BSIM2 requires about 90 DC parameters. As a result, these models become more appropriate for the design of digital than analog circuits.

The next version, referred to as BSIM3, uses a different approach based on physical behavior. There are more parameters with physical meaning and less fitting parameters. In this case, about 40 parameters are needed for DC analysis. The derivatives are continuous and the important improvement is that there is one single expression for the drain current. Concerning the design of analog circuits, two main problems are reported: less accurate modeling of weak/moderate inversion behavior and incorrect intrinsic capacitances in some regions. Despite this, the version BSIM3v3 (level 7) became an *industrial*

standard, and is one of the most widely used MOS models for submicron CMOS technologies.

BSIM4 was developed as an extension of BSIM3 and it models more physical effects characterizing the sub-100 nm devices. Numerous high-frequency and parasitic effects are also added, making it possible to use this model for the design of RF circuits.

EKV model [7], [8], [19] - The EKV model has been developed at EPFL (Ecole Polytechnique Fédérale de Lausanne) as a charge-based physical model. Its version 2.6 (level 5) is dedicated to the design of low voltage and low power analog circuits using submicron CMOS technologies. The most recent version 3.0 also includes modeling of all relevant effects for RF circuit design.

This continuous and compact model shows good accuracy, even though it has a small number of parameters (e.g. for DC analysis only 9 physical, 2 temperature and 3 fitting parameters are needed). It respects the intrinsic source/drain symmetry, introduces the inversion factor as a transistor parameter, and shows the correct behavior in all inversion regions. Version 3.0 is validated for nanometer technologies down to 65 nm.

PHILIPS models [9]-[11] - The MOS model 9 is a compact threshold-voltage based model intended to be used with both analog and digital circuits. It models all significant effects, and has continuous first and second-order derivatives.

The more recent MOS model 11 is a compact surface potential-based model dedicated to design of digital, analog and RF circuits. It has continuous first and higher-order derivatives and also models the effects present in advanced technologies such as gate leakage, gate-induced drain leakage, polydepletion, quantum-mechanical effects and bias-dependent overlap capacitances.

SP model [12] - The SP model is an "advanced physics-based" model developed at the University of Pennsylvania. This surface potential-based model has about 65 parameters and shows correct behavior in all inversion regions. As any good analog design model, it respects the intrinsic source/drain symmetry, it is continuous and has one single expression for the drain current.

PSP model [13], [14] - The PSP model is a combination of the best features of the PHILIPS MOS model 11 and the SP model. It is a surface potential-based model, containing all relevant physical effects. It offers accurate modeling of current, charges and their first and higher-order derivatives, as well as distortion behavior modeling and non-quasi static effects simulation option. It has been adopted as an *industrial standard* for nanometer CMOS technologies.

2.1.2 MOS model oriented towards transistor-level design

As stated previously, having a good MOS model that ensures accurate simulation results and models all relevant effects is mandatory. However, this is not the only requirement. The design of analog circuits requires a lot of manipulation at the transistor level, even before starting the simulations loops. Since simulators are not design tools, the transistor sizes must be obtained by hand-calculation, and only then confirmed by the simulator. Therefore, a small number of model parameters with physical meaning and hierarchical model structure are important advantages. In this case, it becomes possible to approximate model equations, without a great loss of accuracy, and use them for transistor-level manipulations. Moreover, a good understanding of transistor behavior and precise hand-calculations are achieved at the same time.

The design of low voltage circuits requires weak, moderate or limit of moderate inversion as an operating region. This means that the approximations of transistor design parameters in weak and strong inversion are no longer acceptable, as they introduce a large error with regard to the simulation results and do not reflect correctly the transistor behavior. Hence, it is more convenient to have every design parameter expressed as a function of the inversion degree (or the inversion level) and using just one single expression. Furthermore, the inversion level itself becomes an important transistor parameter and it needs to be easily calculated from the model.

Taking this into account, we have chosen the EKV model as the most suitable candidate for transistor-level design. Its hierarchical structure allows us to pass easily from hand-calculations to the complete model without introducing large errors. All important design parameters, such as: the inversion level, the saturation voltage, the Early voltage, the small-signal parameters, the parasitic capacitances, and the transconductance efficiency factor can be directly calculated as functions of the inversion level parameter. Furthermore, the basic concepts of this model can be used to develop a design approach at the transistor level.

2.2 Transistor design parameters

The transistor parameters that are used for the design of analog structures are summarized here. Even though this is not the original contribution of this work, the definitions are introduced with the intention of presenting the transistor parameters and their interdependences in the way they are needed for analog design purposes. Special emphasis is given to the definition of the inversion regions and the inversion level parameter calculation. Later, it is

shown that all design parameters can be expressed as a function of the inversion level parameter. On this basis, the variables for the design at the transistor level are identified and transistor-level design approach based on the device inversion level is proposed in the next section.

It is important to note that the relations that are given in the next subsections represent the approximations for the hand-calculations, with no intention of defining or introducing a new MOS model. The parameter definitions are taken (or derived) from the work of Ch. Enz, F. Krummenacher, and E. Vittoz presented in [8], describing the basics of the EKV modeling concept, and from the work of M. Bucher presented in [17], that lead to the EKV model implementation in commercial simulators. The complete charge-based EKV model equations, version 2.6, such as implemented in a simulator, are given in [16]. Finally, "a rigorous, complete and coherent derivation of the relations between charges and potentials" and the derivation of the EKV model as its approximation can be found in [18].

2.2.1 Transistor symmetry and voltage definitions

A MOS transistor is a four-terminal device presenting an intrinsic source/drain symmetry.

To respect this symmetry, the gate V_G, source V_S and drain V_D voltages are referred to the substrate (bulk), as depicted in Figure 2.1. If V_{Gt}, V_{St}, V_{Dt}, and V_{Bt} are voltages connected to the gate, source, drain and substrate terminal, it follows that

$$V_G = |V_{Gt} - V_{Bt}| \tag{2.1}$$

$$V_S = |V_{St} - V_{Bt}| \tag{2.2}$$

$$V_D = |V_{Dt} - V_{Bt}| \tag{2.3}$$

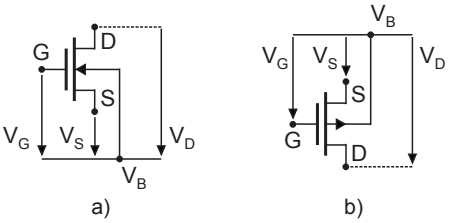


Figure 2.1 Four-terminal symbols of a) NMOS and b) PMOS devices and voltage definitions

Figure 2.2 Simplified three-terminal a) NMOS and b) PMOS transistor symbols when the bulk terminal is connected to the appropriate supply voltage

Table 2.1 Definitions of operating regions with respect to the surface potential

$\Psi_s < 0$	accumulation
$\Psi_s = 0$	flat-band condition ($V_G = V_{FB}$)
$0 < \Psi_s < \Phi_F$	depletion (Φ_F is the Fermi potential)
$\Phi_F < \Psi_s < 2\Phi_F + V_{ch}$	weak inversion
$2\Phi_F + V_{ch} < \Psi_s < 2\Phi_F + V_{ch} + mV_t$	moderate inversion, m = 2-3, V_t = kT/q
$\Psi_s > 2\Phi_F + V_{ch} + mV_t$	strong inversion

Figure 2.2 shows simplified transistor symbols often used in circuit schematics. The unshown substrate terminal is connected to the negative supply voltage in the case of the NMOS transistor, and to the positive supply voltage in the case of the PMOS transistor.

2.2.2 Operating regions

The transistor operating regions are defined as in Table 2.1 with respect to the surface potential Ψ_s and the channel potential V_{ch}. The channel potential V_{ch} corresponds to difference between quasi-Fermi potentials of majority and minority carriers along the channel.

This definition is correct, but cannot be used directly for analog design purposes. However, when the result of the integration of Poisson's equation (2.4) is combined with the relation between the surface potential, the inversion charge and the gate voltage (2.5), it is possible to approximate the inversion mobile charge in each operating region. The integration of this mobile charge along the channel gives the drain current. Finally, the normalized forward current component of the drain current gives the inversion level indication that can be used as a design parameter, as will be shown in subsection 2.2.7.

The result of the integration of Poisson's equation (for an n-channel device) is given here without entering into details

$$Q_{inv} = -\gamma C_{ox}\sqrt{V_t}\left(\sqrt{\frac{\Psi_s}{V_t} + exp\left(\frac{\Psi_s - 2\Phi_F - V_{ch}}{V_t}\right)} - \sqrt{\frac{\Psi_s}{V_t}}\right). \tag{2.4}$$

The relation between the surface potential Ψ_s, the inversion charge Q_{inv} and the gate voltage V_G, valid in all operating regions, is

$$V_G = V_{FB} + \psi_s + \gamma\sqrt{\psi_s} - \frac{Q_{inv}}{C_{ox}}, \tag{2.5}$$

and is developed in [17]. V_{FB} is the flat-band voltage, the gate-oxide capacitance $C_{ox} = \varepsilon_{ox}/t_{ox}$ is calculated from the dielectric constant of SiO_2 ε_{ox} and the oxide thickness t_{ox}, and the parameter γ is called the substrate factor or the body effect factor. It depends on technology, i.e. on the substrate doping concentration N_{sub}, and is equal to

$$\gamma = \frac{\sqrt{2qN_{sub}\varepsilon_s}}{C_{ox}}. \tag{2.6}$$

2.2.3 Strong inversion, pinch-off voltage and slope factor

The definitions of three important design parameters: threshold voltage V_{T0}, pinch-off voltage V_P and slope factor n are related to the strong inversion operation. They are derived from the approximation of the inversion charge in strong inversion given in [8].

The surface potential in strong inversion is $\Psi_s = 2\Phi_F + mV_t + V_{ch}$. It can be rewritten as $\Psi_s = \Psi_0 + V_{ch}$, where $\Psi_0 = 2\Phi_F + mV_t$. Then, equation (2.5) becomes

$$V_G = V_{FB} + \psi_0 + V_{ch} + \gamma\sqrt{\psi_0 + V_{ch}} - \frac{Q_{inv}}{C_{ox}}. \tag{2.7}$$

If now the gate threshold voltage referred to bulk is defined as

$$V_{TB}(V_{ch}) = V_{FB} + \psi_0 + V_{ch} + \gamma\sqrt{\psi_0 + V_{ch}}, \tag{2.8}$$

then the inversion charge can be expressed by

$$Q_{inv} = -C_{ox}(V_G - V_{TB}(V_{ch})). \tag{2.9}$$

"The **threshold voltage V_{T0}** is a gate voltage for which the channel is at equilibrium ($V_{ch} = 0$) and the inversion charge is equal to zero", as stated in [8]. Hence, it follows

$$V_{T0} = V_G\big|_{V_{ch}=0,\,Q_{inv}=0} = V_{TB}\big|_{V_{ch}=0} = V_{FB} + \psi_0 + \gamma\sqrt{\psi_0}\,. \tag{2.10}$$

"The **pinch-off voltage V_P** is a channel potential for which, at a given gate voltage, the inversion charge Q_{inv} becomes zero (i.e. the channel is pinched-off)", as stated in [8]. From

$$V_G\big|_{V_{ch}=V_P,\,Q_{inv}=0} = V_{TB}\big|_{V_{ch}=V_P} = V_{FB} + \psi_0 + V_P + \gamma\sqrt{\psi_0 + V_P} \tag{2.11}$$

it follows that

$$V_P = V_G - V_{T0} - \gamma\left[\sqrt{V_G - V_{T0} - \left(\sqrt{\psi_0} + \frac{\gamma}{2}\right)^2} - \left(\sqrt{\psi_0} + \frac{\gamma}{2}\right)\right]. \tag{2.12}$$

The pinch-off voltage as a function of the gate voltage is depicted in Figure 2.3. If the gate voltage is equal to V_{T0}, the pinch-off voltage is zero.

The slope of $V_P = f(V_G)$ is not constant and is usually written as

$$\frac{dV_P}{dV_G} = \left(\frac{dV_G}{dV_P}\right)^{-1} = \frac{1}{n}, \tag{2.13}$$

where n is the **slope factor**. It is calculated from equation (2.11) as

$$n = 1 + \frac{\gamma}{2\sqrt{\psi_0 + V_P}}, \tag{2.14}$$

and since it depends on γ, it is often called the **substrate effect factor**. Its value usually varies between 1.2 to 1.8.

However, for hand-calculations the slope factor is approximated by its asymptotic value in strong inversion, and given as a technology parameter. Now, the pinch-off voltage is obtained from the first-order development of the relation (2.12) as

$$V_P \approx \frac{V_G - V_{T0}}{n(V_G)} \approx \frac{V_G - V_{T0}}{n}, \tag{2.15}$$

and this simple relation can be used for hand-calculations without assuming the strong inversion operating condition. From this relation, it can be seen that

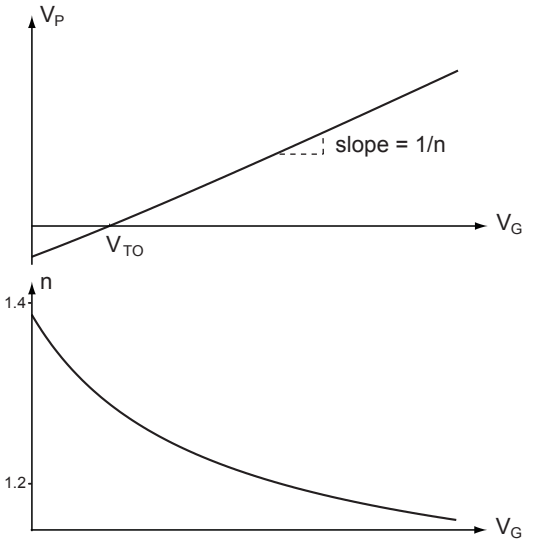


Figure 2.3 Pinch-off voltage and slope factor as functions of the gate voltage

$V_G = V_{T0} + nV_P$, and then, in more general terms, that $V_{TB} = V_{T0} + nV_{ch}$. When this is substituted in equation (2.9), the inversion charge in strong inversion is finally approximated as

$$Q_{inv} = -C_{ox} n (V_P - V_{ch}). \tag{2.16}$$

2.2.4 Weak inversion

When the channel potential becomes smaller than the pinch-off voltage, the inversion charge Q_{inv} becomes much smaller than the depletion charge Q_{dep} and the channel is in weak inversion.

Since $Q_{inv} \ll Q_{dep}$, equation (2.5) can be rewritten as

$$V_G = V_{FB} + \psi_s + \gamma\sqrt{\psi_s}\,. \tag{2.17}$$

If the flat-band voltage is replaced using the definition of V_{T0} given by (2.10), then the gate voltage is

$$V_G = V_{T0} + (\psi_s - \psi_0) + \gamma(\sqrt{\psi_s} - \sqrt{\psi_0})\,. \tag{2.18}$$

Comparing this expression to the relation (2.11), it follows that in weak inversion the surface potential can be approximated as $\psi_s = V_P + \psi_0$.

When this result is introduced in the solution of the Poisson equation, expressed by (2.4), and the Taylor expansion is applied, the inversion charge in weak inversion is approximated as

$$Q_{inv} = -C_{ox}(n-1)V_t \cdot exp\left(\frac{\Psi_0 - 2\Phi_F}{V_t}\right) \cdot exp\left(\frac{V_P - V_{ch}}{V_t}\right). \tag{2.19}$$

2.2.5 Drain current and specific current

Figure 2.4 depicts the inversion charge as a function of the channel potential. As shown previously, it is linearly dependent on (V_P-V_{ch}) in strong inversion, and it depends exponentially on (V_P-V_{ch}) in weak inversion. The transition region corresponds to moderate inversion. The drain current is determined by integrating the inversion charge along the channel, and it is proportional to the shaded surface in the same figure.

If it is assumed that the mobility does not depend on the position along the channel, the drain current is determined as

$$I_D = \beta \int_{V_S}^{V_D} -\frac{Q_{inv}}{C_{ox}} dV_{ch}. \tag{2.20}$$

The integration starts at source ($V_{ch} = V_S$) and finishes at drain ($V_{ch} = V_D$), and

$$\beta = \mu_0 C_{ox} \frac{W}{L}, \tag{2.21}$$

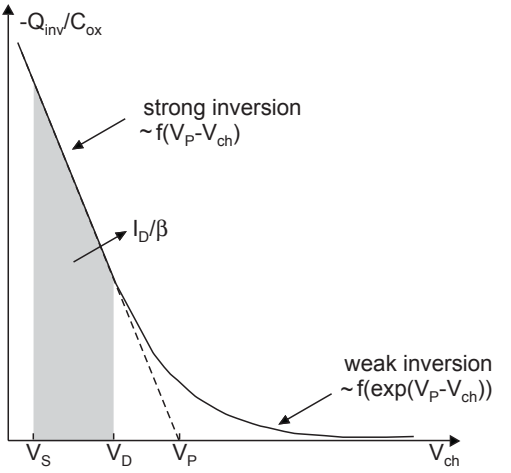


Figure 2.4 Inversion charge as a function of the channel potential

where μ_0 is the mobility, W is the transistor width and L is the transistor length. The integral can be rewritten in the following way

$$I_D = \beta \int_{V_S}^{V_D} -\frac{Q_{inv}}{C_{ox}} dV_{ch} = \beta \int_{V_S}^{\infty} -\frac{Q_{inv}}{C_{ox}} dV_{ch} - \beta \int_{V_D}^{\infty} -\frac{Q_{inv}}{C_{ox}} dV_{ch} = I_F - I_R. \tag{2.22}$$

This introduces the concept of forward (I_F) and reverse (I_R) currents, that perfectly match the transistor symmetry: if the drain and the source voltages are replaced, only the sign of the drain current changes.

Using the previously determined approximations for the inversion charge, the expressions for $I_{F,R}$ in weak/strong inversion can be determined. Nevertheless, for the design of low voltage, low power analog circuits, it is essential to have one expression covering all inversion regions. Therefore, using mathematical interpolation, a continuous function valid over all inversion levels is obtained

$$I_{F,R} = 2n\beta V_t^2 \cdot ln^2\left[1 + exp\left(\frac{V_P - V_{S,D}}{2V_t}\right)\right], \tag{2.23}$$

where the forward current depends on (V_P-V_S), and the reverse current depends on (V_P-V_D). Further, the forward and reverse currents can be normalized by

$$I_S = 2n\beta V_t^2 \tag{2.24}$$

which usually is denoted as the **specific current**.

The specific current is an important design parameter and represents the drain current when the transistor operates in the center of moderate inversion. The parameter β is usually determined as $\beta = KP \cdot (W/L)$, where $KP \approx C_{ox} \cdot \mu_0$ is given as a model parameter, called the transconductance parameter. Obviously, the specific current depends on technology and transistor geometry. However, for analog design purposes, the specific current can be interpreted for the given technology as a function of transistor geometry only. Finally, the **drain current** is written as

$$I_D = I_S \cdot (i_F - i_R), \tag{2.25}$$

where the normalized forward and reverse currents are

$$i_{F,R} = ln^2\left[1 + exp\left(\frac{V_P - V_{S,D}}{2V_t}\right)\right], \tag{2.26}$$

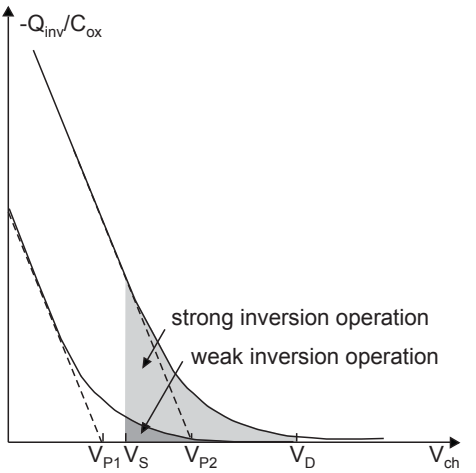


Figure 2.5 The strong and weak inversion operation

and calculated using V_P approximated as $V_P = (V_G - V_{T0})/n$.

The important point here is that the case when the source or the drain voltage is smaller than V_P corresponds to the strong inversion operation. The case when both the source and the drain voltage are larger than V_P corresponds to the weak inversion operation, as illustrated in Figure 2.5.

2.2.6 Saturation drain current and saturation voltage

On the basis of previously developed expressions, the conduction and saturation regions can be interpreted as follows: when the drain voltage is changed, as long as the forward and reverse currents are of the same order of magnitude, the transistor is in conduction. When $I_F \gg I_R$, the influence of the drain voltage becomes negligible, i.e. the transistor is in the saturation region of operation, and the drain current is equal to the **saturation drain current** $\boldsymbol{I_{Dsat}}$

$$I_D = I_S \cdot (i_F - i_R)\Big|_{i_F \gg i_R} = I_S \cdot i_F = I_{Dsat}. \tag{2.27}$$

This rather "mathematical" definition of conduction and saturation operation is not easily used for the design of analog circuits. A more convenient way to deal with the conduction/saturation operation is to compare the difference between the drain and source voltages to the design parameter called the **saturation voltage** $\boldsymbol{V_{DSsat}}$. To determine how to calculate the saturation voltage, the conduction/saturation conditions are analyzed again in strong and weak inversion.

When the transistor is operating in strong inversion, the conduction/ saturation operation can be defined with respect to the voltage V_P as:

- if $V_S, V_D < V_P$, the transistor is in conduction,
- if $V_S < V_P < V_D$, the transistor is in saturation.

The saturation voltage V_{DSsat} is calculated as $V_{DSsat} = V_P - V_S$, as shown in Figure 2.6a.

When the transistor is operating in weak inversion both $V_S, V_D > V_P$. Since the forward drain current is now a few orders of magnitude lower than in the strong inversion case and the condition $I_F >> I_R$ can not be easily verified, the saturation voltage is determined according to the acceptable error between the drain current and the expected saturation drain current (Figure 2.6b). This value is estimated, according to [19], to be several V_t.

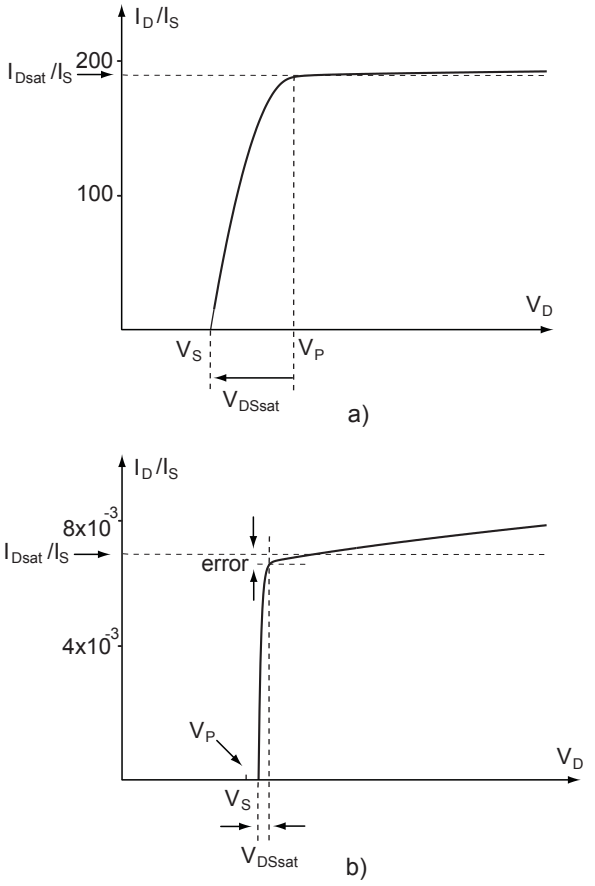


Figure 2.6 Saturation voltage definition in a) strong and b) weak inversion

On the basis of these saturation voltage asymptotes, the EKV model proposes a simple relation

$$V_{DSsat} = V_t \cdot (2\sqrt{i_F} + 4), \tag{2.28}$$

where the saturation voltage depends only on normalized forward drain current, or the inversion level as will be shown next.

2.2.7 Inversion factor as a measure of the device inversion level

The inversion level can be determined from the forward drain current I_F, which is directly proportional to the mobile inversion charge along the channel. The design parameter is called the **inversion factor *IF*** and is equal to

$$IF = \frac{I_{Dsat}}{I_S} = i_F = ln^2\left[1 + exp\left(\frac{V_P - V_S}{2V_t}\right)\right]. \tag{2.29}$$

This is an important relation because it shows that the inversion factor is determined by

- the saturation current (i.e. the transistor bias current, as it will be shown later) and the transistor geometry (via the specific current) or
- the transistor voltage bias (V_G, V_S).

When the saturation drain current is equal to the specific current, the transistor operates in the center of moderate inversion, that is $IF = 1$. The transition region of moderate inversion operation is accordingly estimated to be in the drain current range that corresponds to $0.1I_S$ up to $10I_S$. Therefore, the strong, moderate and weak operating regions are defined as in Table 2.2. This enables easy manipulation of inversion level information during the analog design procedure.

2.2.8 Transconductances

The small-signal transconductances are defined as changes in the drain current due to small variations of gate, source and drain voltages.

Table 2.2 Operating region definitions with respect to the inversion factor

IF < 0.1	weak inversion
0.1 < IF < 10	moderate inversion
IF > 10	strong inversion

$$i_D = \frac{dI_D}{dV_G} \cdot v_G + \frac{dI_D}{dV_S} \cdot v_S + \frac{dI_D}{dV_D} \cdot v_D = g_{mG} \cdot v_G - g_{mS} \cdot v_S + g_{mD} \cdot v_D. \tag{2.30}$$

The **small-signal gate transconductance g_{mG}**, usually denoted as g_m and referred to as the transistor transconductance, represents the first derivative of the drain current in terms of the gate voltage

$$g_{mG} = \frac{dI_D}{dV_G} = \frac{d(I_F - I_R)}{dV_P} \cdot \frac{dV_P}{dV_G} = g_m \,. \tag{2.31}$$

The **small-signal source transconductance g_{ms}** is determined as the first derivative of the drain current in terms of the source voltage

$$g_{mS} = \frac{dI_D}{dV_S} = \frac{dI_F}{dV_S}. \tag{2.32}$$

At this point, it is important to mention that the conventional "Spice-like" small-signal *substrate* transconductance, obtained as a current derivative in the case when the source is considered as a reference, is

$$g_{mB} = \frac{dI_D}{dV_{BS}} = g_{mS} - g_{mG} - g_{mD}, \tag{2.33}$$

since in our case the gate, source and drain voltages are determined with respect to the substrate.

The **small-signal drain transconductance**, representing the first derivative of the drain current in terms of the drain voltage

$$g_{mD} = \frac{dI_D}{dV_D} = -\frac{dI_R}{dV_D} \tag{2.34}$$

is usually denoted as the small-signal drain-source transconductance g_{DS}. This is correct because the influence of the source voltage can be neglected when the reverse current is derived.

If we now recall that the forward/reverse currents $I_{F,R}$ are functions of $(V_P - V_{S,D})$ it follows that

$$\frac{dI_{F,R}}{dV_P} = -\frac{dI_{F,R}}{dV_{S,D}}. \tag{2.35}$$

On the other hand, the pinch-off voltage is approximated as $V_P = (V_G - V_{T0})/n$ and then

$$\frac{dV_P}{dV_G} = \frac{1}{n}, \tag{2.36}$$

and, in the saturation region,

$$g_m = \frac{d(I_F - I_R)}{dV_P} \cdot \frac{dV_P}{dV_G} = \frac{1}{n} \cdot (g_{mS} - g_{DS}) \approx \frac{g_{mS}}{n}. \tag{2.37}$$

Therefore, equation (2.30) can be rewritten as

$$i_D = g_m \cdot v_G - n \cdot g_m \cdot v_S + g_{DS} \cdot v_{DS}, \tag{2.38}$$

and gives the small-signal equivalent scheme drawn in Figure 2.7.

The gate-source transconductance is represented as a resistor, because it is controlled by the voltage on its terminals. Hence, it represents the transistor **output conductance**, which is an important design parameter. For the analog design hand-calculations, the output conductance is approximated using the Early voltage V_a as

$$g_{DS} = \frac{I_{Dsat}}{L \cdot V_a}. \tag{2.39}$$

2.2.9 Normalized transconductance

Deriving the drain current, and then, the EKV model interpolation function, proposed in [17], that links weak, moderate and strong inversion, we obtain

$$\frac{g_m}{I_{Dsat}} = \frac{1}{nV_t} \cdot \frac{1}{\frac{1}{2} + \sqrt{i_F + \frac{1}{4}}}, \tag{2.40}$$

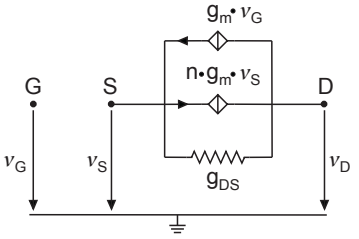


Figure 2.7 Small-signal MOS transistor equivalent scheme with transconductances

$$\frac{g_{mS}}{I_F} = \frac{1}{V_t} \cdot \frac{1}{\frac{1}{2} + \sqrt{i_F + \frac{1}{4}}}, \text{ and} \tag{2.41}$$

$$\frac{g_{DS}}{I_R} = \frac{1}{V_t} \cdot \frac{1}{\frac{1}{2} + \sqrt{i_R + \frac{1}{4}}}. \tag{2.42}$$

The normalized transconductance, derived from equation (2.40) and defined as

$$\frac{g_m}{I_{Dsat}} n V_t = \frac{1}{\frac{1}{2} + \sqrt{i_F + \frac{1}{4}}} \tag{2.43}$$

becomes a fundamental relation for analog design. It provides the relation between the small-signal parameter (the transistor transconductance) and the DC parameters (the drain saturation current and the inversion factor). In fact, it represents a measure of the translation of polarization current into transconductance for different inversion levels. Hence, it is often called the **transconductance efficiency factor** *TEF*. The relation (2.43) is universal, and does not depend on technology. In weak inversion it approaches 1 asymptotically, whereas the strong inversion asymptote is $1/\sqrt{i_F}$ (Figure 2.8).

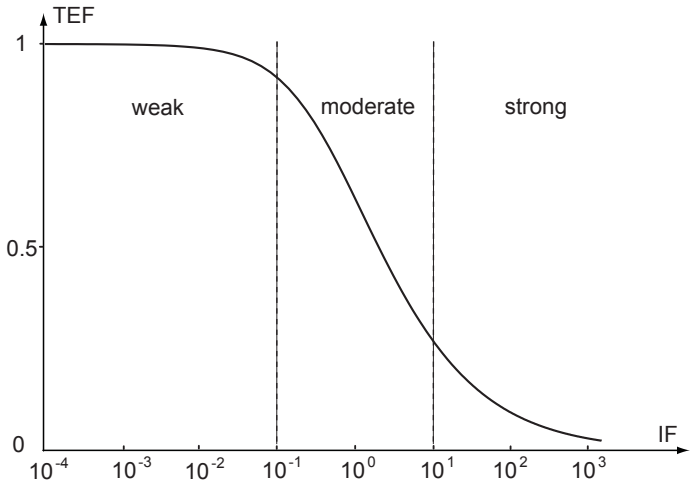


Figure 2.8 Normalized transconductance as a function of the inversion factor

2.2.10 Capacitances

The variation of global charge (Q_{inv}, Q_G, Q_d) with respect to the variations of terminal voltages is modeled through the intrinsic capacitances. They are called "intrinsic" because they refer to the intrinsic part of the device (consisting of the inversion layer, the depletion region, the oxide and the gate) that determines the transistor behavior. The rest of the device represents the extrinsic part and influences the parasitic effects that usually limit the overall response.

The **intrinsic capacitances** are:

- the capacitances between the gate and the other terminals, defined as the derivatives of the gate charge with respect to the source, drain and bulk voltage:

$$C_{GSi} = -\frac{dQ_G}{dV_S}, \tag{2.44}$$

$$C_{GDi} = -\frac{dQ_G}{dV_D}, \text{ and} \tag{2.45}$$

$$C_{GBi} = -\left.\frac{dQ_G}{dV_B}\right|_{V_{S,D,G}=const} \tag{2.46}$$

- the capacitances between the substrate and the source/drain terminal, defined as the derivatives of the depletion charge with respect to the source and drain voltage:

$$C_{BSi} = -\frac{dQ_{dep}}{dV_S}, \text{ and} \tag{2.47}$$

$$C_{BDi} = -\frac{dQ_{dep}}{dV_D}. \tag{2.48}$$

This way of modeling is valid only for the quasi-static behavior, that is "if the terminal voltages vary sufficiently slow, so that the charge distribution in the channel can follow with the negligible inertia", as stated in [15]. According to [8] and [15], the quasi-static behavior is valid up to

$$\omega \ll \omega_0 = \frac{\mu_n}{L^2} \cdot (V_P - V_S). \tag{2.49}$$

For higher frequencies, a different modeling approach is required. However, the design methodology, discussed in this book, is dedicated to the analog circuits for which the non-quasi static model is not required.

The intrinsic capacitances are calculated in [17] from the normalized forward and reverse drain currents as

$$C_{GSi} = C_{ox}WLc_{GSi}, \text{ where } c_{GSi} = \frac{2}{3}\left(1 - \frac{x_R^2 + x_R + \frac{1}{2}x_F}{(x_F + x_R)^2}\right), \tag{2.50}$$

$$C_{GDi} = C_{ox}WLc_{GDi}, \text{ where } c_{GDi} = \frac{2}{3}\left(1 - \frac{x_F^2 + x_F + \frac{1}{2}x_R}{(x_F + x_R)^2}\right), \tag{2.51}$$

$$C_{GBi} = C_{ox}WLc_{GBi}, \text{ where } c_{GBi} = \frac{n-1}{n}(1 - c_{GSi} - c_{GDi}), \tag{2.52}$$

$$C_{BSi} = C_{ox}WLc_{BSi}, \text{ where } c_{BSi} = (n-1)c_{GSi}, \tag{2.53}$$

$$C_{BDi} = C_{ox}WLc_{BDi}, \text{ where } c_{BDi} = (n-1)c_{GDi}, \tag{2.54}$$

and $x_F = \sqrt{i_F + \frac{1}{4}}$, $x_R = \sqrt{i_R + \frac{1}{4}}$. In the saturation region, we can write $i_R \ll \frac{1}{4}$, and $i_F \gg i_R$. As a result,

$$c_{GSi} = \frac{2}{3}\frac{(x_F + 1)\left(x_F - \frac{1}{2}\right)}{\left(x_F + \frac{1}{2}\right)^2}, \; c_{GDi} = 0. \tag{2.55}$$

The **extrinsic capacitances** are:

- the overlap capacitances between the gate and the source, drain or substrate are determined in the following way

$$C_{GSov} = C_{GSO}W, \tag{2.56}$$

$$C_{GDov} = C_{GDO}W, \tag{2.57}$$

$$C_{GBov} = C_{GBO}L\,, \tag{2.58}$$

where C_{GSO}, C_{GDO} and C_{GBO} are the parameters of the technology;

- and the source/drain junction capacitances, C_{jS} and C_{jD}.

The complete small-signal equivalent scheme, including transconductances and capacitances, is shown in Figure 2.9. For an "analog" transistor operating in saturation, the equivalent scheme is simplified as shown in Figure 2.10.

2.2.11 Intrinsic gain

The transistor intrinsic gain is defined as the ratio of its transconductance to its output conductance

$$A_i = \frac{g_m}{g_{DS}} = \frac{g_m}{I_{Dsat}} \cdot \frac{I_{Dsat}}{g_{DS}} = \left(\frac{g_m}{I_{Dsat}}\right) \cdot L \cdot V_a\,. \tag{2.59}$$

This parameter shows the technology limit, that is the gain that can be achieved with a one-gain stage under given bias conditions. It depends on the g_m/I_{Dsat} ratio, and thus the inversion level, as well as on transistor length.

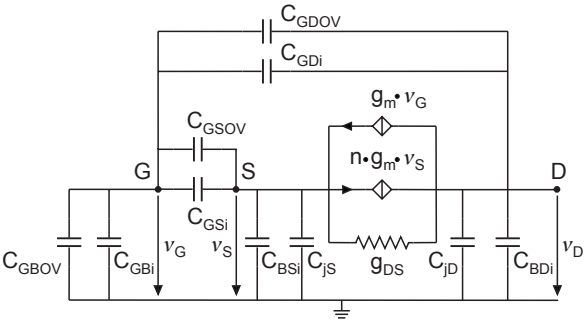


Figure 2.9 Complete small-signal MOS transistor equivalent scheme including transconductances and capacitances

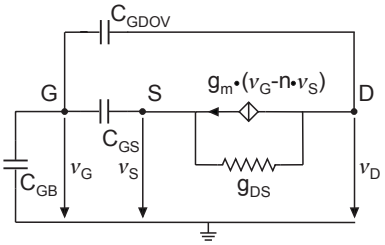


Figure 2.10 Simplified small-signal MOS transistor equivalent scheme

2.2.12 Transition frequency

The transition frequency is the frequency where the magnitude of the transistor current gain becomes unity. It can be approximated as

$$f_t = \frac{g_m}{2\pi(C_{GS}+C_{GB}+C_{GD})}. \tag{2.60}$$

It represents the maximal frequency for which the transistor can be used, and it is a measure of transistor speed.

2.2.13 Intrinsic noise

The intrinsic noise is modeled as either a spectral density of drain current fluctuations or a spectral density of gate voltage fluctuations, as drawn in Figure 2.11.

It consists of the thermal noise component and the flicker noise component. The thermal noise current spectral density, according to [8], can be approximated as a function of the inversion factor by

$$i_{n,th}^2 = 4KT \cdot g_{mS} \cdot \frac{1}{1+i_F} \cdot \left(\frac{1+\alpha}{2} + \frac{2}{3} i_F \frac{1+\sqrt{\alpha}+\alpha}{1+\sqrt{\alpha}} \right) \tag{2.61}$$

where $\alpha = i_R / i_F$. If the transistor operates in saturation, it becomes

$$v_{n,th}^2 = 4KT \cdot g_{mS} \cdot \frac{1}{1+i_F} \cdot \left(\frac{1}{2} + \frac{2}{3} i_F \right). \tag{2.62}$$

The **thermal noise voltage spectral density** is then calculated as

$$v_{n,th}^2 = \frac{S_{Ith}}{g_m^2} = \frac{4KT \cdot n \cdot \frac{1}{1+i_F} \cdot \left(\frac{1}{2} + \frac{2}{3} i_F \right)}{g_m}. \tag{2.63}$$

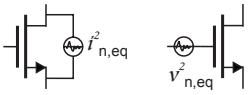


Figure 2.11 Intrinsic noise modeled as spectral density of a) current or b) voltage fluctuations

The **flicker noise voltage spectral density** is determined as

$$v_{n,fl}^2 = \frac{KF}{WLC_{ox} f^{AF}}, \tag{2.64}$$

where f is the frequency and KF, AF are the technology parameters. The equivalent input-referred noise voltage spectral density is then

$$v_{n,eq}^2 = v_{n,th}^2 + v_{n,fl}^2. \tag{2.65}$$

The frequency where the flicker noise component becomes equal to the thermal noise component is called the **corner frequency** and is equal to

$$f_{cor} = \sqrt[AF]{\frac{KF \cdot g_m}{4KT \cdot WLC_{ox} \cdot n \cdot \frac{1}{1+i_F} \cdot \left(\frac{1}{2} + \frac{2}{3} i_F\right)}}. \tag{2.66}$$

2.3 Design approach

Each transistor in a circuit is seen as a combination of one or several design parameters. When the circuit is partitioned into basic analog structures, the specifications for each basic analog structure are derived from the circuit specifications. All polarization currents are determined from the requirement for either maximal current dissipation or minimal acceptable speed. Since each basic analog structure consists of one or several transistors that realize a certain analog function, the required parameters for each transistor are extracted from the analog structure specifications. The design at the transistor level thus consists in the calculation of the values of the transistor design variables to achieve the specified design parameters of the analog structure with the given polarization current.

2.3.1 Design parameters and design variables

Transistor parameters that are required for the design of basic analog structures are:

- saturation voltage V_{DSsat},
- transconductance g_m,
- output conductance g_{DS},
- parasitic capacitances (intrinsic/extrinsic),
- gain A_i,

- transition frequency f_t, and
- equivalent noise.

They are all defined in section 2.2, and closer inspection shows that they depend on some of the following variables:

- saturation current I_{Dsat},
- inversion factor IF,
- transistor width W,
- transistor length L,
- ratio W/L, and
- area WL.

These variables are obviously interdependent, and this complicates the task of transistor sizing which entails transistor width and length calculation in the traditional design approach. Therefore, the problem of the design on transistor level must be posed differently.

Since a transistor in an analog circuit usually operates in the saturation region, the given polarization current represents the saturation drain current

$$I_{pol} = I_{Dsat} = 2nKP\left(\frac{W}{L}\right)V_t^2 \cdot IF. \tag{2.67}$$

Hence, the saturation drain current is not an independent variable at the transistor level, as it is imposed by the circuit-level design. Now if the inversion factor IF is specified, the ratio W/L can be calculated directly as

$$\frac{W}{L} = \frac{I_{Dsat}}{2nKPV_t^2 \cdot IF} \tag{2.68}$$

and vice versa. Since the transistor length affects important design parameters such as the output conductance and the transition frequency, we propose to express the transistor width as a function of the inversion factor and the transistor length as

$$W = \left(\frac{W}{L}\right) \cdot L = \frac{I_{Dsat}}{2nKPV_t^2 \cdot IF} \cdot L, \tag{2.69}$$

and thus to eliminate the transistor width as a design variable. In the same way, the transistor surface can be rewritten as

$$WL = \left(\frac{W}{L}\right) \cdot L^2 = \frac{I_{Dsat}}{2nKPV_t^2 \cdot IF} \cdot L^2. \tag{2.70}$$

As a result, the problem of transistor sizing can be treated as the calculation of the inversion factor and the transistor length to achieve the required design parameters with the given polarization current.

2.3.2 Table of design parameters as functions of the design variables

Using the conclusions from the previous subsection, all important transistor design parameters are summarized in Table 2.3. They are expressed in terms of the two independent design variables. The ratio g_m/I_{Dsat} is used instead of the transconductance g_m as a design parameter, since it is a function of the inversion factor only, and thus more convenient for the analog design as demonstrated in [20]-[22]. On the other hand, the sum of all intrinsic capacitances is approximated as

$$\Sigma C_i = C_{GSi} + C_{GDi} + C_{GBi} \approx C_{ox} \cdot WL = (\Sigma C_i)_{MAX}, \tag{2.71}$$

where WL is replaced by the expression (2.70). In a similar way, the transition frequency is recalculated as a function of the inversion factor and the transistor length. The summary of the design parameters and their dependence on design variables is indicated in Figure 2.12. The g_m/I_{Dsat} ratio, the saturation voltage and the output conductance depend on only one design variable and when these parameters are given as a design specification, the transistor length or the inversion factor is calculated directly. In contrast, the two-variable design parameters demand either two design conditions or the optimization strategy.

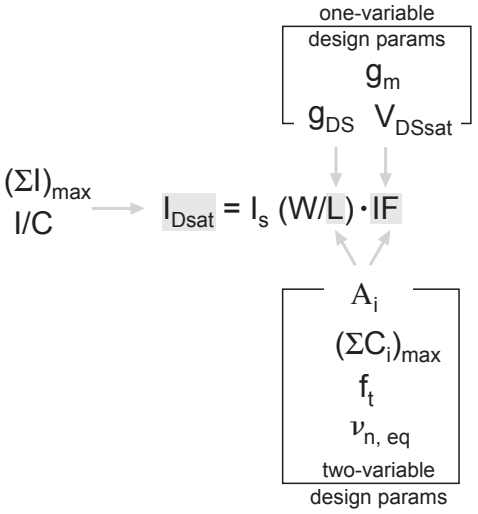


Figure 2.12 Relations between the design parameters and the design variables

Table 2.3 Table of transistor design parameters given as functions of the inversion factor and the transistor length

saturation voltage	$$V_{DSsat} = V_t \cdot (2\sqrt{IF} + 4)$$
g_m/I_{Dsat} ratio	$$\frac{g_m}{I_{Dsat}} = \frac{1}{nV_t} \cdot \frac{1}{\frac{1}{2} + \sqrt{IF + \frac{1}{4}}}$$
output conductance	$$g_{DS} = \frac{I_{Dsat}}{L \cdot V_a}$$
capacitances (in saturation)	$$C_{GS} = C_{ox} \cdot \frac{I_{Dsat}}{2nKPV_t^2 \cdot IF} \cdot L^2 \cdot \frac{2}{3} \frac{(x_F + 1)\left(x_F - \frac{1}{2}\right)}{\left(x_F + \frac{1}{2}\right)^2}$$ $$+ C_{GSO} \cdot \frac{I_{Dsat}}{2nKPV_t^2 \cdot IF} \cdot L$$ $$C_{GD} = C_{GDO} \cdot \frac{I_{Dsat}}{2nKPV_t^2 \cdot IF} \cdot L$$ $$(\Sigma C_i)_{MAX} = C_{ox} \cdot \frac{I_{Dsat}}{2nKPV_t^2 \cdot IF} \cdot L^2$$
intrinsic gain	$$A_i = \left(\frac{g_m}{I_{Dsat}}\right) \cdot L \cdot V_a$$
transition frequency	$$f_t = \frac{g_m}{2\pi(\Sigma C_i)_{MAX}} = \left(\frac{g_m}{I_{Dsat}}\right) \cdot \frac{nKPV_t^2 \cdot IF}{\pi C_{ox} \cdot L^2}$$
noise	$$v_{n,th}^2 = \frac{4KT \cdot n \cdot \frac{1}{1 + IF} \cdot \left(\frac{1}{2} + \frac{2}{3}IF\right)}{\left(\frac{g_m}{I_{Dsat}}\right) \cdot I_{Dsat}}$$ $$v_{n,fl}^2 = \frac{2nKPV_t^2 \cdot KF \cdot IF}{L^2 C_{ox} f^{AF} \cdot I_{Dsat}}$$

2.3.3 Analyses of design parameters as functions of the design variables

The design parameters are analyzed in this subsection using the example of one 0.35 μm CMOS technology with the objectives to identify the parameter bounds (maximal and minimal values), and to identify the possible optimal choices (for the situations when there are several design requirements).

Saturation voltage - Figure 2.13 shows the saturation voltage as a function of the inversion factor. The smallest saturation voltage can be achieved when a transistor is operating in weak inversion, and its value is estimated to several V_t as stated in subsection 2.2.6. It is important to note that this is the minimal possible value that can be expected with any technology. Therefore, to ensure that the minimal saturation voltage is achieved, it is sufficient to chose the transistor operating region at the limit of weak inversion ($IF = 0.1$). When it is necessary to move towards strong inversion because of some other parameter (e.g. to improve the current matching), the best design choice is to place the transistor at the limit of strong inversion ($IF = 10$), since the saturation voltage starts to increase quadratically with the inversion factor in strong inversion.

***g_m/I_{Dsat}* ratio** - The ratio g_m/I_{Dsat}, illustrated as a function of the inversion factor in Figure 2.14, achieves a maximum in weak inversion, whereas it is proportional to $1/(\sqrt{IF})$ in strong inversion. The maximal value $1/(nV_t)$ depends on technology, since it depends on n, and varies from 27 up to 32 at ambient temperature. This is a very important parameter because it shows the maximal transconductance that can be achieved with the given polarization current.

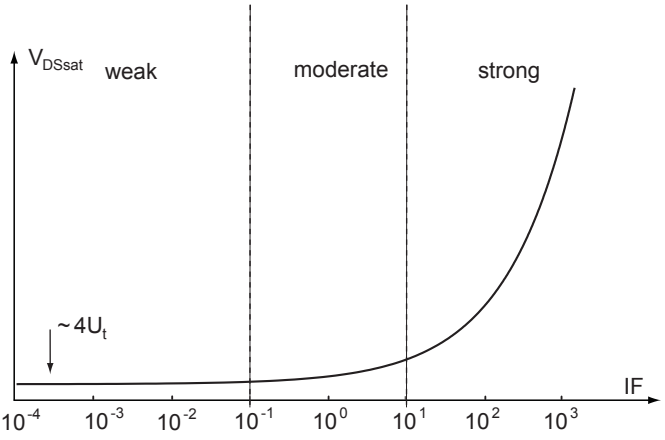


Figure 2.13 Saturation voltage as a function of the inversion factor

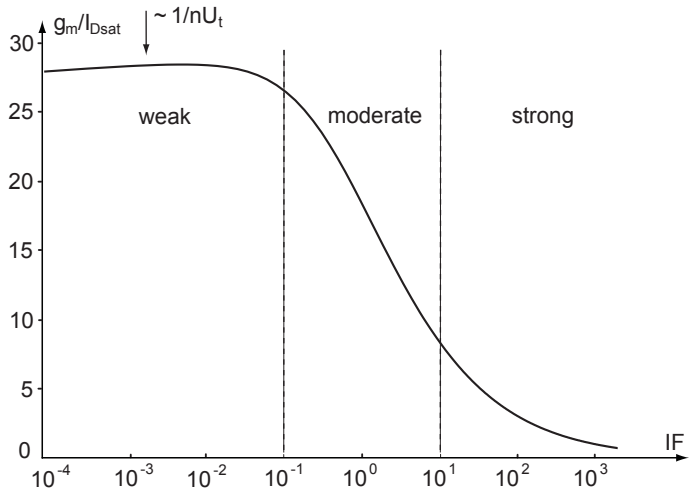


Figure 2.14 g_m/I_{Dsat} ratio as a function of the inversion factor (0.35 μm CMOS technology)

Intrinsic gain - As it is directly proportional to the g_m/I_{Dsat} ratio, the maximal gain is obtained when the transistor operates in weak inversion. Figure 2.15 gives an example of the intrinsic gain as a function of the inversion factor for different transistor lengths for one 0.35 μm CMOS technology. Obviously, a longer L provides a larger gain. A key point here is to notice that when the transistor operates in the center of moderate inversion ($IF = 1$), the gain value almost approaches its maximum, and thus it may often be an optimal design choice.

It is also of interest to observe the gain characteristics in Figure 2.16 which confirm that the intrinsic gain does not depend on the polarization current, when the inversion factor is kept constant. We can see from equation (2.59) that even though the transconductance value increases with the saturation drain current (since the g_m/I_{Dsat} ratio stays the same), the output conductance also increases with the polarization current, and thus the intrinsic gain can be considered the same.

The most common design situation is depicted in Figure 2.17. Here, the polarization current is kept constant, and the transistor length is changed from the minimal value allowed by technology L_{MIN} up to $100L_{MIN}$, moving in this way the transistor operating region towards strong inversion. Nevertheless, when L increases, the overall gain also increases.

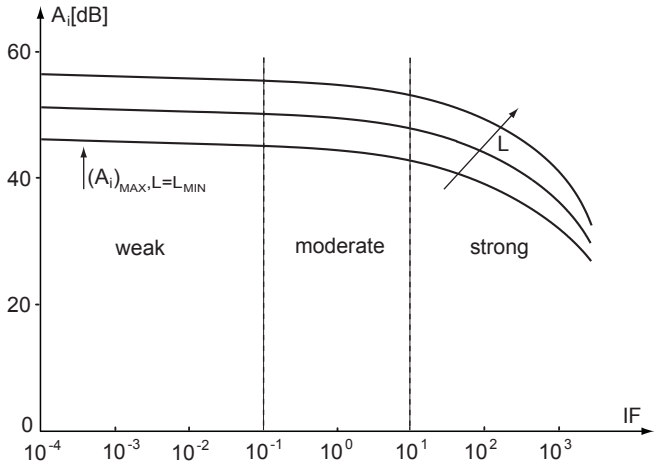


Figure 2.15 Intrinsic gain as a function of the inversion factor for different transistor lengths: $1.5L_{MIN}$, $3L_{MIN}$, $6L_{MIN}$ (0.35 μm CMOS technology)

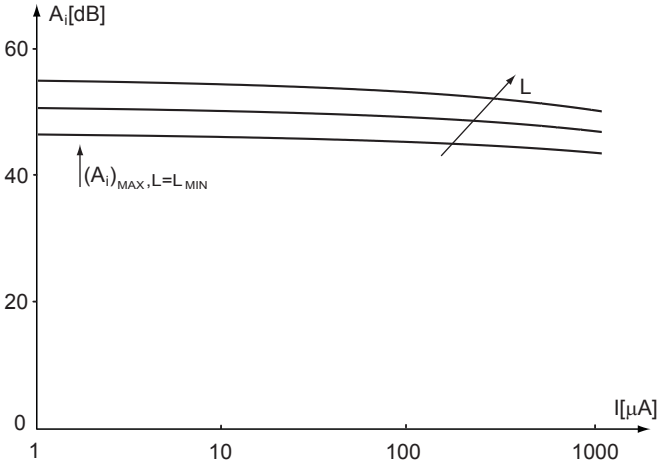


Figure 2.16 Intrinsic gain as a function of the polarization current (when *IF* = const.) for different transistor lengths: $1.5L_{MIN}$, $3L_{MIN}$, $6L_{MIN}$ (0.35 μm CMOS technology)

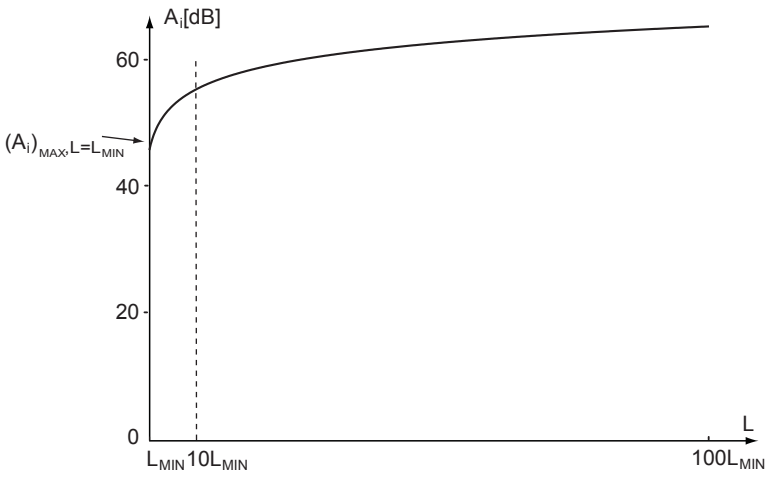


Figure 2.17 Intrinsic gain as a function of the transistor length when the polarization current is constant (0.35 μm CMOS technology)

Capacitances - Figure 2.18 presents an example of the gate-source C_{GS} and the gate-drain C_{GD} capacitance trends when the inversion factor is changed, and the transistor area (WL) is kept constant. As expected, the minimal C_{GS} is in weak inversion. The bottom value is the overlap capacitance C_{GSov} and when the transistor is in the saturation region, C_{GDi} is negligible with respect to the overlap capacitance C_{GDov}.

Unfortunately, the described situation rarely occurs in real design. More often, the saturation drain current is constant (it is imposed by the circuit requirements), and the transistor length is determined by another design parameter. When the inversion factor decreases, the transistor width and the transistor area augment. Hence, as Figure 2.19 clearly shows, weak inversion is the worst operating region, because the gate-source capacitance achieves there its maximum.

Figure 2.20 depicts another situation: the saturation current is constant, and the transistor width is fixed. When the transistor length is changed, the inversion factor is consequently changed. In this case, a longer L means moving towards strong inversion, but also a larger surface and thus a larger C_{GS}.

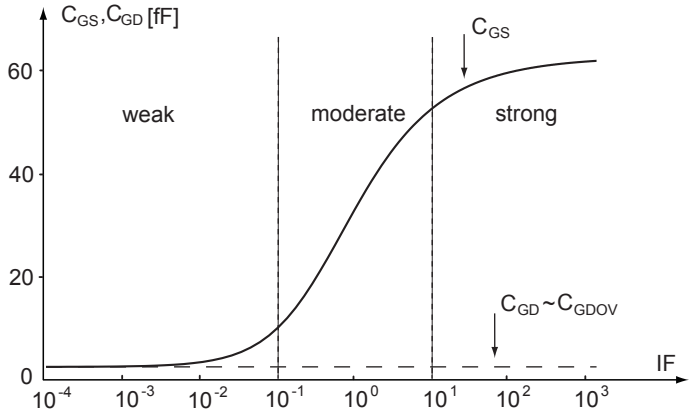


Figure 2.18 Gate-source and gate-drain capacitances as functions of the inversion factor when *WL* = const. (0.35 μm CMOS technology, $L = L_{MIN}$, W = 50 μm, COX = 4.48 mF/m²)

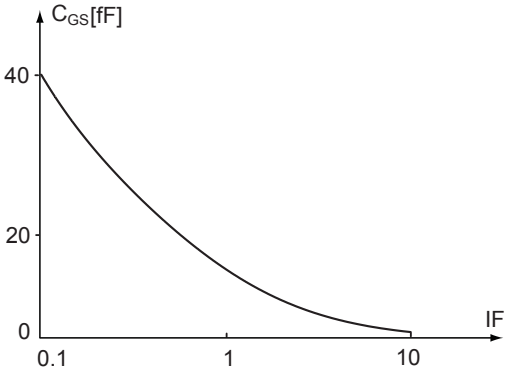


Figure 2.19 Gate-source capacitance as a function of the inversion factor, when the polarization current and the transistor length are constant (0.35 μm CMOS technology, $L = L_{MIN}$, COX = 4.48 mF/m²)

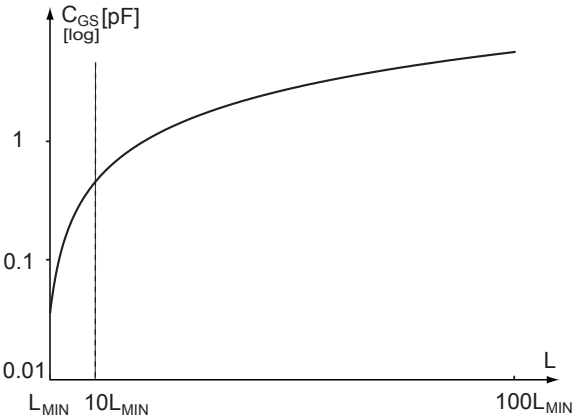


Figure 2.20 Gate-source capacitance as a function of the transistor length when the polarization current and the transistor width are kept constant (0.35 μm CMOS technology, *W* = 50 μm, *COX* = 4.48 mF/m²)

Transition frequency - Figure 2.21 depicts the transition frequency as a function of the inversion factor for different transistor lengths. Higher transition frequency is achieved when the transistor operates in strong inversion and its length is as small as possible.

The situation when the polarization current is constant and the transistor length is changed from L_{MIN} up to $100L_{MIN}$ is shown in Figure 2.22. Clearly, to obtain the maximal f_t, the minimal transistor length is a must. In the case when the transistor length has to be increased due to some other design parameter, the acceptable limit that we propose here is $L = 10L_{MIN}$.

Noise - The thermal noise voltage root spectral density as a function of the inversion factor, when the transistor area is kept constant, is depicted in Figure 2.23. It decreases when the transistor operating region moves towards strong inversion.

The flicker noise voltage root spectral density is inversely proportional to the transistor area. Thus, it is of interest to observe the situation when the transistor length is imposed by another design parameter, and the polarization current is constant. In this case, the weak inversion operation means a larger transistor width and a larger surface, and thus a smaller flicker noise contribution (Figure 2.24).

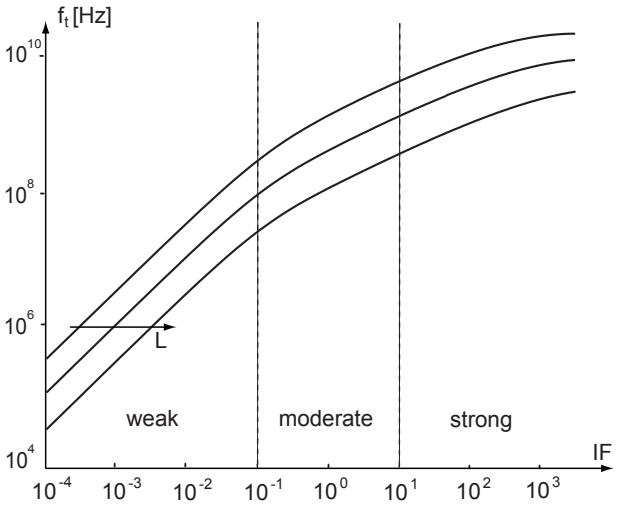


Figure 2.21 Transition frequency as a function of the inversion factor for different transistor lengths: 1.5L_{MIN}, 3L_{MIN}, 6L_{MIN} (0.35 μm CMOS technology, *KP* = 243 μA/V²)

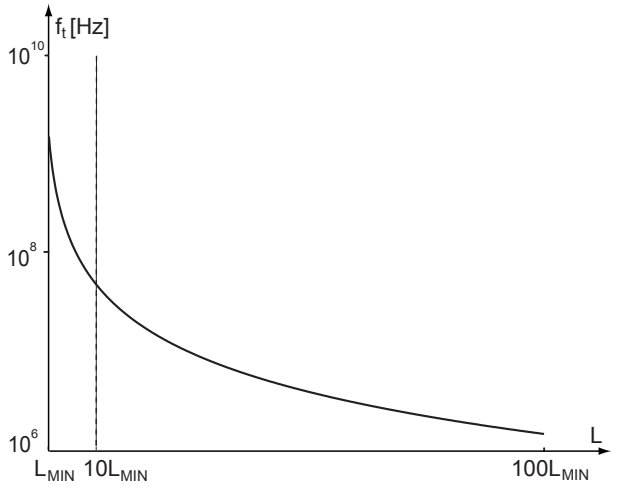


Figure 2.22 Transition frequency as a function of the transistor length when the polarization current is constant (0.35 μm CMOS technology, *KP* = 243 μA/V²)

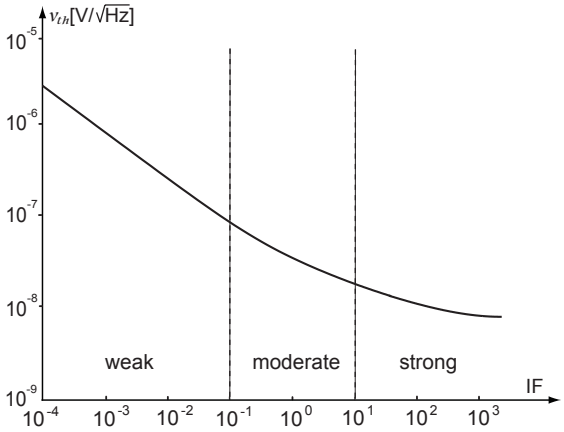


Figure 2.23 Thermal noise voltage root spectral density as a function of the inversion factor, *WL* = const. (0.35 μm CMOS technology, *L* = L_{MIN}, *W* = 50 μm, *KP* = 243 μA/V²)

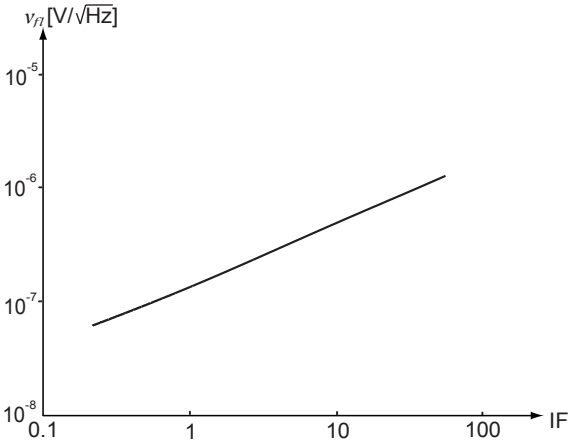


Figure 2.24 Flicker noise voltage root spectral density at *f* = 1 Hz as a function of the inversion factor when the transistor length and the polarization current are constant (0.35 μm CMOS technology, *L* = L_{MIN}, *I* = 10 μA, *KP* = 243 μA/V²)

2.3.4 Design steps

Design at the transistor level includes the following steps:

- check of design parameters limits,
- determination of the bounds of design variables,
- calculation of the design variables to achieve the specified or the closest value of each specified design parameter.

Check of design parameters limits - Before any calculation at the transistor level, the first thing to be verified is if the specified parameter is within the technology limits. For example, it is not possible to expect a saturation voltage smaller than $4V_t$, or parasitic capacitances smaller than the overlap capacitances calculated for the minimal transistor width/length. Similarly, if the required transconductance is larger than $I_{Dsat}/(nV_t)$, the problem is not in the transistor sizing. In this case, to make it possible to achieve the required value, the polarization current must be augmented.

If the specified design parameters are outside the technology limits, the design must be revised at the circuit level. This topic will be treated in the following chapters.

Determination of the bounds of design variables - Minimal transistor length and width are defined by the fabrication process. However, minimal transistor lengths are often avoided because of short channel effects, and it is proposed not to use transistor lengths smaller than $2L_{MIN}$. In addition, to achieve better current/voltage matching, it could be decided to increase the minimal acceptable transistor length for some transistors to $5L_{MIN}$.

On the other hand, maximal transistor area is imposed by the total circuit layout area constraint. For example, if L is fixed, moving towards deep weak inversion means increasing the transistor width, which can become “too wide”. As a result, maximal transistor width or length may be defined for certain design situations.

Calculation from two design specifications - If the polarization current and two design parameters are specified from the design at the higher level, the transistor length and the inversion factor can be calculated directly.

The most common design situations include the following combinations:

- output conductance + saturation voltage,
- output conductance + transconductance,
- output conductance + gain,
- transconductance + gain,
- saturation voltage + gain,
- output conductance + sum of parasitic capacitances,
- transconductance + sum of parasitic capacitances,
- output conductance + transition frequency,
- transconductance + equivalent noise,

- gain + transition frequency,
- gain + sum of parasitic capacitances, and
- gain + equivalent input referred noise.

The design recipe for each case is given in the next subsection.

Optimization strategy - When one or both of the specifications are expressed as requirements to minimize or maximize a certain design parameter, the design task consists in searching for the best design variables' values. A similar situation may occur when there are more than two design specifications. Following the analyses in subsection 2.3.3, the optimal design regions are proposed for each design recipe in the next subsection.

2.3.5 Design recipes

Output conductance + saturation voltage - The transistor length is calculated using the approximation (2.39), whereas the inversion factor is calculated using (2.28). Thus,

$$g_{DS} \rightarrow L = \frac{I_{Dsat}}{g_{DS} \cdot V_a}, \text{ and} \tag{2.72}$$

$$V_{DSsat} \rightarrow IF = \left[0.5 \cdot \left(\frac{V_{DSsat}}{V_t} - 4\right)\right]^2. \tag{2.73}$$

Output conductance + transconductance - The transistor length is determined as in (2.72), and the inversion factor is calculated from the g_m/I_{Dsat} ratio. Thus,

$$g_m \rightarrow \frac{g_m}{I_{Dsat}} \rightarrow IF = \left[\frac{1}{nV_t \cdot (g_m/I_{Dsat})} - \frac{1}{2}\right]^2 - \frac{1}{4}. \tag{2.74}$$

Output conductance + gain - Since the intrinsic gain is defined as the ratio of the transconductance and the output conductance, it follows that

$$A_i \rightarrow g_m = A_i \cdot g_{DS}, \tag{2.75}$$

which is equivalent to the case: output conductance + transconductance.

Transconductance + gain - Similarly to the previous case,

$$A_i \rightarrow g_{DS} = A_i/g_m, \tag{2.76}$$

which is equivalent to: output conductance + transconductance.

Saturation voltage + gain - The inversion factor is determined using the equation (2.73). Further, the g_m/I_{Dsat} ratio is calculated from the inversion factor as in (2.40). Thus,

$$V_{DSsat} \rightarrow IF \rightarrow \frac{g_m}{I_{Dsat}}. \tag{2.77}$$

The transistor length is now easily determined from the development (2.59) as

$$A_i \rightarrow L = \frac{A_i}{(g_m/I_{Dsat}) \cdot V_a}. \tag{2.78}$$

Optimization strategy - The design conditions to:
- minimize the saturation voltage, or
- achieve the maximal gain

require weak inversion as the operating region for both parameters.

Output conductance + sum of parasitic capacitances - The transistor length is calculated from the output conductance using the expression (2.72). On the other hand, the sum of parasitic capacitances is largely approximated by the equation (2.71). Therefore,

$$(\Sigma C_i) \rightarrow IF = C_{ox} \cdot L^2 \cdot \frac{I_{Dsat}}{2nKPV_t^2 \cdot (\Sigma C_i)}. \tag{2.79}$$

Optimization strategy - The design conditions can be to:
- maximize the output resistance, or
- minimize the sum of parasitic capacitances.

Since larger transistor length increases the output resistance (decreases the output conductance), the parasitic capacitances are augmented consequently. Thus, a design trade-off must be found.

According to the analysis of source-gate capacitance behavior shown in Figure 2.19, the optimal solution is in the region $1 < IF < 10$ where the capacitance value approaches its minimum. If the inversion factor is set in these bounds, the maximal transistor length is

$$(L)_{MAX} = \sqrt{\frac{(\Sigma C_i)_a}{C_{ox} \cdot \frac{I_{Dsat}}{2nKPV_t^2 \cdot IF}}}, \tag{2.80}$$

where the maximal acceptable value $(\Sigma C_i)_a$ must be estimated from the design at a higher level.

The other solution is to decrease the saturation drain current, i.e. the polarization current. As a result, the output resistance increases, whereas the

parasitic capacitances diminish. However, this decision must be confirmed at the circuit level, and may not always be possible.

A special case here can be the combination: **output conductance + sum of parasitic capacitances + saturation voltage**. When the saturation voltage has to be minimized, the output resistance maximized and the parasitic capacitances minimized, the optimal solution is, again, in the region $1 < IF < 10$.

Transconductance + sum of parasitic capacitances - The required transconductance value imposes the inversion factor as in (2.74), and the transistor length is then computed as

$$(\Sigma C_i) \rightarrow L = \sqrt{\frac{(\Sigma C_i)}{C_{ox} \cdot \frac{I_{Dsat}}{2nKPV_t^2 \cdot IF}}}. \tag{2.81}$$

Optimization strategy - To fulfill the requirements to:

- maximize the transconductance, or
- minimize the parasitic capacitances

a design trade-off needs to be found. To achieve the highest g_m the transistor must operate in weak inversion, which is the most unfavorable region for minimization of the parasitic capacitances. According to Figure 2.14 and Figure 2.19, the optimum is around $IF = 1$.

The first possibility is to set the transistor length to L_{MIN}, and then calculate IF according to the required transconductance value. However, this solution is acceptable only if this places the transistor in moderate inversion operation. The better solution is to set the inversion factor to $IF = 1$, and this gives the transistor length as

$$L = \sqrt{\frac{(\Sigma C_i)_a}{C_{ox} \cdot \frac{I_{Dsat}}{2nKPV_t^2}}}, \tag{2.82}$$

where $(\Sigma C_i)_a$ is the minimal acceptable value.

Another special case is the combination: transconductance + sum of parasitic capacitances + saturation voltage. The minimization/maximization trade-off is again around $IF = 1$.

Output conductance + transition frequency - The required output conductance gives the transistor length, as in (2.72). If the transition frequency is approximated as

$$f_t = \left(\frac{g_m}{I_{Dsat}}\right) \cdot \frac{nKPV_t^2 \cdot IF}{\pi C_{ox} \cdot L^2}, \tag{2.83}$$

and the g_m/I_{Dsat} ratio is replaced by (2.40), the inversion factor is given by

$$f_t \to IF = \frac{C_{ox} f_t L^2 \pi (C_{ox} f_t^2 L^2 \pi + KPV_t)}{KP^2 V_t^2}. \tag{2.84}$$

Optimization strategy - The design conditions in this case can be to:
- maximize the transition frequency, or
- maximize the output resistance.

According to the analysis in subsection 2.3.3, maximal transition frequency is achieved when the transistor length is minimal. It is also observed that the transition frequency increases in strong inversion. Thus, a possible solution is to set $IF = 10$, and calculate the transistor length from the minimal acceptable f_t value using

$$L = \sqrt{\left(\frac{g_m}{I_{Dsat}}\right) \cdot \frac{nKPV_t^2 \cdot IF}{\pi C_{ox} \cdot f_t}}. \tag{2.85}$$

The other way to find a design trade-off is to use the result shown in Figure 2.22. On this basis, putting $L < 10L_{MIN}$ still improves the output resistance and may be acceptable. The inversion factor is then calculated from the minimal acceptable f_t value using equation (2.84).

Transconductance + equivalent input referred noise - The equivalent input noise can be approximated by its flicker noise component, since its contribution is critical and dominant in low frequencies. Therefore, when the inversion factor is determined from the g_m/I_{Dsat} ratio, the transistor length can be calculated as

$$v_{n,fl} \to L = \sqrt{\frac{KF \cdot 2nKPV_t^2 \cdot IF}{v_{n,fl} \cdot I_{Dsat} \cdot C_{ox} f^{AF}}}. \tag{2.86}$$

Optimization strategy - There is no need for trade-off decisions here, since weak inversion operation provides maximal transconductance and minimal flicker noise. Total noise can be further decreased at the cost of polarization current. Higher saturation current, when the inversion factor is kept constant, results in higher g_m minimizing the thermal noise component, and thus, the overall noise level. Nevertheless, this decision must be reconsidered from the point of view of circuit-level design and total current consumption requirement, and may not always be applicable.

Gain + transition frequency - If the gain is determined as in equation (2.59) and the transition frequency is approximated by the expression (2.83), the transistor length can be calculated as

$$L = \frac{KPV_a + \sqrt{KP^2V_a^2 - 4A_i^2C_{ox}f_tKPn^2\pi V_t}}{2A_iC_{ox}f_tn\pi}, \tag{2.87}$$

$$\text{if } KP^2V_a^2 \geq 4A_i^2C_{ox}f_tKPn^2\pi V_t. \tag{2.88}$$

Then, the g_m/I_{Dsat} ratio is calculated with respect to the required gain, which finally gives the inversion factor.

$$\frac{g_m}{I_{Dsat}} = \frac{A_i}{L \cdot V_a} \rightarrow IF. \tag{2.89}$$

Optimization strategy - When it is required to:
- maximize the gain, or
- maximize the transition frequency,

the design optimum can be found in moderate inversion, and a good initial value is $IF = 1$. If the transition frequency has higher design priority, the transistor length is determined from the approximation (2.83). Otherwise, it is calculated from the gain equation (2.59).

Gain + sum of parasitic capacitances - There are two equations: the gain given by the equation (2.59) and the sum of intrinsic capacitances approximated as (2.71). The mathematical solution is

$$IF = \frac{2A_i^2C_{ox}I_{Dsat}KPn(\Sigma C_i)V_a^2}{(A_i^2C_{ox}I_{Dsat}n - 2KP(\Sigma C_i)V_a^2)^2}, \text{ and} \tag{2.90}$$

$$L = \frac{2A_iKPn(\Sigma C_i)V_aV_t}{2KP(\Sigma C_i)V_a^2 - A_i^2C_{ox}I_{Dsat}n}, \tag{2.91}$$

giving an additional requirement

$$2KP(\Sigma C_i)V_a^2 > A_i^2C_{ox}I_{Dsat}n \tag{2.92}$$

since the transistor length cannot be negative.

Optimization strategy - The design conditions can be to:
- maximize the gain, or

- minimize the parasitic capacitances.

The optimal solution is, as in the previous case, in the center of moderate inversion. The transistor length is calculated from the acceptable value that represents higher design priority: gain or (ΣC_i).

Gain + equivalent input referred noise - When the equation (2.59) expressing the intrinsic gain, and the equation approximating flicker noise contribution

$$v_{n,fl} = \frac{KF \cdot 2nKPV_t^2 \cdot IF}{L^2 \cdot I_{Dsat} \cdot C_{ox} f^{AF}} \tag{2.93}$$

are solved, the transistor length and the inversion factor are given by

$$IF = \frac{2A_i^2 C_{ox} f^{AF} I_{Dsat} KFKPnV_a^2 v_{n,th}}{(A_i^2 C_{ox} f^{AF} I_{Dsat} nv_{n,th} - 2KFKPV_a^2)^2}, \text{ and} \tag{2.94}$$

$$L = \frac{2A_i KFKPnV_a V_t}{A_i^2 C_{ox} f^{AF} I_{Dsat} nv_{n,th} - 2KFKPV_a^2}, \tag{2.95}$$

with an additional requirement

$$A_i^2 C_{ox} f^{AF} I_{Dsat} nv_{n,th} > 2KFKPV_a^2. \tag{2.96}$$

Optimization strategy - There is no possible design problem here, since:

- maximal gain, or
- minimal flicker noise contribution

are both achieved either in weak inversion operation, or when the transistor length is augmented.

2.4 Conclusion

The inversion factor and the transistor length are identified as the only independent variables for design at the transistor level. The design, therefore, consists in the choice of the inversion factor and the transistor length values to achieve the transistor parameters specified by higher-level design. The design recipes for most common design situations are proposed. They include direct calculations, when two design specifications are given, as well as the discussion of optimal design regions in the case when it is required to

minimize/maximize a certain design parameter. Table 2.4 summarizes the most important optimization guidelines concerning the choice of the inversion operating region on the basis of the proposed transistor-level design recipes.

Table 2.4 Transistor-level design guidelines

	g_m	g_{DS}	V_{DSsat}	A_i	f_t	(ΣC)	v_n
g_m	x	direct calc	$IF \leq 0.1$	$IF \leq 0.1$	$IF = 1$	$IF = 1$	$IF \leq 0.1$
g_{DS}	direct calc	x	direct calc	direct calc	$IF \geq 10$	$1 < IF < 10$	$IF \leq 0.1$
V_{DSsat}	$IF \leq 0.1$	direct calc	x	$IF \leq 0.1$	$IF = 1$	$IF = 1$	$IF \leq 0.1$
A_i	$IF \leq 0.1$	direct calc	$IF \leq 0.1$	x	$IF = 1$	$IF = 1$	$IF \leq 0.1$
f_t	$IF = 1$	$IF = 10$	$IF = 1$	$IF = 1$	x	$IF = 10$	$IF = 1$
(ΣC)	$IF = 1$	$1 < IF < 10$	$IF = 1$	$IF = 1$	$IF = 10$	x	$IF = 1$
v_n	$IF \leq 0.1$	$IF \leq 0.1$	$IF \leq 0.1$	$IF \leq 0.1$	$IF = 1$	$IF = 1$	x

Bibliography

[1] P. Bendix, “Spice model quality: process development viewpoint”, International Symposium on Quality Electronic Design, pp. 477-481, March 2001

[2] P. Bendix, D. Foty, D. Pachura, “Practical aspects of MOS transistor model "accuracy" in modern CMOS technology”, IEEE International Conference on Electronics, Circuits and Systems, pp. 642-645, December 2004

[3] Y. Tsividis, K. Suyama, “MOSFET modeling for analog circuit CAD: problems and prospects”, IEEE Journal of Solid-State Circuits, Volume 29, pp. 210-216, March 1994

[4] SMARTSpice/UTMOST III, Modeling Manual, Volume 1, SILVACO International, March 1995

[5] BSIM model official site: www-device.eecs.berkeley.edu/~bsim3

[6] Y. Cheng, C. Hu, “MOSFET Modeling and BSIM3 User's Guide”, Kluwer Academic Publishers, 1999, ISBN 0792385756

[7] EKV model official site: legwww.epfl.ch/ekv

[8] Ch. Enz, F. Krummenacher, E. Vittoz, “An analytical MOS transistor model valid in all regions of operation and dedicated to low-voltage and low-current applications”, Analog Integrated Circuits and Signal Processing, Kluwer Academic Publishers, pp. 83-114, July 1995

[9] PHILIPS models site: www.nxp.com/Philips_Models/mos_models

[10] R. Velghe, D. Klaassen, F. Klaassen, “Compact MOS modeling for analog circuit simulation”, International Electron Devices Meeting, Technical Digest, pp. 485-488, December 1993

[11] P. Woerlee, M. Knitel, R. van Langevelde, D. Klaassen, L. Tiemeijer, A. Scholten, A. Zegers-van Duijnhoven, “RF-CMOS performance trends”, IEEE Transactions on Electron Devices, Volume 48, pp. 1776-1782, August 2001

[12] G. Gildenblat, W. Hailing, C. Ten-Lon, G. Xin, C. Xiaowen, “SP: an advanced surface-potential-based compact MOSFET model”, IEEE

Journal of Solid-State Circuits, Volume 39, pp. 1394-1406, September 2004

[13] PSP model official site: pspmodel.asu.edu

[14] W. Grabinski, B. Nauwelaers, D. Schreurs, “Transistor Level Modeling for Analog/RF IC Design”, chapter 2, Springer, 2006, ISBN 1402045557

[15] Y. Tsividis, “Operation and Modeling of the MOS Transistor”, Oxford University Press, second edition, 2003, ISBN 0195170148

[16] “The EPFL-EKV MOSFET Model Equations for Simulations”, Technical Report, Model Version 2.6, June 1997

[17] M. Bucher, “Analytical MOS Transistor Modelling for Analog Circuit Simulation”, PhD thesis N°2114, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 1999

[18] J.-M. Sallese, M. Bucher, F. Krummenacher, P. Fazan, “Inversion charge linearization in MOSFET modeling and rigorous derivation of the EKV compact model”, Solid-State Electronics, Volume 47, pp. 677-683, April 2003

[19] Ch. C. Enz, E. A. Vitoz, “Charge-based MOS Transistor Modeling: The EKV Model for Low-power and RF IC Design”, John Wiley & Sons, first edition, 2006, ISBN 9780470855416

[20] D. Binkley, M. Bucher, D. Foty, “Design-oriented characterization of CMOS over the continuum of inversion level and channel length”, IEEE International Conference on Electronics, Circuits and Systems, pp. 161-164, December 2000

[21] M. Bucher, D. Kazazis, F. Krummenacher, D. Binkley, D. Foty, Y. Papananos, “Analysis of transconductances at all levels of inversion in deep submicron CMOS”, International Conference on Electronics, Circuits and Systems, Volume 3, pp. 1183-1186, September 2002

[22] D. Foty, M. Bucher, “Re-interpreting the MOS transistor via the inversion coefficient and the continuum of g_{ms}/I_d”, International Conference on Electronics, Circuits and Systems, Volume 3, pp. 1179-1182, September 2002

CHAPTER 3

BSIM to EKV conversion

The BSIM2EKV converter has been developed as a stand-alone application that works in interaction with an external simulator. It makes it possible to extract the parameters of the EKV model from the parameters of the BSIM model using the algorithm based on standard EKV model parameters' extraction procedure and the simulated transistor characteristics. This chapter provides the details on BSIM to EKV conversion procedure, and presents the developed tool.

3.1 Motivation and purpose

The main motivation to develop the program for automatic conversion of BSIM to EKV model parameters was to allow the analog designers to use the PAD tool [1]-[2], presented in chapter 6. Since the BSIM model is an industrial standard, very often only the BSIM model parameters are provided by the foundries. On the other hand, the transistor-level calculator implemented in PAD is based on the EKV model. Therefore, the EKV model parameters are required as input data for each analog design scenario. The problem is solved by creating the BSIM2EKV conversion program as a separate module. The BSIM2EKV converter extracts automatically the parameters of the EKV model (version 2.6) from the parameters of the BSIM model (version 3.3) and creates the EKV model library file. The generated model library file can be used not only as a PAD tool input file, but also as a PSpice or HSpice-like simulator input file.

The extraction algorithm described in this chapter can be also used for the extraction of the EKV model parameters manually. It is enough to perform the first five steps of the extraction procedure, since only the intrinsic EKV model parameters are needed for the hand-calculations. This facilitates the use of g_m/I_D based methodologies [3]-[7], as well as the calculation of the initial design sizes using the EKV model approximations. Moreover, the transistor-level design approach presented in this book is based on the choice of the inversion factor whose calculation requires the intrinsic EKV model parameters.

3.2 Conversion concept

The parameter extraction algorithm is based on the "efficient parameter extraction methodology" described by M. Bucher et al. [8]. The basic idea behind the BSIM to EKV conversion is to replace the measured characteristics required for the standard extraction procedure by transistor characteristics simulated using the BSIM model library file. The EKV model parameters are then extracted in their logical order, as proposed by the standard procedure. The transistor characteristics where the corresponding parameter is dominant are simulated only in the operating regions where the two models can be compared. In this way, the parameters are determined with minimal error propagation and with maximal respect of the fundamental concepts of the two models. The extraction concept is illustrated in Figure 3.1.

The transistor characteristics are simulated using the external simulator. The present version supports the interaction with PSpice (BSIM level 7) and SMASH (BSIM level 49, i.e. HSpice-like library files). The important point is that simulator demonstration ("demo") versions are sufficient, since all simulated circuits consist of one transistor and several ideal voltage or current sources. When the simulator executive file path is specified, the BSIM2EKV program creates netlists and stimulus files and automatically runs all required simulations. The transistor characteristics are then extracted from the simulation output files and saved as arrays of points that are used later in the extraction procedure.

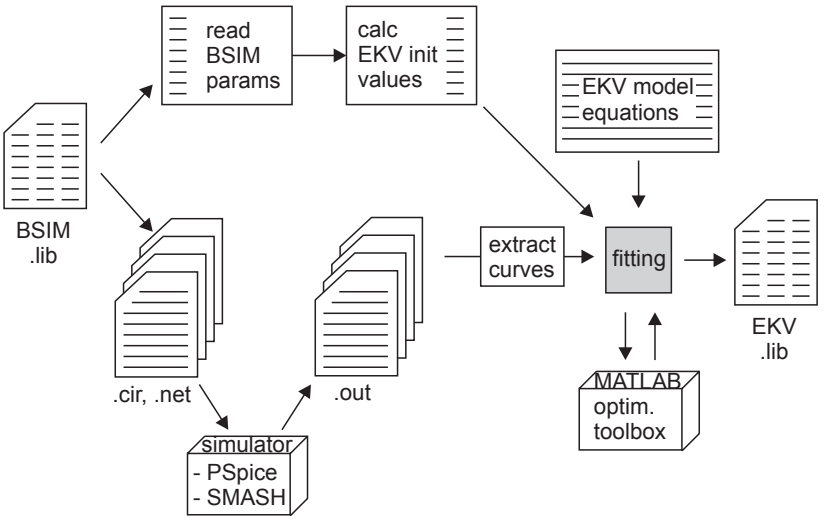


Figure 3.1 BSIM to EKV extraction concept

The conversion algorithm has been developed as a sequence of simulations of the transistor characteristics, and parameter extraction steps consisting in fitting the simulated curves by the EKV model equations. Due to the hierarchical structure of the EKV model, the optimization loops are not required. In each step, the EKV model equations, used for curve fitting, include only the previously extracted parameters. Thus, the global approach, which means extracting all parameters from all characteristics and setting the default parameters' values for non-extracted parameters, is avoided. Each parameter is determined from the set of curves where its impact is dominant. The curve fitting is based on the Levenberg-Marquardt non-linear least squares method. The initial value for each parameter is calculated from either corresponding BSIM parameters or using graphical methods. The convergence parameters and the model parameters limits are set according to [9]. The extracted EKV model library file includes the parameter sets listed in Table 3.1.

The conversion algorithm was first developed using the MATLAB optimization toolbox. The MATLAB scripts are compiled as C++ files and merged with the rest of the program that is also written in C++. As a result, the BSIM2EKV converter works as a stand-alone application.

3.3 BSIM versus EKV, the fundamental differences and the conversion algorithm guidelines

Several studies have been undertaken by P. Bendix [10] and S. Terry et al. [11] in order to determine the fundamental differences between BSIM and EKV model behavior. The parameters of each model were extracted from the measured curves, and the results compared in different operating regions. The main purpose was to draw conclusions in terms of the feasibility of simulation results. Nevertheless, the conclusions of the aforementioned authors in fundamental comparison points allowed us to determine the important guidelines for the BSIM to EKV conversion algorithm.

The principal difference between the two models is observed for the g_m/I_{Dsat} ratio as a function of the inversion level. Although the EKV model behavior is considered more idealistic, the prediction in all inversion regions is good. In the case of the BSIM model, the behavior is correct in strong and deep weak inversion, while an error of 40% is detected at the limit of weak and moderate inversion. Consequently, the strong inversion region is chosen wherever possible as the region where the parameters of the EKV model are extracted from the curves simulated using the BSIM model. The only exceptions are the $V_P = f(V_G)$ simulations. In these cases, transistors are polarized in the center of the moderate region, since it is the fundamental point of the standard EKV model parameters' extraction procedure.

Table 3.1 Extracted/calculated EKV model parameters' list

RSH	sheet resistance	[Ω/m^2]
	process related parameters	
COX	gate oxide capacitance per unit area	[F/m^2]
XJ	junction depth	[m]
DW	channel width correction	[m]
DL	channel length correction	[m]
	basic (intrinsic) model parameters	
VT0	long-channel threshold voltage	[V]
GAMMA	body effect parameter	[$\sqrt{V}$]
PHI	bulk Fermi potential	[V]
KP	transconductance parameter	[A/V^2]
E0	mobility reduction coefficient	[V/m]
UCRIT	longitudinal critical field	[V/m]
	charge sharing and channel length modulation parameters	
LAMBDA	channel length modulation	-
WETA	narrow-channel effect coefficient	-
LETA	short-channel effect coefficient	-
	reverse short-channel effect parameters	
Q0	reverse short-channel effect peak charge density	[As/m^2]
LK	reverse short-channel effect characteristic length	[m]
	impact ionization related parameters	
IBA	first impact ionization coefficient	[$1/m$]
IBB	second impact ionization coefficient	[V/m]
IBN	saturation voltage factor for impact ionization	-

Table 3.1 (continued) Extracted/calculated EKV model parameters' list

	temperature parameters	
TCV	threshold voltage temperature coefficient	[*V/K*]
BEX	mobility temperature coefficient	-
UCEX	longitudinal critical field temperature exponent	-
	noise parameters	
KF	flicker noise coefficient	-
AF	flicker noise exponent	-

The comparison of drain currents as a function of V_{GS} and their derivatives shows better behavior of the EKV model, even though the characteristics are slightly underestimated for small L. Concerning the output conductance, the conclusion is that the BSIM model has a correct trend for both long and short transistors, where there exist a relatively small error for short L. The error occurring in the case of the EKV model is explained by simplified modeling of the drain induced barrier lowering.

Therefore, the most critical step in the conversion algorithm becomes the extraction of the parameters *UCRIT* and *LAMBDA*, as will be shown in the next section. To overcome the model differences, only the initial values are determined from the characteristics $I_D = f(V_D)$ in strong, moderate and weak inversion. Using these initial values as a starting fitting point, the parameter *UCRIT* is determined from $I_D = f(V_G)$ characteristics, and the final value of *LAMBDA* is extracted from $I_D = f(V_D)$ in strong inversion only.

3.4 Conversion algorithm

The conversion algorithm starts with the initialization step where all necessary information is gathered:

- simulator executive file path,
- working directory path,
- technology information,
- BSIM model library file path.

The BSIM model library file is parsed and the parameters required for the calculation of the initial fitting points are read. The detailed diagram of all conversion steps is depicted in Figure 3.2. When all EKV parameters are extracted, the EKV model library file is created together with the conversion

report text file. The corresponding rms error is calculated for each fitting step and written in the conversion report text file. Moreover, the testbench simulation option becomes enabled. This makes it possible to compare the transistor characteristics simulated with the BSIM model and the just extracted EKV model.

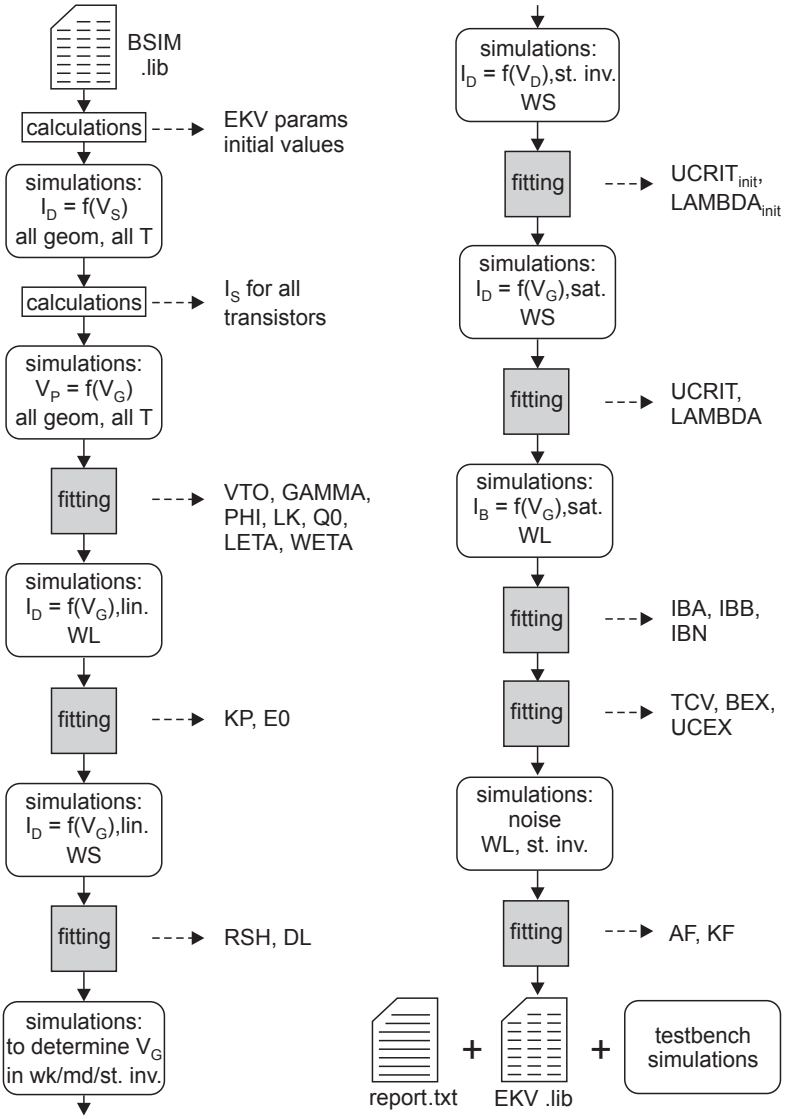


Figure 3.2 Flow-chart of BSIM to EKV conversion steps

3.4.1 Initialization

The technology information required for the conversion is:

- minimal transistor width W_{MIN},
- minimal transistor length L_{MIN},
- supply voltages V_{DD}/V_{SS},
- and parameter *HDIF*.

The parameter *HDIF* in the EKV model is calculated from layout rules as

$$HDIF = D_{cont,psedge} + 0.5W_{cont} \tag{3.1}$$

where $D_{cont,psedge}$ is the minimal distance from the contact to the polysilicon edge and W_{cont} is the minimal contact width.

According to minimal transistor width and length, the transistor geometries of interest are determined in the following way:

- wide/long transistor **WL** - W_{MAX}/L_{MAX}, where if $L_{MIN} < 0.35$ µm, then $W_{MAX} = L_{MAX} = 10$ µm, otherwise $W_{MAX} = L_{MAX} = 25$ µm,
- set of wide/short transistors **WS** - W_{MAX}/L_{MIN}, $1.2L_{MIN}$, $1.4L_{MIN}$, $1.6L_{MIN}$, $1.8L_{MIN}$, $2L_{MIN}$, $2.4L_{MIN}$, $2.8L_{MIN}$, $3.2L_{MIN}$, $3.6L_{MIN}$, $4L_{MIN}$,
- set of narrow/long transistors **NL** - W_{MIN}, $2W_{MIN}/L_{MAX}$.

The parameters that are read from the BSIM model library are:

- nominal temperature (*TNOM*),
- process parameters (*TOX, XJ, NCH*),
- parameters for the initial values' calculation (*U0, VSAT, WINT, LINT, BETA0, ALPHA1*),
- gate-overlap capacitances (*CGSO, CGDO, CGBO*),
- and junction capacitance parameters (*CJ, CJSW, MJ, MJSW*).

Using these data, the EKV and BSIM parameters are then mapped as in Table 3.2. Finally, the initial values for the fitting steps are calculated as in Table 3.3.

Table 3.2 BSIM and EKV parameter mapping

TNOM	$= TNOM_{BSIM}$
TOX	$= TOX_{BSIM}$
NSUB	$= NCH_{BSIM}$
XJ	$= XJ_{BSIM}$
DW	$= -2WINT_{BSIM}$

Table 3.2 (continued) BSIM and EKV parameter mapping

CGSO	= $CGSO_{BSIM}$
CGDO	=$CGDO_{BSIM}$
CGBO	= $CGBO_{BSIM}$ or $\varepsilon_{OX}LINT/TOX$
U0	= $U0_{BSIM}$
VMAX	= $VSAT_{BSIM}$

Table 3.3 Calculation of initial values of EKV parameters

DL	$init = -2LINT_{BSIM}$
GAMMA	$init = \frac{\sqrt{2q\varepsilon_{Si} \cdot (NSUB \cdot 10^6)}}{COX}$, $COX = e_{OX}/TOX$
PHI	$init = 2V_t(TNOM) \cdot ln\left(\frac{NSUB \cdot 10^6}{n_i(TNOM)}\right)$, where $n_i(T) = 1.45 \cdot 10^{16} \cdot \frac{T}{T_{REF}} \cdot exp\left(\frac{E_g(T_{REF})}{2V_t(T_{REF})} - \frac{E_g(T)}{2V_t(T)}\right)$, $E_g(T) = 1.16 - \left(0.000702 \cdot \frac{T^2}{T + 1108}\right)$, $V_t(T) = \frac{kT}{q}$, T_{REF} = 300.15K
KP	$init = (U0 \cdot 10^{-4}) \cdot COX$
UCRIT	$init = \frac{VMAX}{U0 \cdot 10^{-4}}$
E0	$init = \frac{0.1}{0.4 \cdot TOX}$
IBB	$init = \frac{BETA0_{BSIM}}{LC}$, $LC = \sqrt{\frac{\varepsilon_{Si}XJ}{COX}}$
IBA	$init = IBB \cdot ALPHA1_{BSIM}$

3.4.2 Specific current calculation

Before starting any simulation or fitting step, the specific current must be determined for each transistor geometry. This information is needed:

- for $V_P = f(V_G)$ simulations, and
- to correctly determine weak, moderate and strong inversion regions of operation.

  Since the saturation drain current in strong inversion is approximated by

$$I_D = \frac{n\beta}{2}(V_P - V_S)^2 = \frac{I_S}{4V_t^2}(V_P - V_S)^2, \tag{3.2}$$

it follows that

$$\frac{d\sqrt{I_D}}{dV_S} = -\frac{\sqrt{I_S}}{2V_t} \tag{3.3}$$

and hence, the specific current can be extracted from the slope of the square-root of the drain current plotted as a function of the source voltage (Figure 3.3). As, in practice, this slope is not constant, the specific current is calculated from the maximal slope as

$$I_S = [2V_t(d\sqrt{I_D}/dV_S)_{MAX}]^2. \tag{3.4}$$

Accordingly, the first simulation step consists in simulations of the drain current as a function of the source voltage for all transistor geometries. Strong inversion and saturation are ensured by setting $V_D = V_G = V_{DD}$ for an NMOS transistor and $V_D = V_G = V_{SS}$ for a PMOS transistor. To extract the temperature parameters, the same simulation set is repeated for the WL transistor for the

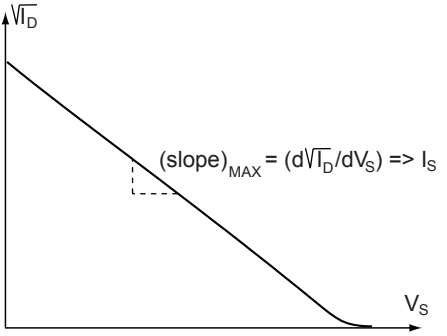


Figure 3.3 Shape of the curve of the square-root of the drain current as a function of source voltage for a WL NMOS transistor operating in strong inversion

temperatures of interest (-40°C, -20°C, 80°C, and 120°C). The specific current in each case is calculated using the equation (3.4).

3.4.3 Extraction of parameters *VTO, GAMMA* and *PHI*

The intrinsic parameters *VTO, GAMMA* and *PHI* are determined by fitting the simulated curves $V_P = f(V_G)$ for the WL transistor to the simplified equations

$$V_G' = V_G - VTO + PHI + GAMMA \cdot \sqrt{PHI} \tag{3.5}$$

and

$$V_P = V_G' - PHI - GAMMA\left(\sqrt{V_G' + \left(\frac{GAMMA}{2}\right)^2} - \frac{GAMMA}{2}\right) \tag{3.6}$$

for $V_G' > 0$, and $V_P = -PHI$ for $V_G' \le 0$.

The $V_P = f(V_G)$ curves are simulated using the scheme, proposed by M. Bucher [8], which is shown in Figure 3.4. According to the expression (2.31), when a transistor operates in saturation region and $V_P = V_S$, it follows that

$$I_D = I_S \cdot [ln(2)]^2 \approx \frac{I_S}{2} \quad . \tag{3.7}$$

Hence, if a transistor is polarized with half of the specific current, and the gate voltage is varied from the positive to the negative supply voltage, then the simulated source voltage V_S is equal to the pinch-off voltage V_P.

3.4.4 Extraction of parameters *LETA, WETA, and LK, Q0*

The parameter *LETA* is determined using the simulated $V_P = f(V_G)$ curves for WS transistors, while the parameter *WETA* is extracted from $V_P = f(V_G)$

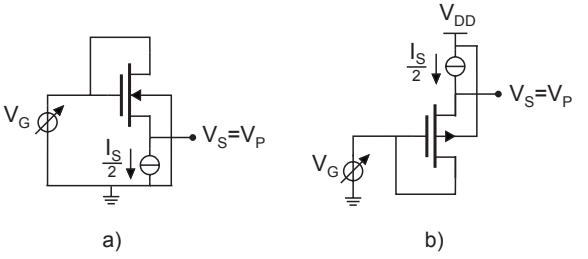


Figure 3.4 Scheme for $V_P = f(V_G)$ simulation for a) NMOS and b) PMOS transistors

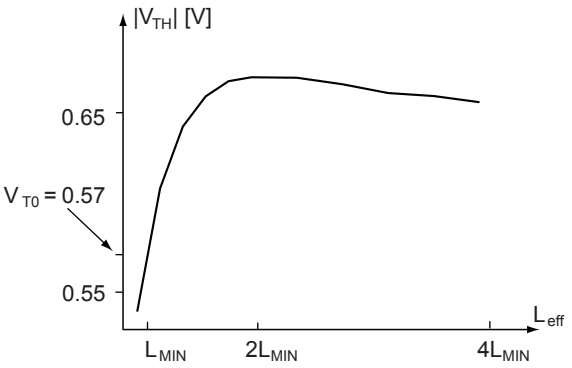


Figure 3.5 $V_{TH} = f(L_{eff})$ curve extracted for PMOS WS transistors (0.5 μm CMOS technology conversion example, $V_{T0,p}$ = 0.57 V)

curves for NL transistors. In this conversion step, the complete $V_P = f(V_G)$ EKV model equations are used for the curve fitting.

The parameters *LK, Q0* are used to model the reverse short-channel effect through the voltage difference $\Delta V_{RSCE} = f(LK, Q0)$. Thus, the equivalent threshold voltage value of each WS transistor is determined as the intersection of the appropriate $V_P = f(V_G)$ curve and the x-axis. Then, the curve $V_{TH} = f(L_{eff})$, whose shape is illustrated in Figure 3.5, is plotted and fitted using

$$V_{TH} = V_{T0} + \Delta V_{RSCE} + \gamma' \sqrt{V_S'} - GAMMA\sqrt{PHI} \tag{3.8}$$

where $V_{S',D'} = f(V_{S,D}, PHI)$, and $\gamma' = f(V_{S'}, V_{D'}, LETA, WETA)$.

3.4.5 Extraction of parameters *KP, E0*

Using the simplified drain current equation, the parameters *KP, E0* are extracted by fitting the curves $I_D = f(V_G)$ for a WL transistor operating in the linear region (V_{DS} = 50 mV) for different bulk voltages V_B. For an NMOS transistor, the V_B values are chosen in the range V_{SS} - $0.5(V_{DD}-V_{SS})$ to V_{SS}. For a PMOS transistor, the V_B values are chosen in the range V_{DD} to V_{DD} + $0.5(V_{DD}-V_{SS})$. Figure 3.6 illustrates the results obtained after fitting the simulated curves with the simplified EKV model on a 0.5 μm CMOS technology conversion example [12], [13].

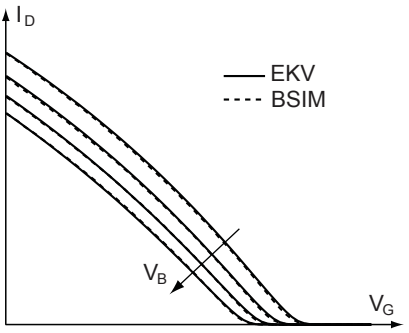


Figure 3.6 $I_D = f(V_G)$ simulated (BSIM) and fitted (EKV) characteristics for a WL PMOS transistor operating in the linear region (0.5 μm CMOS technology conversion example, KP_p = 45 μV/A², $E0_p$ = 43.3 V/m)

3.4.6 Extraction of parameters *RSH, DL*

The initial value of the sheet resistance RSH_{init} is determined using the method of Ghibaudo, described in detail in [14]. The RSH_{init} is calculated from the factor of the mobility reduction that is determined graphically from $I_D = f(V_G)$ curves simulated for WS transistors operating in the linear region. Then, fitting the same set of simulated curves for WS transistors, the parameters RSH and DL are extracted. The EKV model equations are still simplified in this step and include only effects modeled with the parameters extracted in the previous steps.

3.4.7 Extraction of parameters *UCRIT, LAMBDA*

The parameters *UCRIT, LAMBDA* are extracted using the complete set of EKV model equations and the simulated characteristics for $I_D = f(V_D)$ and $I_D = f(V_G)$. In order to determine the appropriate gate voltage value in strong, moderate and weak inversion, the WS transistors are simulated as diode-connected and polarized by $100I_S$, I_S and $0.01I_S$, respectively. However, due to the differences reported between the two models, only the strong inversion region is valid for the parameter extraction. Therefore, $I_D = f(V_D)$ curves are simulated for WS transistors in strong inversion, and the parameters are extracted using a different strategy compared to the previous steps (where each pair of model parameters is obtained by direct fitting of the appropriate set of curves).

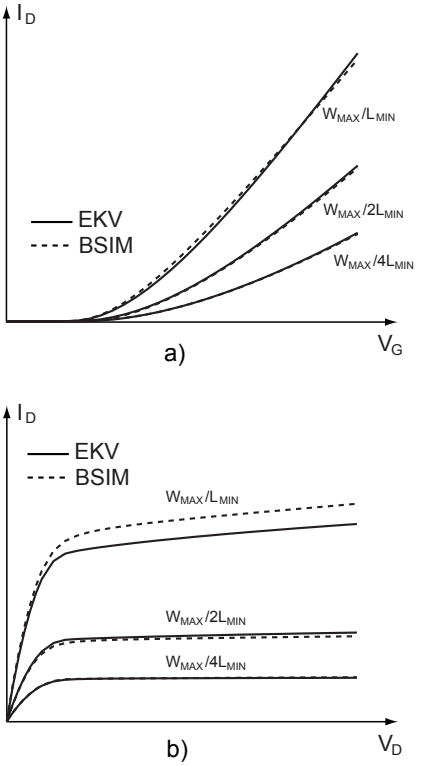


Figure 3.7 a) $I_D = f(V_G)$ in the saturation region and b) $I_D = f(V_D)$ in strong inversion operating region simulated (BSIM) and fitted (EKV) characteristics for WS NMOS transistors (0.5 μm CMOS technology conversion example, $LAMBDA_n$ = 0.27, $UCRIT_n$ = 5.25 MV/m)

First, the initial values of *UCRIT*, *LAMBDA* are estimated by fitting $I_D = f(V_D)$ curves simulated for WS transistors in strong inversion. The parameter *LAMBDA* is fixed to its initial value, and the parameter *UCRIT* is extracted by fitting $I_D = f(V_G)$ curves simulated for WS transistors operating in saturation. The *UCRIT* value is then fixed to this extracted value, and the parameter *LAMDA* is extracted by fitting once more $I_D = f(V_D)$ curves simulated for WS transistors in strong inversion. The final value of the parameter *LAMBDA* is evaluated giving the highest weight to the fitting result obtained in the case of the shortest transistor, i.e. W_{MAX}/L_{MIN}. Figure 3.7 depicts the results obtained after a fitting of $I_D = f(V_D)$ and $I_D = f(V_G)$ curves simulated with the BSIM model by the complete EKV model on a 0.5 μm CMOS technology conversion example.

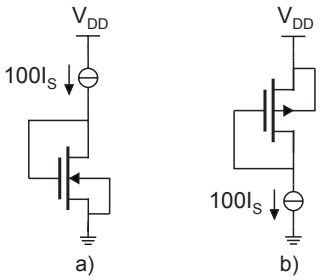


Figure 3.8 Scheme for the simulation of equivalent output noise current for WL a) NMOS and b) PMOS transistors operating in strong inversion

3.4.8 Extraction of parameters *IBA, IBB, IBN*

The parameters that are modeling the substrate current are extracted from $I_B = f(V_G)$ curves simulated for the WL transistor operating in saturation.

3.4.9 Extraction of parameters *TCV, BEX,* and *UCEX*

The temperature parameters are extracted by fitting the simulated curves for temperatures -40°C, -20°C, 80°C and 120°C in the following order:

- *TCV* (related to *VTO*) from $V_P = f(V_G)$ for a WL transistor,
- *BEX* (related to *KP*) from $I_D = f(V_G)$ for a WL transistor operating in the linear region,
- and *UCEX* (related to *UCRIT*) from $I_D = f(V_G)$ for WS transistors operating in saturation.

3.4.10 Extraction of parameters *KF, AF*

The noise parameters are extracted from the equivalent output noise current spectral density for WL transistors operating in strong inversion. The transistors are connected and polarized as presented in Figure 3.8, and the AC noise simulations are performed in this case.

3.5 Conversion results

After all conversion steps have been accomplished successfully, the BSIM2EKV program generates the EKV model library file and the conversion report text file. The conversion report file consists of the description of the conversion procedure, i.e. the details on each simulation and fitting step and the rms errors calculated for each set of fitted curves. The rms error is calculated as is defined in [15]. At the same time, the testbench simulation menu becomes

active in the program window (Figure 3.9). It allows the user to compare the different transistor characteristics simulated with the BSIM library file and the extracted EKV library file. In this way, the user can inspect visually the differences between the two models and estimate the quality of the conversion results. The available simulation sets are listed in Table 3.4.

Table 3.4 Testbench simulation options

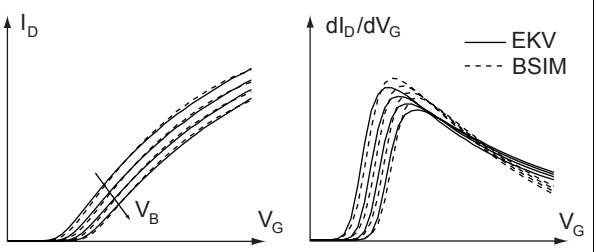
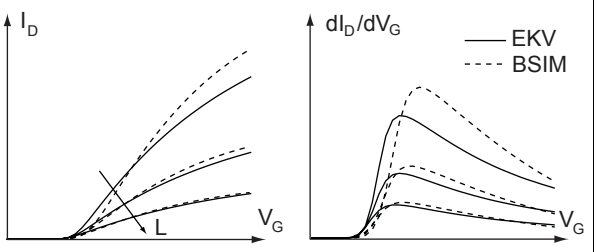
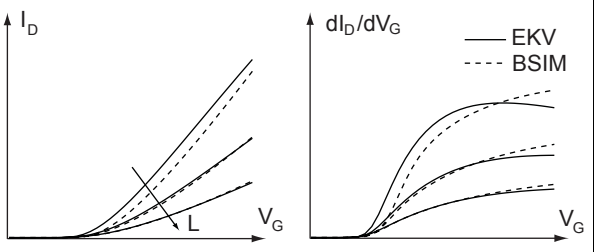
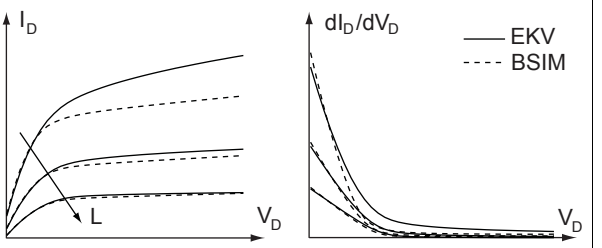
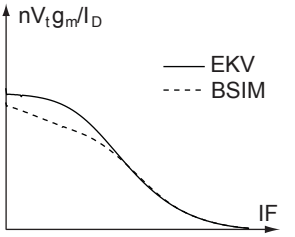
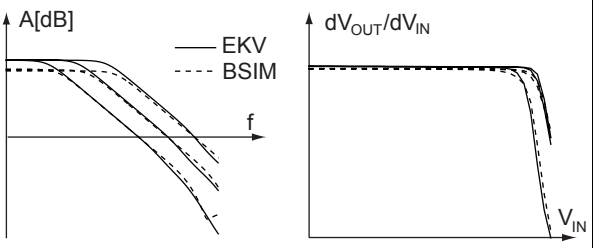
$I_D = f(V_G)$ WL transistor, linear region, different V_B	
$I_D = f(V_G)$ WL, WS and NL transistors, linear region	
$I_D = f(V_G)$ WL, WS and NL transistors, saturation region	

Table 3.4 (continued) Testbench simulation options

$I_D = f(V_D)$ WL, WS transistors, strong/ moderate/weak inversion	
normalized g_m/I_{Dsat} ratio	
simple circuit example: **OTA** (with differential pair in strong/ moderate/ weak inversion) - AC transfer function - DC input range	

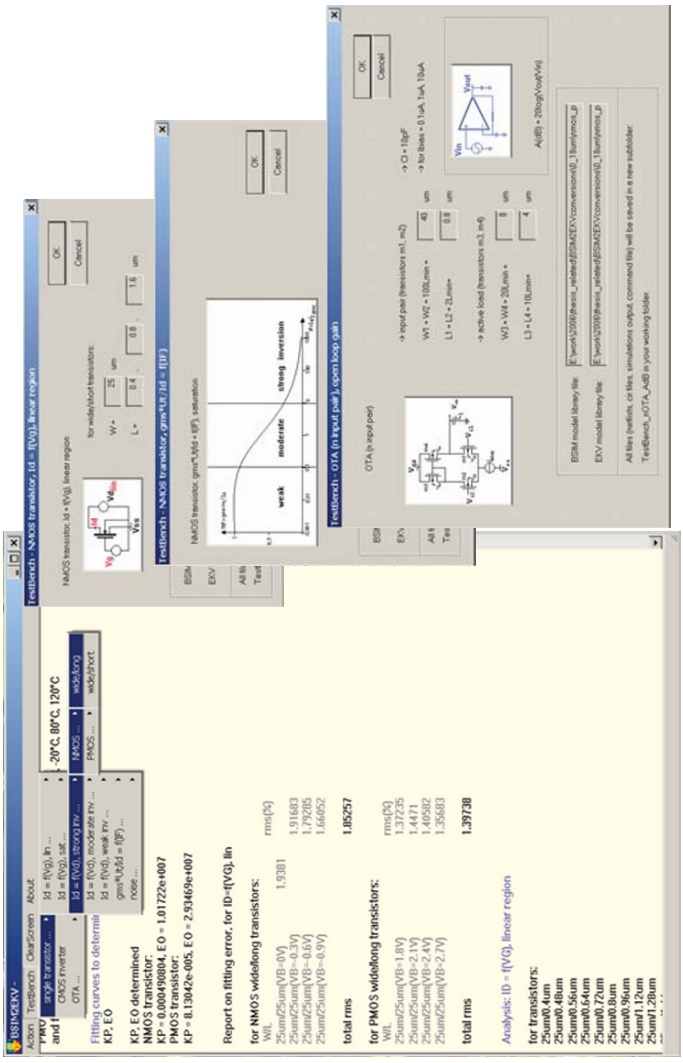


Figure 3.9 Snapshot of BSIM2EKV program windows

3.5.1 The 0.18 μm CMOS technology conversion example

The results of the BSIM to EKV model parameters conversion are illustrated here for an example of a 0.18 μm CMOS technology. The conversion time with an Intel Pentium 4 CPU at 2.4 GHz is about 5 minutes. The rms errors are given in Table 3.5. The comparison of transistor characteristics in the linear region and in the saturation, as well as in the strong inversion, are presented in Figures 3.10, 3.11, and 3.12, respectively.

Table 3.5 0.18 μm CMOS technology conversion example: rms error

compared characteristics	rms error [%] NMOS	rms error [%] PMOS
$V_P = f(V_G)$, **WL**	1.63	0.38
$I_D = f(V_G)$, **lin, different** V_B, **WL**	1.85	1.40
$V_P = f(V_G)$, **lin, WS**	5.61	2.14
$V_P = f(V_G)$, **sat, WS**	4.14	1.54
$V_P = f(V_G)$, **strong, WS**	7.48	2.12

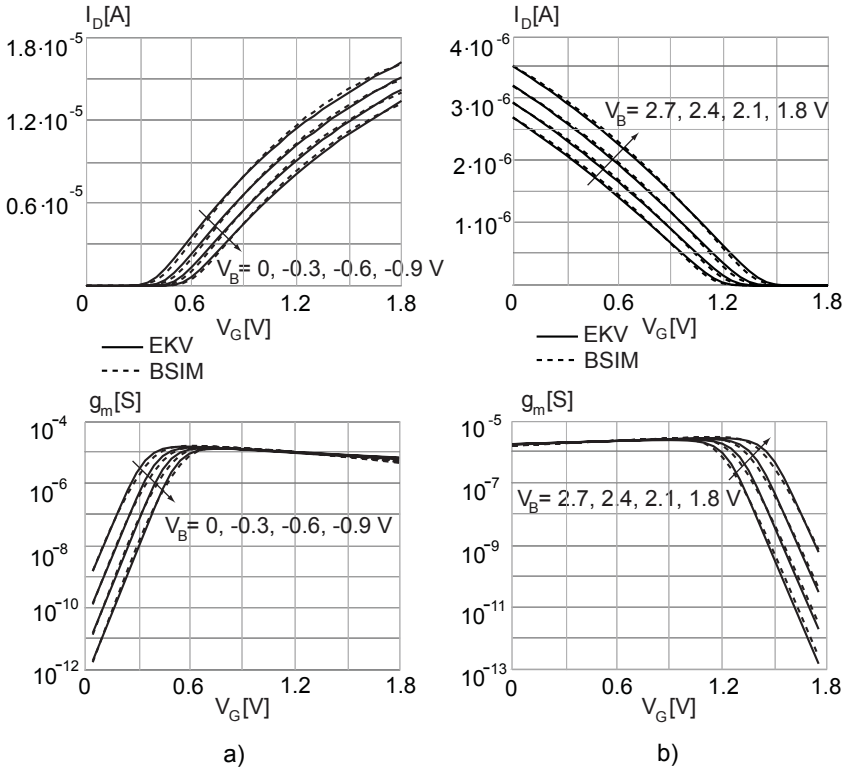


Figure 3.10 a) NMOS and b) PMOS $I_D = f(V_G)$ characteristics in the linear region for different V_B, simulated with BSIM and extracted EKV model parameters in a 0.18 μm CMOS technology conversion example

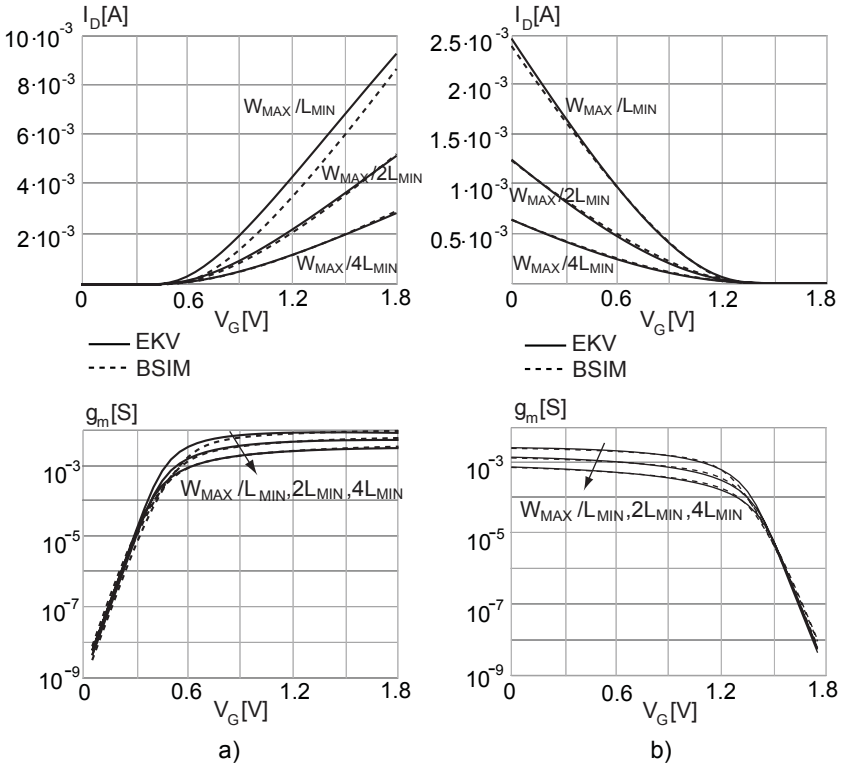


Figure 3.11 a) NMOS and b) PMOS $I_D = f(V_G)$ characteristics in the saturation region, simulated with BSIM and extracted EKV model parameters in a 0.18 μm CMOS technology conversion example

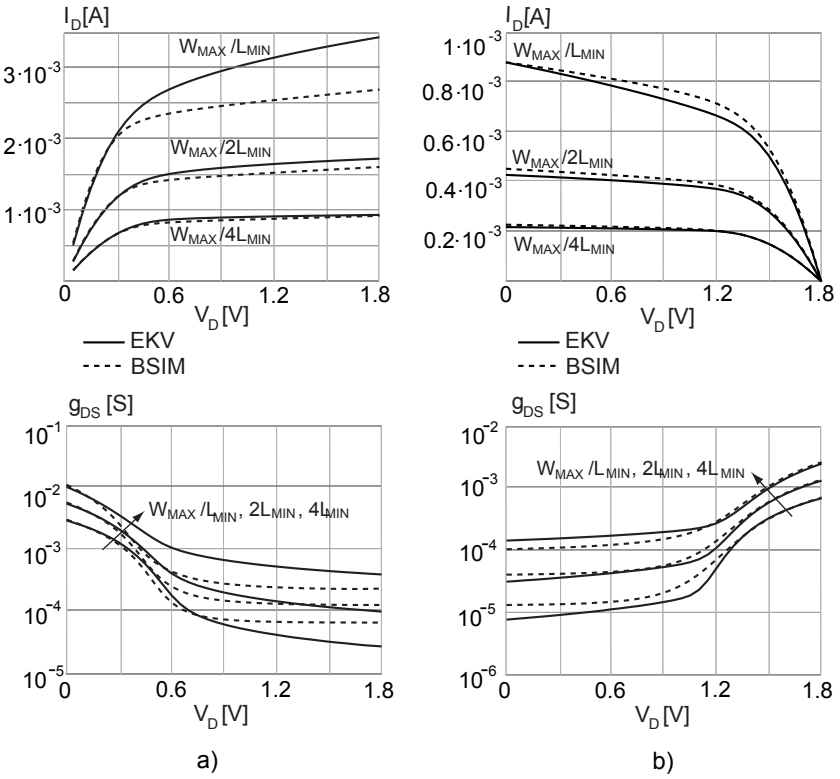


Figure 3.12 a) NMOS and b) PMOS $I_D = f(V_D)$ characteristics in strong inversion, simulated with BSIM and extracted EKV model parameters in a 0.18 μm CMOS technology conversion example

3.5.2 Results interpretation for analog design

The comparison of transistor characteristics simulated with the BSIM and the extracted EKV model parameters is important in order to identify the regions where special care must be taken because of significant differences between the results. A key point is that the differences do not occur because of problems of the fitting method, but as a result of the differences in the two modeling concepts. The best illustration is given in Figure 3.13, presenting the normalized g_m/I_{Dsat} ratio.

As stated previously, the fundamental difference appears in weak and moderate inversion. Hence, another testbench is available in order to estimate what are the gain differences for a single-stage amplifier structure. Figure 3.14 depicts the gain differences obtained in a 0.18 μm CMOS technology conversion example. Since the final design must be confirmed by the BSIM model simulation, this means that the appropriate margin must be added or subtracted when the analog blocks are designed either using PAD or using hand-calculation approximations.

In the similar way, all other comparison curves can be used to investigate the region of operation of interest. For example, if an analog switch is designed, then the simulations in the linear region may be very useful. Or, if an active load or a current mirror is designed, the comparison of the output conductances in weak and moderate inversion is mandatory. As a result, each analog structure can be sized using the advantages of the EKV model, and after taking into account the security margin, the result can be confirmed with the design kit provided by the foundry (i.e. the BSIM model library file).

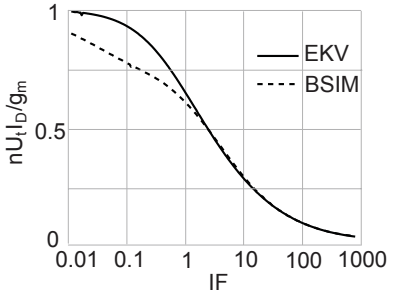


Figure 3.13 Normalized g_m/I_{Dsat} ratio simulated with BSIM and with the extracted EKV model parameters for a 0.18 μm CMOS technology conversion example

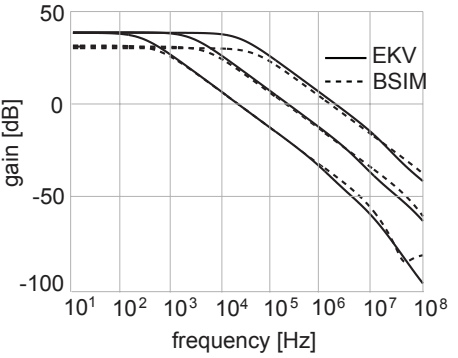


Figure 3.14 OTA transfer function with a differential pair operating in weak/ moderate/strong inversion simulated with BSIM and the extracted EKV model parameters for a 0.18 μm CMOS technology conversion example

3.6 Conclusion

BSIM2EKV program is a stand-alone application that enables to convert automatically the parameters of the BSIMv3.3 model to the parameters of the EKVv2.6 model. The conversion algorithm is developed in such a way as to minimize the error propagation and to respect the fundamental model differences, and no optimization loops are necessary. The BSIM2EKV has been successfully tested on different CMOS technologies. The conversion is usually performed in approximately 5 to 10 minutes with an average rms error within 5%.

3.7 Download

BSIM2EKV program is developed in the frame of the Ph.D. project that was performed at Electronics Labs, EPFL, Lausanne, Switzerland. The BSIM2EKV trial version and supporting documentation can be downloaded from the Electronics Labs web site: **http://analog.epfl.ch**.

Bibliography

[1] D. Stefanović, M. Kayal, M. Pastre, V. Litovski, "Procedural analog design (PAD) tool", Fourth International Symposium on Quality Electronic Design, pp. 313-318, March 2003

[2] D. Stefanović, M. Kayal, M. Pastre,"PAD: A new interactive knowledge-based analog design approach", Analog Integrated Circuits and Signal Processing Journal, Volume 42, pp. 291-299, March 2005

[3] D. Foty, M. Bucher, "Re-interpreting the MOS transistor via the inversion coefficient and the continuum of g_{ms}/I_d", International Conference on Electronics, Circuits and Systems, Volume 3, pp. 1179-1182, September 2002

[4] M. Bucher, D. Kazazis, F. Krummenacher, D. Binkley, D. Foty, Y. Papananos, "Analysis of transconductances at all levels of inversion in deep submicron CMOS", International Conference on Electronics, Circuits and Systems, Volume 3, pp. 1183-1186, September 2002

[5] F. Silveira, D. Flandre, P. G. A. Jespers, "A gm/ID based methodology for the design of CMOS analog circuits and its application to the synthesis of a silicon-on-insulator micropower OTA", IEEE Journal of Solid-State Circuits, Volume 31, pp. 1314-1319, September 1996

[6] D. Flandre, A. Viviani, J.-P. Eggermont, B. Gentinne, P. G. A. Jespers, "Improved synthesis of gain-boosted regulated-cascode CMOS stages using symbolic analysis and gm/ID methodology", IEEE Journal of Solid-State Circuits, Volume 32, pp. 1006-1012, July 1997

[7] D. Binkley, M. Bucher, D. Foty, "Design-oriented characterization of CMOS over the continuum of inversion level and channel length", IEEE International Conference on Electronics, Circuits and Systems, pp. 161-164, December 2000

[8] M. Bucher, C. Lallement, C. C. Enz, "An efficient parameter extraction methodology for the EKV MOST model", IEEE International Conference on Microelectronic Test Structures, pp. 145-150, March 1996

[9] M. Bucher, "Analytical MOS Transistor Modelling for Analog Circuit Simulation", PhD thesis N°2114, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 1999

[10] P. Bendix, “Detailed comparison of the SP2001, EKV, and BSIM3 models”, Technical Proceedings of the 2002 International Conference on Modelling and Simulation of Microsystems, Nanotech 2002, Volume 1, pp. 649-652

[11] S. C. Terry, J. M. Rochelle, D. M. Binkley, B. J. Blalock, D. P. Foty, M. Bucher, “Comparison of a BSIM3V3 and EKV MOSFET model for a 0.5 µm CMOS process and implications for analog circuit design”, IEEE Transactions on Nuclear Science, Volume 50, August 2003

[12] D. Stefanović, F. Krummenacher, M. Pastre, M. Kayal, “BSIM2EKV: un outil pour la conversion automatique des paramètres du modèle BSIM aux paramètres du modèle EKV”, TAISA’04 5ème Colloque sur le Traitement Analogique de l’Information, du Signal et ses Applications, pp. 85-88, Septembre 2004

[13] D. Stefanović, F. Krummenacher, M. Paste, M. Kayal, “BSIM2EKV: BSIM3.3 to EKV 2.6 model library file automatic conversion”, EKV Users' Meeting/Workshop, November 2004

[14] G. Ghibaudo, “New method for the extraction of MOSFET parameters”, Electronics Letters, Volume 24, pp. 543, 1998

[15] HP IC-CAP Modelling Software, “User’s Guide”, chapter 7, Hewlett-Packard, 1997

CHAPTER 4

Basic analog structures

A library including the most commonly used basic analog building blocks is defined. Then, circuit partitioning and derivation of the specifications from the circuit-level design requirements are discussed. On this foundation, the basic analog structures are classified as transconductance, load and bias structures. In addition, it is shown that the optimization of circuit performances is possible only at this level. Finally, the recipes for the design of basic analog structures based on the inversion-level methodology are presented.

4.1 Introduction

Structured analog design is based on the concept that each analog cell can be divided into basic analog structures. A basic analog structure is the smallest analog building block consisting of one or several transistors. It can be described by a set of design parameters that affect circuit-level performances. From this point of view, the design procedure becomes the same for both simple and extremely complicated circuits. It consists in:

- analog cell partitioning into basic analog structures,
- derivation of the specifications for each basic analog structure from the circuit-level specifications, and
- design of these basic analog structures in a specific sequence.

This chapter is focused on the basic analog structures, discussing the circuit partitioning, the derivation of design specifications, and the design of basic analog structures at the transistor level. Then, the next chapter shows several design scenarios in which the design sequence is topology dependent, yet still based on the same procedural approach for any circuit complexity.

4.2 Basic analog structures library

A basic analog structure is defined as a set of one or several transistors connected in a specific way to realize voltage-to-current conversion, current-

to-voltage conversion, or both [1]. Such a structure obviously demands current or voltage bias, which is again realized by another basic analog structure. Several basic analog structures connected together represent a complex analog structure that is often called an analog cell. An analog cell realizes an analog function such as: voltage or current amplifier, voltage follower, comparator, multiplier.

Table 4.1 shows the basic analog structures that are most commonly used as building blocks of the analog cells. It is important to note that a basic analog structure is not any combination of transistors. In fact, there is a finite number of basic analog structures and they create an analog library that can be used for the design of a large number of analog circuits.

Table 4.1 Basic analog structures library

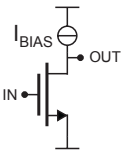
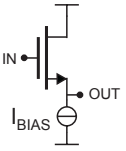
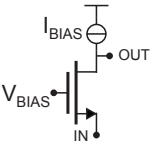
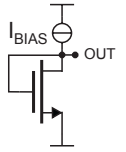
name	schematic view	basic features
common source		• voltage to current conversion
common drain		• voltage to current conversion • the simplest voltage follower
common gate		• voltage to current conversion • small input impedance
diode-connected transistor		• current to voltage conversion

Table 4.1 (continued) Basic analog structures library

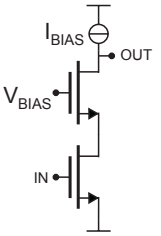
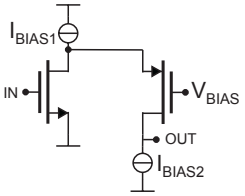
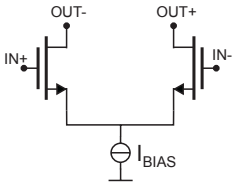
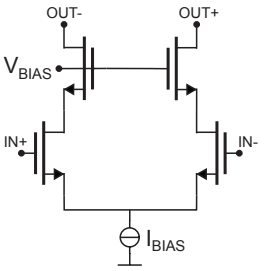
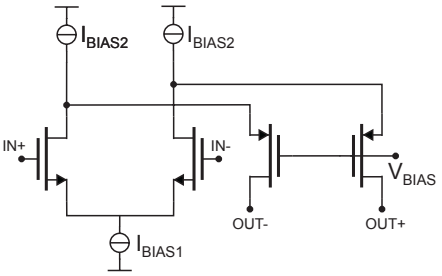
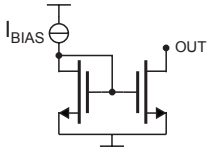
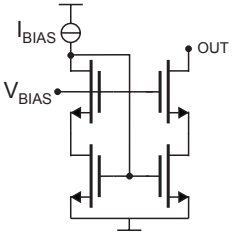
name	schematic view	basic features
cascode		• voltage to current conversion • common source + common gate stage • high output impedance
folded cascode		• voltage to current conversion • a cascode stage variant
differential pair		• differential input voltage to current conversion
cascoded differential pair		• differential input voltage to current conversion • differential pair variant

Table 4.1 (continued) Basic analog structures library

name	schematic view	basic features
folded cascode differential pair		• differential input voltage to current conversion • differential pair variant
current mirror		• current mirroring • current multiplication or division
cascode current mirror		• precise current mirroring • current multiplication or division • high output impedance

4.2.1 Transistor in a design environment

Closer inspection shows that a transistor in a basic analog structure can be connected in one of the following ways:

- source connected to the DC voltage supply, the input at the gate, and the output at the drain terminal (i.e. common source),
- drain connected to the DC voltage supply, the input at the gate, and output at the source terminal (i.e. common drain),
- gate connected to the DC voltage bias, the input at the source, and the output at the drain terminal (i.e. common gate),

- gate and drain connected together (i.e. transistor connected like a diode),
- transistor in a differential pair.

This implies that besides the finite number of different basic analog structures, there is also a finite number of transistor design situations. Hence, the design of a basic analog structure represents a special case of transistor-level design, where each transistor works in the given environment and only few transistor design parameters are dominant.

4.3 Structured design approach

Structured design will be demonstrated in this section on the example of an operational amplifier. The same approach can be used for the design of any other complex analog structure.

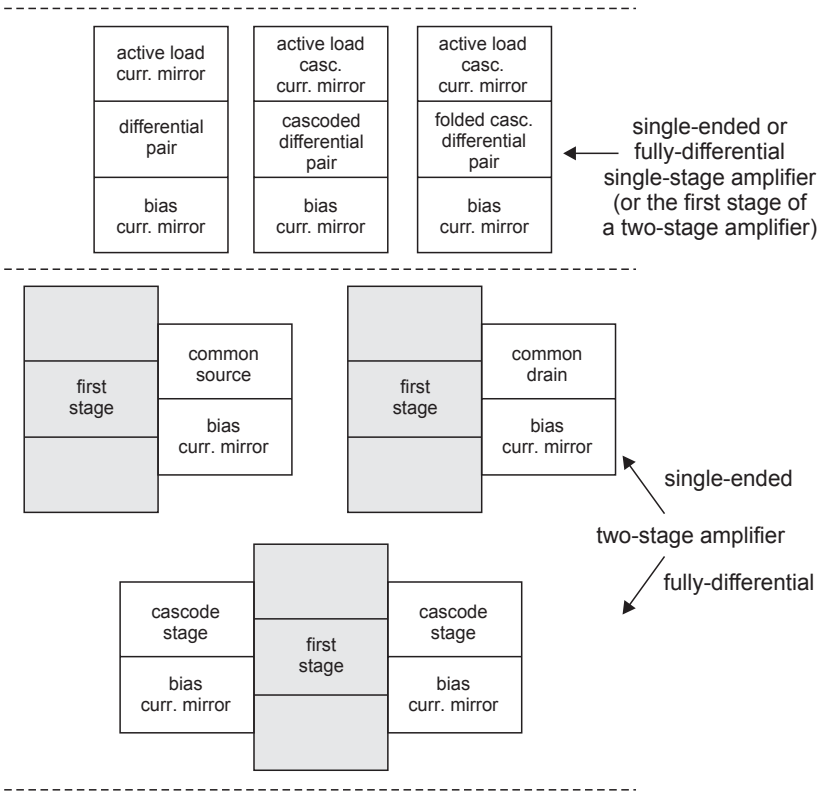


Figure 4.1 Partitioning of analog amplifiers into basic analog structures

4.3.1 Circuit partitioning

The most common operational transconductance amplifier and operational amplifier topologies implemented as either a single-ended or a fully-differential topology and partitioned into basic building blocks are shown in Figure 4.1. The operational transconductance amplifiers, realized as single-stage amplifiers, are shown in the following order: simple OTA, telescopic OTA, and folded cascode OTA. A two-stage operational amplifiers consist of an OTA in the first stage followed by a common source, a common drain or a cascode as illustrated in the same figure.

Classification of basic analog structures - Each stage of the analog amplifier consists of a transconductance (g_m) structure that converts the input voltage to the output current, followed by a load structure that converts the output current to the output voltage. The transconductance structure requires a current bias, whereas any cascoded structure requires a voltage bias. Therefore, all basic analog structures can be classified in the following way:

- **transconductance structures**: common source, common drain, cascode, differential pair and its cascoded variants,
- **load structures**: simple and cascode current mirrors,
- and **bias structures**: simple and cascode current mirrors, diode-connected transistor.

A key point is that all transconductance structures have the same set of design parameters that affect the circuit performances. The same holds for the bias stages and the load stages. This facilitates the derivation of the specifications and also simplifies the design procedure.

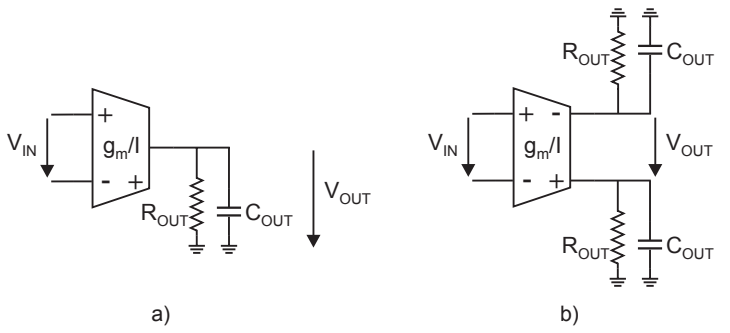


Figure 4.2 Behavioral model of an analog amplifier with a) single-ended and b) fully-differential outputs

Behavioral model of an analog amplifier - Since each transconductance structure requires a current bias structure, they can be seen together as a g_m-cell providing the required transconductance and the output current. This g_m-cell followed by a load gives the simplest behavioral model of an analog amplifier (Figure 4.2). This behavioral model is useful for system-level simulations, as well as for the derivation of the amplifier specifications from the system-level specifications. It can be more complex, depending on the system and the application.

4.3.2 Derivation of the specifications

Figure 4.3 summarizes the structured analog design procedure. First, the analog circuit is partitioned into transconductance, load and bias structures. Then, the specifications are derived for each analog structure from the circuit-level specifications. At this point, it is possible to check technology limits, verify circuit-level specifications using the behavioral model or make topology changes. Since each analog structure represents a set of transistors in a design environment, the final step is to choose the appropriate design recipe for each transistor with regard to the basic analog structures' specifications. Each transistor can be then sized as proposed in chapter 2 using the inversion level methodology. An example of the derivation of the specifications for the analog amplifiers illustrated in Figure 4.1 is given in Table 4.2.

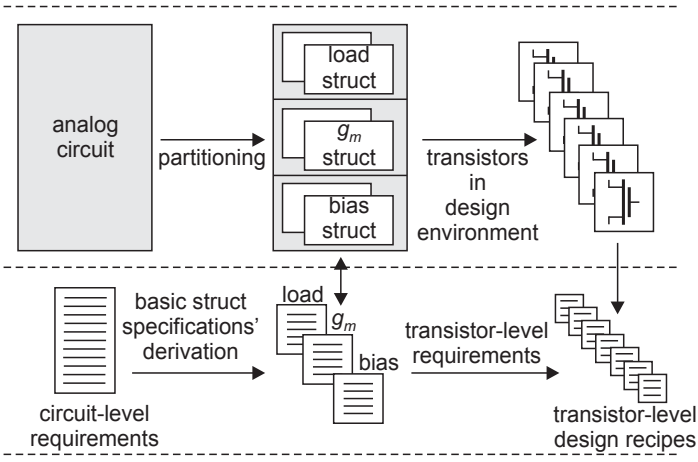


Figure 4.3 Structured analog design procedure and derivation of the specifications

Table 4.2 Derivation of the specifications, example of an analog amplifier

circuit-level specification	approximative expression	basic analog structure specification
gain	$g_m R_{OUT}$	• the transconductance of transconduct. struct. • output resistance of load struct.
gain bandwidth	$\frac{g_m}{2\pi C_L}$	• the transconductance of transconduct. struct. • maximal output capacitance of load struct.
phase margin	$180° - atan\left(\frac{f_{GBW}}{f_{dp}}\right) - \Sigma atan\left(\frac{f_{GBW}}{f_{ndp}}\right)$	• the ΣC in nodes in the signal gain path
common-mode input range (CMIR)	$V_{sup} - V_{GS} - V_{DSsat}$	• the saturation voltage of bias struct. • the saturation voltage and the V_{GS} voltage headroom of transconduc. struct.
differential-mode input range (DMIR)	$\sim V_{GS}$	• the V_{GS} voltage headroom of transconduc. struct.
output swing	$V_{sup} - \Sigma V_{DSsat}$	• the saturation voltage of load struct. (and, in some cases, of transconduct. struct.)
static current consumption	ΣI	• all circuit branch currents

Table 4.2 (continued) Derivation of the specifications, example of an analog amplifier

circuit-level specification	approximative expression	basic analog structure specification
slew-rate	$\frac{I_{OUT}}{C_L}$	• the bias current of transconduct. struct. • the branch currents of current bias struct. • maximal output capacitance of load struct.
speed	• equivalent RC const. • current sourcing/ sinking capability	• the intrinsic capacitance and the transition frequency of all transistors • all circuit branch currents
input-referred equivalent noise	$v_{n,\,dp}^2 - \left(\frac{g_{m,\,mirr}}{g_m}\right)^2 \Sigma v_{n,\,mirr}^2$	• the noise contrib. of all transistors except cascodes • the ratio of the transconductances of transconduct. and load struct.
input-referred equivalent offset	$v_{m,\,dp}^2 - \left(\frac{g_{m,\,mirr}}{g_m}\right)^2 \Sigma v_{m,\,mirr}^2$	• the voltage mismatch contrib. of all transistors except cascodes • the ratio of the transconductances of transconduct. and load struct.
rejection ratios	• $\frac{A_D}{A_{DD}}, \frac{A_D}{A_{SS}}, \frac{A_D}{A_{CMi}}$ • $\frac{A_{CMFB}}{A_{DD}}, \frac{A_{CMFB}}{A_{SS}}, \frac{A_{CMFB}}{A_{CMi}}$	• the output resistance of bias struct. • the output conductance of mirror transistors of load struct.

Design parameters' trade-offs - Sometimes the design specifications can be difficult to achieve because they may demand incommensurate values of transistor design variables. When the transistor-level optimization strategies, described in chapter 2, cannot provide satisfactory results, the design optimization consists in finding a balance between the design specifications that can be relaxed and the design specifications that have to be respected at the level of basic analog structures. In fact, this becomes the only abstraction level where the design parameters may be traded off in order to provide the required circuit-level performances. It must be pointed out that the design optimization can be treated as a mathematical problem, (for example [2], [3]), in a limited number of cases, but this does not provide a general insight into the solution of the problem. An attempt to tune the transistor sizes in order to achieve the circuit performances results in a complicated and time-consuming procedure, and has no real analog design background. Thanks to the structured design methodology, the analog design optimization problem is approached in this book in a different way.

In the first place, since each basic analog structure has its own role in the circuit, only the parameters that influence the circuit-level behavior need to be taken into account. Second, according to the circuit-level design requirements, it is possible to determine the level of priority of the derived design specification, as well as the acceptable parameters' bounds. This results in a topology dependent design sequence, as will be presented in chapter 5, and hence all basic analog structures have to be designed in a specific order. In this way, the design specifications are derived not only from the circuit specifications, but also with regard to the design parameters of the previously designed blocks. Indeed, it becomes possible to find out which specification can be relaxed. Moreover, using the classification of analog blocks in transconductance, load and bias structures, and the behavioral model of the amplifier, the circuit-level specifications can be correctly interpreted. For example, if a certain value of the amplifier gain has to be achieved, the required values of the equivalent transconductance and the output resistance can be confirmed by the behavioral system-level simulations. Therefore, some important design decisions can be taken at the system level, without taking into consideration the transistor-level design.

Another important point is the possibility of checking the technology limits at the level of basic analog structures. For example, since the transconductance of the transconductance structure is derived from the gain bandwidth specification, and the output current is derived from the slew-rate specification, we can write

$$\frac{f_{GBW}}{SR} = \frac{g_m/(2\pi C_L)}{I_{OUT}/C_L} = \frac{g_m}{2\pi I_{OUT}} = \frac{g_m}{2\pi I_{Dsat}} = \frac{1}{2\pi}\cdot\left(\frac{g_m}{I_{Dsat}}\right). \tag{4.1}$$

Now, if this ratio is larger than the maximal (g_m/I_{Dsat}) ratio for the given technology, there in no design solution, and there is no need to look for the design solution at the transistor level.

Finally, when it is difficult (or impossible) to achieve a certain design requirement for the given technology, there is a solution, presented in chapter 7, that is again based on the structured design concept. It consists in replacing only the basic analog structure that affects the parameter in question by its more appropriate version. For example, if a high output resistance has to be achieved, the transconductance and the load structure have to be replaced by their cascoded versions. The rest of the circuit remains the same, as well as the design sequence.

Analog cell design steps are the central topic of the next chapter. However, before defining the analog cell design sequence, it is important to understand the characteristics of all basic analog structures, as well as being able to identify their design parameters. This is analyzed in the following sections of this chapter.

4.4 Transconductance structures

Single-input-single-output transconductance structures that convert an input voltage in an output current are: common source, common drain, common gate and cascode. On the other hand, differential pair and its cascoded variants convert a differential input voltage to a differential output current. Nevertheless, in both cases the important design parameters are the following:

- transconductance,
- output resistance,
- input range,
- output swing,
- noise contribution, and
- parasitic capacitances seen at the input and at the output.

In the case of differential structures the additional parameter is input offset, i.e. voltage mismatch contribution, since they consist of two or four transistors. In this section, the design parameters and the appropriate design cases leading to transistor-level design recipes are determined for all the aforementioned transconductance structures.

4.4.1 Common source

Common source requires a current bias that behaves as a load structure at the same time. This bias-load structure is often a simple or cascode current

source. Figure 4.4 depicts the common source stage without entering into the bias implementation details.

Design parameters - According to the small-signal equivalent scheme shown in Figure 4.5, the small-signal output current is determined by the transconductance of transistor M1 as

$$i_{OUT} = -g_{m1}v_{IN}. \tag{4.2}$$

Assuming that the output resistance R_L of the bias structure is larger than the output resistance of the common source $R_L \gg 1/g_{DS1}$, the voltage gain is approximated as the intrinsic gain

$$A = \frac{v_{OUT}}{v_{IN}} = -\frac{g_{m1}}{g_{DS1}}. \tag{4.3}$$

Obviously, the output conductance g_{DS1} becomes an important parameter, since it limits the total output resistance.

The output swing is limited on the one hand by the saturation voltage of transistor M1 V_{DSsat1}, and on the other hand by the voltage headroom required by the bias structure.

The parasitic capacitance seen at the input (Figure 4.6) is estimated as

$$C_{IN} \approx C_{GB1} + C_{GS1} + \frac{g_{m1}}{g_{DS1}} \cdot C_{GD1}, \tag{4.4}$$

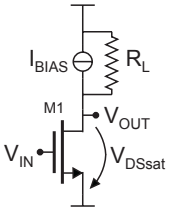


Figure 4.4 Common source with its bias/load structure

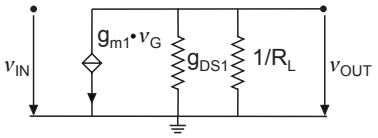


Figure 4.5 Common source small-signal equivalent scheme

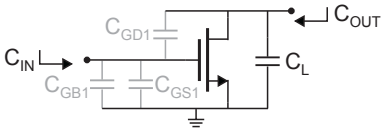


Figure 4.6 Common source parasitic capacitances

while the output capacitance is $\approx C_{GD1}$. The capacitance C_L is not taken into consideration since it is not an intrinsic part of this basic analog structure. In design practice, the input capacitance is more critical since it limits the transient response (speed) of the gain stage. Therefore, it is usually largely approximated by the transistor intrinsic gate-oxide capacitance $C_{ox}W_1L_1$, that becomes an important design parameter.

Hence, the common source design parameters are the following parameters of transistor M1:

- transconductance g_{m1},
- output conductance g_{DS1},
- saturation voltage V_{DSsat1}, and
- intrinsic gate-oxide capacitance.

Design cases - The following design situations are possible:

- output conductance + transconductance,
- output conductance + saturation voltage,
- transconductance + parasitic capacitances, and
- output conductance + saturation voltage + parasitic capacitances.

The design solutions are given in chapter 2 as transistor-level design recipes, and will not be further discussed here.

4.4.2 Common drain

Common drain with its bias is depicted in Figure 4.7, again without entering into the bias implementation details.

Design parameters - Figure 4.8 depicts the small-signal equivalent scheme in two cases:

- $S = B$, that is the source terminal is connected to the bulk terminal,
- $S \neq B$, that is the source terminal is not connected to the bulk terminal.

The small-signal output current is given by:

$$i_{OUT} = g_{m1}v_{IN}\,. \tag{4.5}$$

In the first case, there is no body effect and the voltage gain is

$$A = \frac{v_{OUT}}{v_{IN}} = \frac{g_{m1}}{g_{m1} + g_{DS1} + (1/R_L)} \approx 1, \tag{4.6}$$

whereas in the second case it becomes

$$A = \frac{v_{OUT}}{v_{IN}} = \frac{g_{m1}}{g_{mS1} + g_{DS1} + (1/R_L)} \approx \frac{g_{m1}}{g_{mS1}} \approx \frac{1}{n}. \tag{4.7}$$

As a result, the common drain stage behaves as a voltage follower, since its gain is less than or equal to unity.

According to Figure 4.9, both the input and the output parasitic capacitances are determined by the gate-source capacitance C_{GS1} that is largely approximated, as in the previous case, by $C_{ox}W_1L_1$.

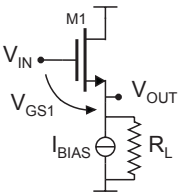


Figure 4.7 Common drain with its bias

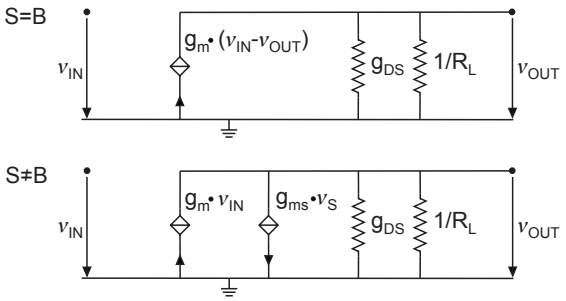


Figure 4.8 Common-drain small-signal equivalent scheme

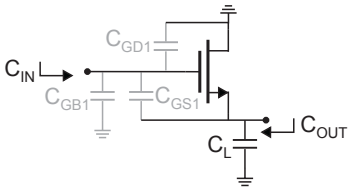


Figure 4.9 Common drain parasitic capacitances

The amplitude of the output voltage depends on the DC level of the signal at the input and the inversion level of transistor M1. The DC output voltage level (Figure 4.7) is calculated with regard to the input as

$$V_{OUT} = V_{IN} - V_{GS1}, \tag{4.8}$$

and thus the output voltage range is limited by the required V_{GS1} voltage headroom as

$$(V_{OUT})_{MAX} = \frac{V_{IN} - V_{T0} - 2nV_t \cdot ln\left[exp\left(\sqrt{\frac{I_{Dsat}}{I_S}}\right) - 1\right]}{n}, \tag{4.9}$$

or in the case when the source terminal is connected to the bulk terminal as

$$(V_{OUT})_{MAX} = V_{IN} - V_{T0} - 2nV_t \cdot ln\left[exp\left(\sqrt{\frac{I_{Dsat}}{I_S}}\right) - 1\right]. \tag{4.10}$$

Therefore, it depends on the inversion operating region, since $\frac{I_{Dsat}}{I_S} = IF$.

Figure 4.10 shows the V_{GS} headroom part $\Delta V = 2V_t \cdot ln[exp(\sqrt{IF}) - 1]$ that is dependent on the inversion operating region as a function of the inversion factor. Similarly, as in the case of the saturation voltage (Figure 2.13), the smallest voltage headroom is achieved in weak and moderate inversion.

The design parameters are thus summarized as follows:

- transconductance g_{m1},
- saturation voltage V_{DSsat1},
- gate-source voltage headroom V_{GS1}, and
- intrinsic gate-oxide capacitance.

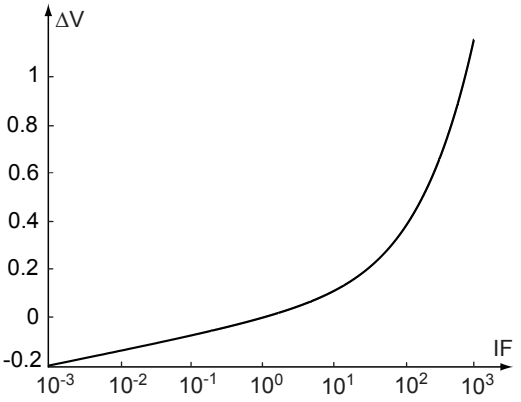


Figure 4.10 V_{GS} headroom portion dependent on the inversion operating region as a function of the inversion factor

Design cases - To ensure that the voltage gain is as close as possible to unity, the source and the bulk terminal are usually connected together, and the transistor length is set to L_{MIN}. Since the saturation voltage and the output range change in the same way with the inversion operating region, the only design case is in fact: saturation voltage + parasitic capacitances. The optimal solution is in the region $1 < IF < 10$, as shown in chapter 2.

4.4.3 Common gate

Common gate with its bias, which behaves as a load at the same time, is shown in Figure 4.11.

The small-signal equivalent scheme is depicted in Figure 4.12, whereas the equivalent scheme for the calculation of the output resistance is illustrated in Figure 4.13. The resistance R_S is the equivalent resistance seen in the source of the transistor M1 and it is determined by another basic analog structure.

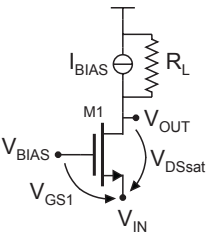


Figure 4.11 Common gate with its bias

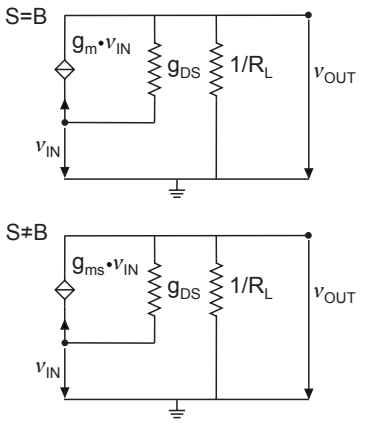


Figure 4.12 Common gate small-signal equivalent scheme

Design parameters - In the case when the source terminal is not connected to bulk, according to Figure 4.12 (case $S \neq B$), the small-signal output current is written as

$$i_{OUT} = g_{mS} v_{IN}, \tag{4.11}$$

the voltage gain is given by

$$A = \frac{v_{OUT}}{v_{IN}} = \frac{g_{mS1}}{g_{DS1} + (1/R_L)} \approx \frac{g_{mS1}}{g_{DS1}} \approx n\frac{g_{m1}}{g_{DS1}}, \tag{4.12}$$

and the output resistance is estimated, according to Figure 4.13 (case $S \neq B$), as

$$R_{OUT} = \frac{v_{OUT}}{i_{OUT}} \approx \frac{g_{mS1}}{g_{DS1}} \cdot R_S \approx n\frac{g_{m1}}{g_{DS1}} \cdot R_S. \tag{4.13}$$

In the case $S = B$, the small-signal output current is

$$i_{OUT} = g_m v_{IN} \tag{4.14}$$

the voltage gain is

$$A = \frac{v_{OUT}}{v_{IN}} = \frac{g_{m1}}{g_{DS1} + (1/R_L)} \approx \frac{g_{m1}}{g_{DS1}}. \tag{4.15}$$

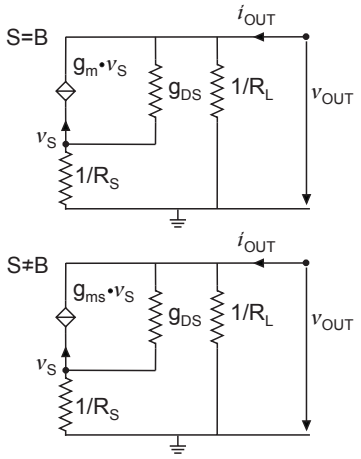


Figure 4.13 Common gate small-signal equivalent scheme for the output resistance calculation

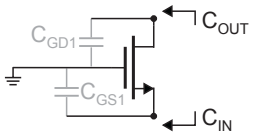


Figure 4.14 Common gate parasitic capacitances

and the output resistance is

$$R_{OUT} = \frac{v_{OUT}}{i_{OUT}} \approx \frac{g_{m1}}{g_{DS1}} \cdot R_S . \tag{4.16}$$

As a result, the design parameters to be taken into account are not only the transconductance, and the output conductance, but also the intrinsic gain of the transistor M1.

Concerning the speed and the frequency response, the input parasitic capacitance is equal to the gate-source capacitance C_{GS1}, whereas the output parasitic capacitance is the gate-drain capacitance C_{DS1}, as is indicated in Figure 4.14. Since the capacitance C_{GS1} is dominant, the intrinsic gate-oxide capacitance of the transistor M1 can be considered as a design parameter.

Similarly to the previous two stages, the output voltage swing is determined by the saturation voltage, as well as by the V_{GS1} voltage headroom.

Therefore, the common gate design parameters are:
- transconductance g_{m1},
- output conductance g_{DS1},
- intrinsic gain A_{i1},
- saturation voltage V_{DSsat1}, and
- intrinsic gate-oxide capacitance.

Design cases - The possible design situations are similar to the common source case, and they have been already discussed in chapter 2:
- output conductance + transconductance,
- output conductance + gain,
- transconductance + gain,
- gain + saturation voltage, and
- gain + saturation voltage + parasitic capacitances.

4.4.4 Cascode and folded cascode

The cascode stage acts as a common source stage followed by a common gate stage, as illustrated in Figure 4.15.

Design parameters - The dominant design parameters and the design cases are defined for both the transistor M1 and the "cascode" transistor M2 from the parameters that affect the circuit-level design (transconductance, output resistance, output swing, parasitic capacitances).

The small-signal cascode structure output current is

$$i_{OUT} = -g_{m1} v_{IN}, \tag{4.17}$$

whereas its output resistance is given by

$$R_{OUT} \approx \frac{1}{g_{DS1}} \cdot \frac{(n) \cdot g_{m2}}{g_{DS2}}. \tag{4.18}$$

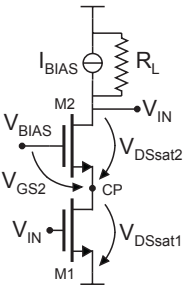


Figure 4.15 Cascode with its bias

Assuming again that the bias output resistance R_L is larger than the output resistance of the cascode, the voltage gain is approximated as

$$A \approx -\frac{g_{m1}}{g_{DS1}} \cdot \frac{(n) \cdot g_{m2}}{g_{DS2}}. \tag{4.19}$$

The slope factor n in equations (4.18) and (4.19) appears in the case when the source and the bulk terminal are not shorted.

Concerning the influence of the parasitic capacitances, the most critical is the capacitance seen in the cascode point CP (Figure 4.16) as will be shown in the next chapter. It is approximated as

$$C_{CP} = C_{GS2} + \frac{g_{m1}}{(n) \cdot g_{m2}} \cdot C_{GD1} \approx C_{GS2}. \tag{4.20}$$

The input parasitic capacitance is equal to gate-source capacitance C_{GS1}, since the influence of the Miller capacitance C_{GD1} can be neglected as the gain g_{m1}/g_{m2} from the input to the cascode point is small. The output parasitic capacitance corresponds to the gate-drain capacitance C_{GD2}.

To ensure maximal output swing, the bias voltage V_{BIAS} is determined so that the transistor M1 is kept at the limit of saturation, that is

$$V_{BIAS} = V_{DSsat1} + V_{GS2} + \Delta V, \tag{4.21}$$

where ΔV represents the security margin (either fixed at 50 mV or calculated to compensate worst/best case or temperature variations). It follows that this bias voltage depends on IF_1 and IF_2 since from the equations (4.8), (4.9) and (4.21) it is given by

$$V_{BIAS} = V_{T0} + (n) \cdot V_{DSsat1} + 2nV_t ln[exp(\sqrt{IF_2}) - 1]. \tag{4.22}$$

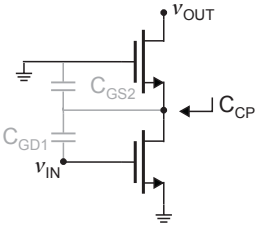


Figure 4.16 Parasitic capacitances in the cascode point

Consequently, the minimal required DC output voltage headroom becomes

$$(V_{OUT})_{MIN} = V_{DSsat1} + V_{DSsat2} + \Delta V, \tag{4.23}$$

and it depends on the inversion operating regions of transistors M1 and M2.

The design parameters of transistor M1 are thus:

- transconductance g_{m1},
- output conductance g_{DS1},
- saturation voltage V_{DSsat1}, and
- intrinsic gate-oxide capacitance.

On the other hand, the design parameters of transistor M2 are:

- transconductance g_{m2},
- output conductance g_{DS2},
- saturation voltage V_{DSsat2},
- intrinsic gain A_{i2}, and
- intrinsic gate-oxide capacitance.

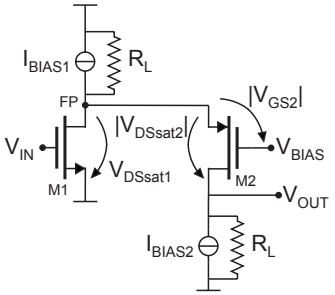


Figure 4.17 Folded cascode with its biases

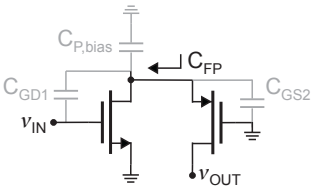


Figure 4.18 Parasitic capacitances in the folded point

Design cases - The possible design situations for transistor M1 are:
- output conductance + transconductance,
- transconductance + saturation voltage,

and they do require an optimization strategy.

Nevertheless, the design of transistor M2 usually demands a search for design trade-offs since the possible design cases are:
- gain + saturation voltage,
- gain + saturation voltage + parasitic capacitances,
- output conductance + saturation voltage + parasitic capacitances.

The cascode variant realized with a complementary cascode transistor is depicted in Figure 4.17. It requires an additional current bias I_{BIAS2}. Here, the DC voltage bias is calculated with regard to the voltage range required by the current bias I_{BIAS1}. The design parameters and the design cases are the same as for the cascode. (Figure 4.18 displays the parasitic capacitances in the folded point.)

4.4.5 Differential pair

A differential pair is depicted in Figure 4.19. As previously stated, it converts a differential input voltage $v_{IN,\,DM} = v_{IN+} - v_{IN-}$ to a differential output current $i_{OUT} = i_{OUT-} - i_{OUT+}$, and ideally should not be sensitive to any variation of the common-mode input voltage $v_{IN,\,CM} = 0.5 \cdot (v_{IN+} + v_{IN-})$.

Differential pair variants are depicted in Figure 4.20, as a) cascoded differential pair and b) folded cascoded differential pair. The cascode (folded cascode) pair transistors M3,4 impose the drain voltage of transistors M1,2, ensuring that they operate in the saturation region. In addition, the output resistance of these variant structures is equal to the output resistance of the cascode, calculated in subsection 4.4.4. This makes it possible to create high-gain amplifiers, if the load structure is also cascoded. Since the design of the cascode (folded cascode) transistors M3,4 is exactly the same as in the case of the cascode structure, it will not be further discussed here. We will focus on the design parameters of the differential pair transistors M1,2.

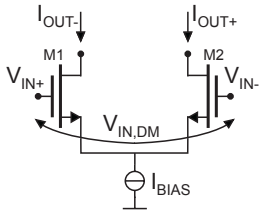


Figure 4.19 Differential pair

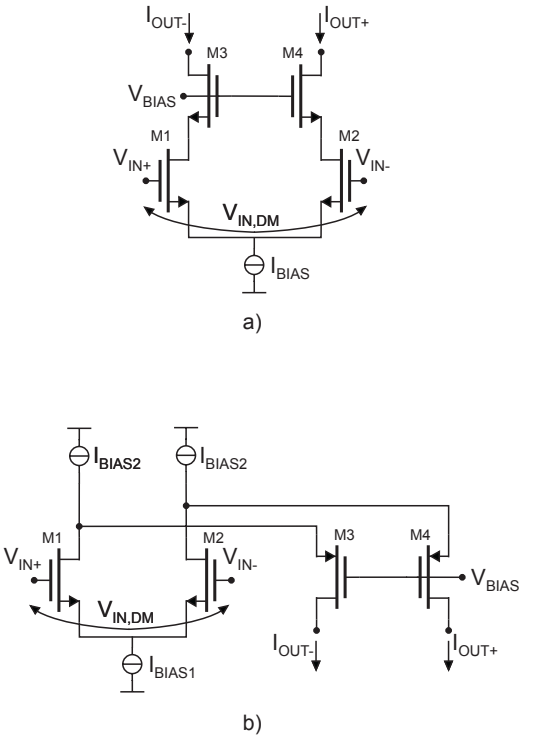


Figure 4.20 Differential pair variants: a) cascoded b) folded cascoded differential pair

Design parameters - The design parameters of a differential pair are:

- equivalent transconductance $g_{m,eq}$,
- saturation voltage of the differential pair V_{DPsat},
- differential input range,
- common-mode input range,
- input capacitance,
- equivalent noise, and
- equivalent offset.

On this basis, the transistor parameters and the design cases will be determined later.

Transfer characteristic of the differential pair is shown in Figure 4.21. The equivalent transconductance is defined as the first-order derivative of the output current

$$g_{m,\,eq} = \frac{d(I_{OUT})}{d(V_{IN,\,DM})}, \tag{4.24}$$

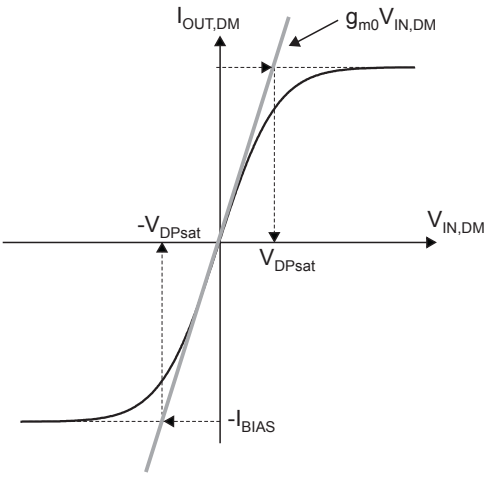


Figure 4.21 Transfer characteristic of the differential pair

and thus represents the slope of the transfer characteristics. The slope at the origin, i.e. for a small differential input voltage, is equal to the transconductance of the transistors M1,2 at the DC operating point

$$g_{m0} = \left.\frac{d(I_{OUT})}{d(V_{IN,\,DM})}\right|_{V_{IN,\,DM}\to 0} = g_{m1} = g_{m2}. \tag{4.25}$$

Hence, when a small differential signal is present at the input, the small-signal output current is given by

$$\begin{aligned} i_{OUT} &= i_{OUT-} - i_{OUT+} = g_{m1}\cdot\frac{v_{IN,\,DM}}{2} + g_{m2}\cdot\frac{v_{IN,\,DM}}{2} \\ &= g_{m0}\cdot v_{IN,\,DM}. \end{aligned} \tag{4.26}$$

On the other hand, if a large differential signal is present at the differential pair input, the whole tail current I_{BIAS} swings to the positive or the negative output terminal, depending on the sign of the input voltage. It is said that the differential pair is in "saturation".

The differential input range is defined as the differential input signal range for which the operating point of the differential pair remains in the linear region of the transfer characteristic. Consequently, the equivalent transconductance remains constant or within a certain error. When this parameter is given as a

design specification, the acceptable error must also be specified to make it possible to correctly evaluate the differential pair performance.

The parameter that is used as a measure of the differential input signal range for design purposes is called the saturation voltage of the differential pair V_{DPsat}. This is a rough approximation calculated as a ratio of the tail current and the equivalent transconductance in the operating point

$$V_{DPsat} = \frac{I_{BIAS}}{g_{m0}} = \frac{2 \cdot I_{Dsat(1,2)}}{g_{m(1,2)}} = \frac{2}{\left(\frac{g_m}{I_{Dsat}}\right)_{(1,2)}}, \tag{4.27}$$

and shown in Figure 4.21. Since V_{DPsat} is directly proportional to the inversion factor, it becomes possible to easily determine the inversion operating region to achieve the required differential input range.

On the other hand, if both input signals decrease or increase at the same time, the common-mode input range must be defined. It represents the range of input common-mode signal in which both transistors operate in the (transistor) saturation region and the tail current remains constant. Again, the equivalent transconductance is constant or within a certain error in the input common-mode signal range, and the acceptable error must be specified for the performance evaluation.

When a differential pair is integrated on a chip, its two transistors never have the same characteristics, and consequently, there exists voltage and current mismatch. The gate voltage mismatch is the most critical one, because the differential pair usually represents the input stage of the circuit and converts the small input signal into a current. Its standard deviation is calculated, according to [4], as

$$\sigma(\delta V_G) = \sqrt{\sigma_T^2 + \left(\frac{I_{Dsat}}{g_m}\right)^2 \cdot \sigma_\beta^2}, \tag{4.28}$$

where $\sigma_T = A_T/\sqrt{WL}$, $\sigma_\beta = A_\beta/\sqrt{WL}$, and A_T and A_β are the parameters of the technology.

Figure 4.22 depicts the standard deviation of the gate voltage mismatch as a function of the inversion factor for different transistor areas. As is expected, a larger area results in a smaller voltage mismatch. In addition, since voltage mismatch is inversely proportional to the g_m/I_{Dsat} ratio, it decreases if the transistor operates in moderate or weak inversion.

Finally, it is important to note that the cascode transistors (M3,4) do not contribute to the input voltage mismatch, as will be shown in the next chapter where the equivalent input offset is calculated at the circuit level. The same holds for the equivalent noise contribution. However, the noise contribution of

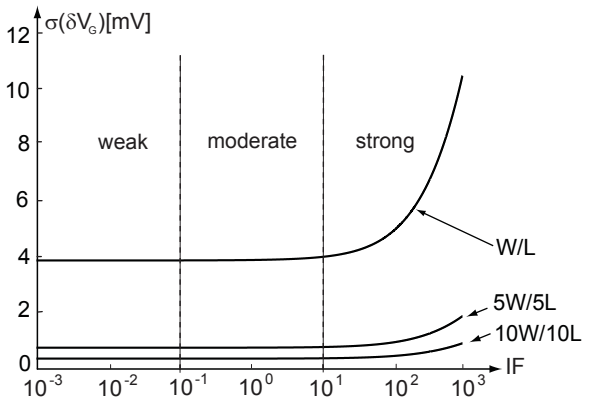


Figure 4.22 Standard deviation of the gate voltage mismatch as a function of the inversion factor for different transistor areas

the differential pair transistors (M1,2) is the most important one at the circuit level, since it represents the input stage contribution.

According to the previous definitions, we can conclude that the transistor design parameters of interest are:

- transconductance g_{m1},
- saturation voltage V_{DSsat1},
- intrinsic gate-oxide capacitance, and
- noise contribution.

Design cases - There are three possible design situations:

- differential pair in the amplifier with negative feedback, that means $v_{IN,\,DM} \to 0$, $V_{IN+} \approx V_{IN-} \approx V_{IN,\,CM}$,
- differential pair in the amplifier in the unity-gain configuration (voltage follower), that means $v_{IN,\,DM} \to 0$, $V_{IN+} \approx V_{IN-}$, whereas $V_{IN,CM}$ has a large swing, and
- differential pair in the CMFB amplifier, which means $v_{IN,DM}$ has a large amplitude, whereas $V_{IN,\,CM} \approx V_{CM,\,REF}$.

This is illustrated in Figure 4.23 a, b, c, respectively.

Differential pair in the amplifier in the negative feedback - In the first design situation, the differential input signal is small due to the negative feedback, and it can be considered that the operating point stays at the origin of the transfer characteristic. The two differential pair transistors have the same inversion factor and the same transconductance. Then, we can write

$$I_{BIAS} = I_{Dsat1} + I_{Dsat2} = I_S \cdot (IF_1 + IF_2) = 2 \cdot I_S \cdot IF. \tag{4.29}$$

The voltage imposed at the source of transistors M1 and M2 is directly proportional to the DC input common-mode voltage and is equal to

$$V_{S0} = \frac{V_{IN,CM} - V_{T0} - 2nV_t \cdot ln\left[exp\left(\sqrt{\frac{I_{BIAS}}{2I_S}}\right) - 1\right]}{(n)}. \tag{4.30}$$

The slope factor n appears in the denominator if the source and the bulk terminal are not shorted. The situation when the source terminal is connected to the bulk terminal is more favorable because the V_{GS} voltage overdrive is smaller in this case. Moreover, the differential pair is in a separate well, and thus isolated from the rest of the circuit.

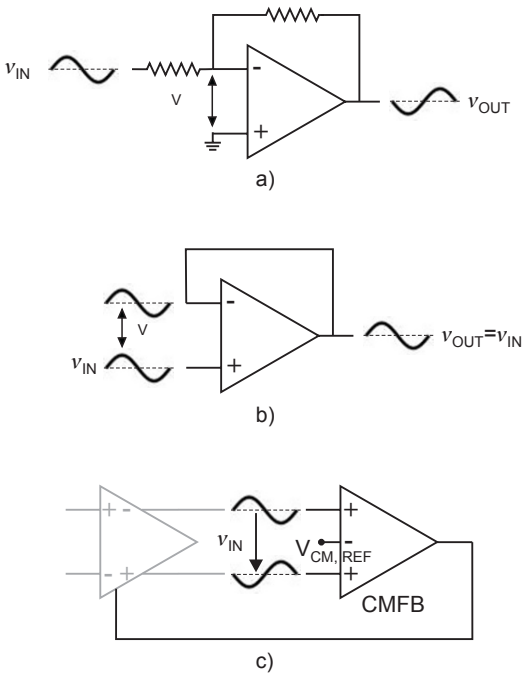


Figure 4.23 Possible design situations for the differential pair used an input stage of the amplifier

The design of the differential pair, in the "negative feedback" design situation, is equivalent to the design of a simple transistor, where the following design cases, already discussed in chapter 2, are possible:

- transconductance + saturation voltage,
- transconductance + equivalent noise, and
- transconductance + parasitic capacitances.

However, there are two additional cases:

- transconductance + voltage mismatch contribution, and
- transconductance + parasitic capacitances + voltage mismatch contribution.

According to Figure 4.22 and Figure 2.14, if a transistor operates in weak inversion, the voltage mismatch contribution is decreased, and the transconductance achieves its maximal value for the given bias current. In contrast, if the transient response of the circuit is a design requirement with high priority, according to Figure 2.19, the compromise can be found around $IF = 1$.

Differential pair in the amplifier in the unity gain configuration - When an amplifier works as voltage follower, the differential input signal is small, but the common-mode input signal has a large swing. In this case, the transistors M1 and M2 work in the same inversion region and the equivalent transconductance is constant as long as the tail current is constant. Since the source voltage of the differential pair depends on input common-mode voltage

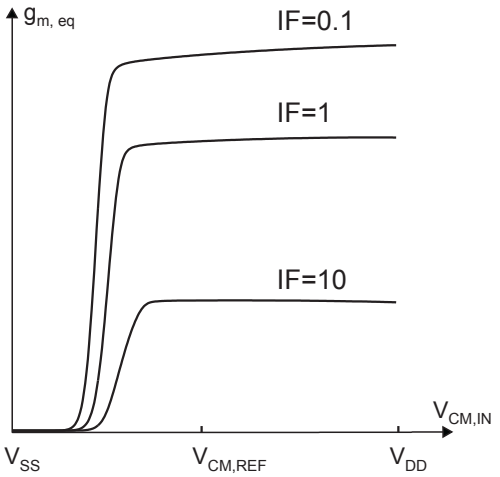


Figure 4.24 Common-mode input range limitation for a transistor operating in weak, moderate and strong inversion

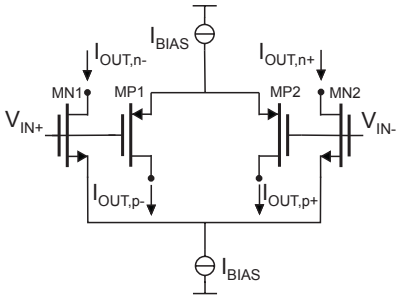


Figure 4.25 Complementary differential pair

(equation (4.30)), the tail current starts to decrease when this voltage starts to reduce the voltage range ΔV_{BIAS} required for the correct operation of the bias stage, as will be illustrated in more detail in chapter 7. At this point, it is only important to note that the common-mode voltage range is limited by ΔV_{BIAS} and the gate-source voltage headroom of the differential pair. The first one is not a design parameter here, while the second one is a design parameter and depends on the inversion operating region of the differential pair transistors (Figure 4.10). As a result, the common-mode input range also depends on the inversion factor. This is presented in Figure 4.24 for the differential pair transistors operating in weak, moderate and strong inversion (using the same bias stage in all three cases).

Hence, the most common design case for the differential pair at the voltage follower input is: transconductance + saturation voltage + CMIR. Obviously, weak inversion is the most favorable operating region for all three design parameters. However, if rail-to-rail operation is required, a possible solution is a differential pair variant called the complementary differential pair, depicted in Figure 4.25 and analyzed in chapter 7.

Differential pair in the CMFB amplifier - Finally, when a large differential input range is required, the saturation voltage of the differential pair has to be improved. This can be done by moving the transistor towards strong inversion operation, as illustrated in Figure 4.26.

The additional design case is thus: transconductance + differential pair saturation voltage, and it may require a search for the design optimum. A good initial solution is $IF = 1$. An alternative solution is a topology variant, analyzed in chapter 7, and illustrated in Figure 4.27.

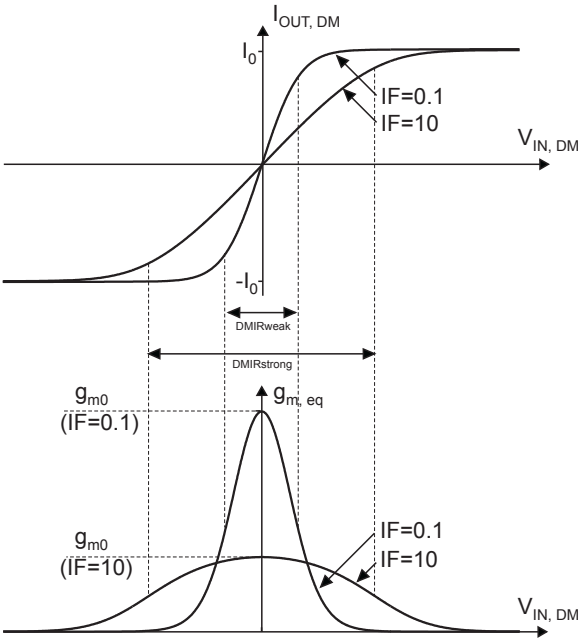


Figure 4.26 Differential input range for a transistor operating in weak and strong inversion

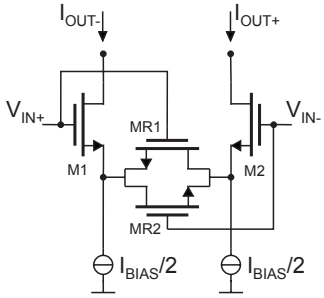


Figure 4.27 Linearized differential pair

4.5 Load structures

Simple and cascode current mirrors are used either as load or as current bias structures. The design parameters and the design cases are basically the same, with the only difference being that the specifications are usually more relaxed for a bias structure. Different voltage and current bias basic analog structures are presented in the next section, whereas this section is dedicated to the important points in the design of load structures.

4.5.1 Simple current mirror

Figure 4.28 shows a simple current mirror used as an active load in the case of single-ended (a) and fully-differential (b) amplifiers. This kind of load stage can be used with the differential pair, or some of its variants. In the fully-differential case (Figure 4.28b), the output current is the difference between the differential pair output currents at the negative and the positive terminals, and then the current mirror transistors act as current sources, requiring an additional voltage (and current) bias. If the amplifier is single-ended, the output current is again the difference between the differential pair output currents at negative and positive terminals. The result is obtained thanks to current mirroring from negative to positive terminal implemented as in Figure 4.28a.

The most important feature of a current mirror is a precise current mirroring. In the case of the simple current mirror, an electrical mismatch occurs as a consequence of the Early effect and finite output conductance. The drain voltages of transistors M1 and M2 are imposed by the circuit environment, and thus they are never exactly the same. This results in a difference in their drain currents, that can be eliminated only if the two drain voltages are kept exactly the same.

In addition, it must be noted that when the current mirror transistors are integrated on a chip, they do not have exactly the same characteristics due to the variation of process parameters, temperature or stress. As a result, there is also gate voltage and current statistical mismatch. According to [4], this is a

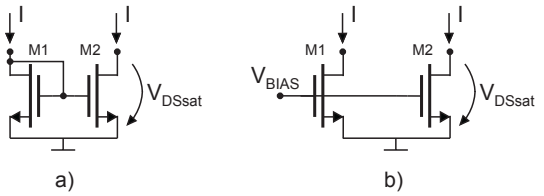


Figure 4.28 Simple current mirror as an active load with a) single b) differential outputs

design parameter of the current mirror, and depends on the transistor area and the inversion operating region. Its standard deviation is evaluated as

$$\sigma\left(\frac{\delta I_{Dsat}}{I_{Dsat}}\right) = \sqrt{\sigma_{\beta}^2 + \left(\frac{g_m}{I_{Dsat}}\right)^2 \cdot \sigma_T^2}, \tag{4.31}$$

where in the case of the same transistor sizes $\sigma_T = A_T/\sqrt{WL}$, $\sigma_\beta = A_\beta/\sqrt{WL}$, and A_T and A_β are parameters of the technology.
In the case when there is an additional voltage bias with different transistor sizes (for example, for load current sources in a fully-differential amplifier as shown in Figure 4.28b),

$$\sigma_T = \frac{A_T}{\sqrt{2}}\sqrt{\frac{1}{(WL)_{BIAS}} + \frac{1}{WL}}, \text{ and } \sigma_\beta = \frac{A_\beta}{\sqrt{2}}\sqrt{\frac{1}{(WL)_{BIAS}} + \frac{1}{WL}}.$$

Design parameters - The parameters that affect the circuit-level performance are:

- output resistance, equal to $1/g_{DS1}$ for the current mirror transistors, and usually designed to be larger than the output resistance of the corresponding transconductance structure,
- saturation voltage, V_{DSsat1}, limiting the output swing,
- parasitic capacitances, i.e. intrinsic gate-oxide capacitance, limiting the speed and the frequency response,
- equivalent noise, and
- statistical current mismatch.

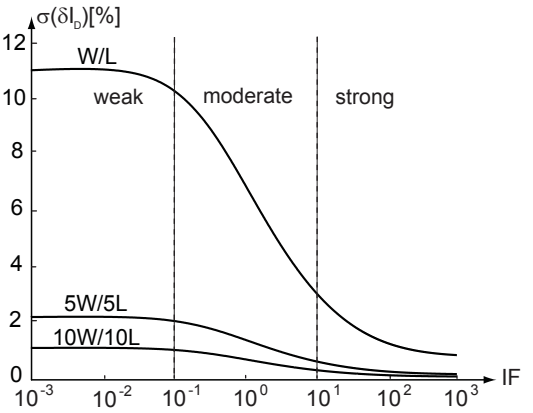


Figure 4.29 Current mismatch standard deviation as a function of the inversion factor for different transistor areas

Design cases - The most common design case is: output conductance (resistance) + saturation voltage, which directly gives the inversion factor and the transistor length of the transistors M1 and M2. The additional design cases are:

- output conductance + saturation voltage + noise,
- output conductance + saturation voltage + parasitic capacitances, and
- output conductance + saturation voltage + current mismatch.

The first two have already been discussed in chapter 2. In the last case, since the current mismatch standard deviation is directly proportional to the g_m/I_{Dsat} ratio, the most favorable operating region is strong inversion, as illustrated in Figure 4.29. If the saturation voltage has to be minimized, the design compromise can be either the limit of strong inversion ($IF = 10$) or moderate inversion operation with an increased transistor area in the following way: nW/nL, where n is a multiplier.

4.5.2 Cascode current mirror

Cascode current mirror (Figure 4.30) enables more precise current mirroring because the drain voltage of transistors M1,2 is imposed by the cascode transistors. It is used as an active load for cascoded or folded cascoded differential pairs, since it provides a high output resistance.

The cascode transistors demand an additional voltage bias. It is usually calculated in the same way as for the cascode structure transistor by keeping the bottom transistor at the limit of saturation.

An important advantage, that will be discussed further in the next chapter, is that cascode transistors do not contribute to the equivalent input noise and offset of the amplifier.

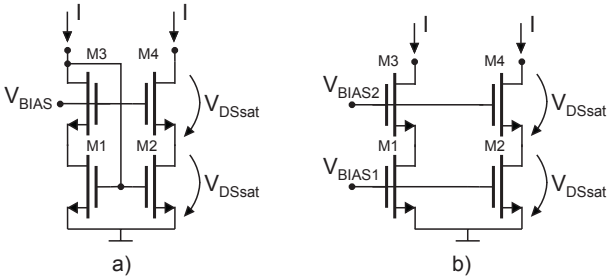


Figure 4.30 Cascode current mirror as an active load with a) single b) differential outputs

Design parameters - The design parameters that affect the circuit performance are:

- output resistance, given by

$$R_{OUT} \approx \frac{1}{g_{DS1}} \cdot \frac{g_{m2}}{g_{DS2}} \tag{4.32}$$

- voltage headroom, approximated as $V_{DSsat1} + V_{DSsat2}$,
- parasitic capacitances, i.e. gate-oxide capacitances of all transistors,
- equivalent noise, and
- current mismatch.

Design cases - The transistors M1,2 are designed in the same way as in the simple current mirror presented in the previous subsection, whereas the transistors M3,4 are designed as transistor M2 in the cascode structure presented in subsection 4.4.4.

4.6 Bias structures

When an analog circuit is integrated, the number of external current sources is limited. Hence, there is one main current bias stage that generates all required bias currents. These bias currents are then used to generate local circuit voltage and current biases, as illustrated in Figure 4.31.

4.6.1 Voltage bias

The simplest way to generate a bias voltage is a diode-connected transistor depicted in Figure 4.32a. It is generally used for a simple current mirror biasing (Figure 4.28b) or a cascode transistor biasing (Figure 4.15). The generated voltage V_{BIAS} depends on the bias current I_{BIAS} and is given by

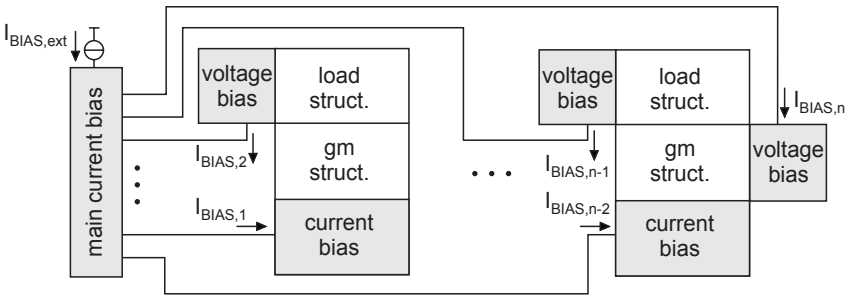


Figure 4.31 Current and voltage biasing on the chip

$$V_{BIAS} = V_{T0} + 2nV_t \cdot ln\left[exp\left(\sqrt{\frac{I_{BIAS}}{I_S}} \right) - 1 \right]. \tag{4.33}$$

If a simple current mirror is biased, the diode-connected transistor sizes are matched with the active load transistors in the following way

$$\frac{(W/L)_{diode}}{(W/L)_{mirror}} = \frac{I_{BIAS}}{I_{mirror}} \tag{4.34}$$

where the transistor length is the same for all transistors, and the ratio of the transistor widths corresponds to the ratio of the currents.

If a large DC bias voltage has to be generated, for example for cascode transistor biasing, two diode-connected transistors can be stacked as shown in Figure 4.32b. Now, the $V_{BIAS} > 2V_{T0}$.

However, the most often used topology is the one presented in Figure 4.32c. This makes it possible to have $V_{T0} < V_{BIAS} < 2V_{T0}$. Here, the transistor M1 operates in the linear region and acts as an equivalent resistance in the source of diode-connected transistor M2.

Finally, when a small bias voltage is needed, the topology depicted in Figure 4.33 can be used. The bias voltage is then calculated as follows. According to [5], the transistors M1 and M2 connected in series act as an equivalent transistor, whose sizes are calculated as

$$\frac{L_{eq}}{W_{eq}} = \frac{L_1}{W_1} + \frac{L_2}{W_2}. \tag{4.35}$$

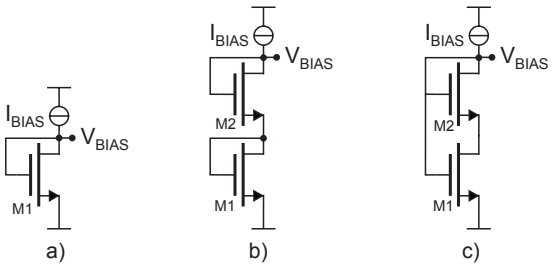


Figure 4.32 Topologies for voltage bias generation: a) diode-connected transistor, b) two stacked diode-connected transistors, c) diode-connected transistor with the equivalent resistance in the source

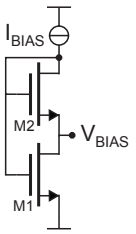


Figure 4.33 Low bias voltage generation

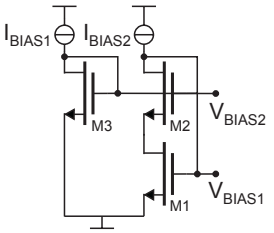


Figure 4.34 Cascode current mirror bias voltage generation

Thus,

$$V_{BIAS} = 2nV_t \cdot ln\left\{\left[exp\left(\sqrt{\frac{I_{BIAS}}{I_{S,eq}}}\right) - 1\right] / \left[exp\left(\sqrt{\frac{I_{BIAS}}{I_{S2}}}\right) - 1\right]\right\}, \quad (4.36)$$

where $I_{S,eq} = 2nKP\left(\frac{W_{eq}}{L_{eq}}\right)V_t^2$ and $I_{S2} = 2nKP\left(\frac{W_2}{L_2}\right)V_t^2$.

Figure 4.34 displays a voltage bias suitable for the cascode current mirror (Figure 4.30b). Transistors M1 and M2 are matched with the cascode current mirror transistors, i.e. the transistor lengths are the same and the transistor widths are calculated from the ratio of the currents. The transistor M3 is sized as explained previously to generate V_{BIAS2} that is required to keep the transistor M1 in the saturation region.

4.6.2 Current bias

The bias currents for the transconductance structures are generated using simple or cascode current mirrors. All bias currents are matched and generated as illustrated in Figure 4.35. A topology for the generation of bias currents for different analog cells from one external source is shown in Figure 4.36. The

resistance R_{BIAS} is necessary in order to avoid start-up problems. When there is no external current source, the topology depicted in Figure 4.37 can be used.

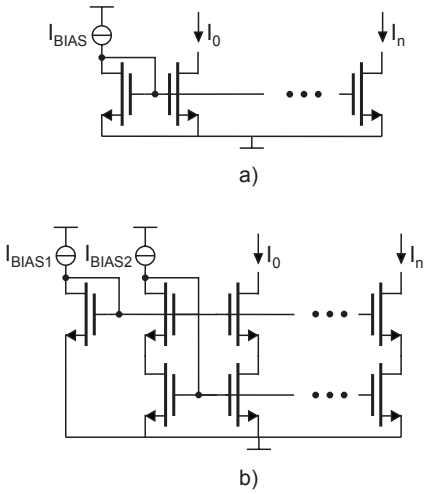


Figure 4.35 Bias currents generation: a) simple and b) cascode current mirrors

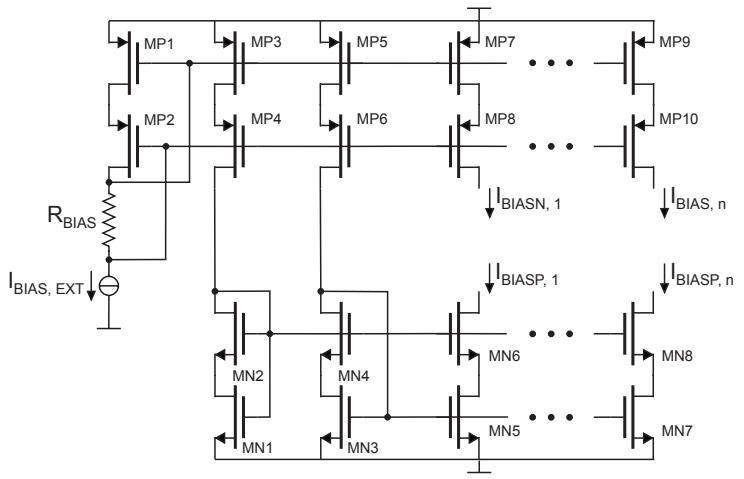


Figure 4.36 Main current biasing from one external current source

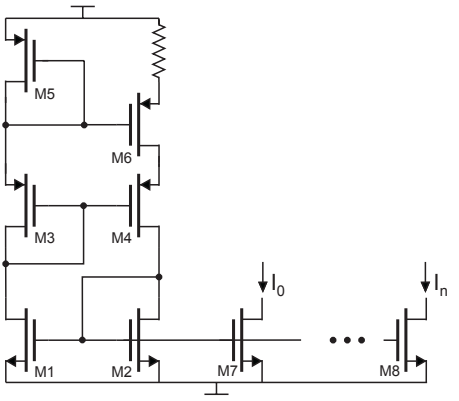


Figure 4.37 Current bias generation without external current source

4.7 Conclusion

The concept of structured design facilitates the design of analog cells because each complex analog structure is partitioned into simple analog building blocks that affect only few circuit-level parameters and demand less design effort. There is a finite number of basic analog structures and they can be classified as transconductance, load and bias structures. Each basic analog structure class has the same dominant design parameters and affects the same circuit-level parameters. Consequently, the design specifications are derived for any circuit complexity in the same way. Moreover, if a certain design parameter cannot be achieved, the basic analog structure in question can be replaced by its more advanced version without any need to change the design procedure. Finally, each basic analog structure can be seen as a set of transistors in the design environment. When the basic analog structure specifications are determined, the design consists in the determination of the appropriate transistor-level design case and the use of a corresponding transistor-level design recipe.

Bibliography

[1] P. Allen, D. Holberg, “CMOS Analog Circuit Design”, Oxford University Press, second edition, January 2002, ISBN10: 0195116445

[2] M. del Mar Hershenson, S. P. Boyd, T. H. Lee, “GPCAD: a tool for CMOS op-amp synthesis”, International Conference on Computer-Aided Design, pp. 296-303, November 1998

[3] M. del Mar Hershenson, S. P. Boyd, T. H. Lee, “Optimal design of a CMOS op-amp via geometric programming”, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, pp. :1-21, January 2001

[4] Ch. C. Enz, E. A. Vitoz, “Charge-based MOS Transistor Modeling: The EKV Model for Low-power and RF IC Design”, John Wiley & Sons, first edition, 2006, ISBN 9780470855416

[5] M. Pastre, “Methodology for the Digital Calibration of Analog Circuits and Systems - Application to a Hall Sensor Microsystem”, PhD thesis N°3210, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2005

CHAPTER 5

Procedural design scenarios

The procedural design sequence for the design of an analog amplifier is introduced first. Then, the procedural design scenarios for three widely used analog cells: folded cascode OTA, fully-differential folded cascode OTA and Miller operational amplifier are presented. For each cell, the circuit partitioning, the derivation of the specifications, the procedural design steps, as well as the local optimization steps are discussed in detail.

5.1 Introduction

After the analog cell partitioning into its building blocks, and the derivation of the design specifications for each basic analog structure, the basic structures have to be designed in a specific order for the following reasons:

- **propagation of the specifications** - the design parameters of already sized basic analog structures may affect the design parameters, i.e. design specifications of the basic analog structures that are not yet designed,
- **verification of the parameters bounds** - the design parameters of already sized basic analog structures may impose lower or upper limits on the design parameters of some other basic analog structures,
- **local optimization** - the design trade-offs can be found only at the level of the basic analog structures; since the design sequence imposes the design priorities, it becomes possible to estimate which parameters can be relaxed and at what cost.

5.1.1 Procedural design sequence for an analog amplifier

Since the analog amplifier is one of the most commonly used analog cells [1]-[6] in analog and mixed-mode systems, a general procedural design sequence for an analog amplifier is defined in this section. Then, the examples of the procedural design scenarios for three different amplifier topologies are presented in the next sections. Nevertheless, any other analog cell can be designed in a similar way and using the same design approach. The procedural

design sequence is always defined with regard to the design priorities and the basic building blocks.

Classification of analog amplifiers - In the most general terms, all analog amplifiers are either current or voltage amplifiers. The so-called transconductance amplifiers can be considered as a special case of voltage amplifiers, since the transconductance structure always requires a load structure that converts an output current into an output voltage.

In this analysis, we will focus on voltage amplifiers that can be classified as proposed in Figure 5.1:

- **Single-input single-output amplifiers** - These amplifiers have one input and one output terminal, and can be used in stand-alone mode or as output stages. If they are realized as class A amplifiers, they consist of a transconductance structure such as common-source, common-drain or cascode and the appropriate load/bias structure. The examples of class B, or class AB amplifiers are inverter or push-pull amplifier, but they will not be taken into consideration here.
- **Operational transconductance amplifiers (OTAs)** - These amplifiers convert the differential input voltage into output current, and they are usually used in negative feedback with a capacitive load or as an input stage of an operational amplifier. The main features are: moderate (or high) gain, high output impedance and high output current sourcing/sinking capability. They consist of a differential transconductance structure (i.e. a differential pair or one of its variants), a current bias structure and a load structure.

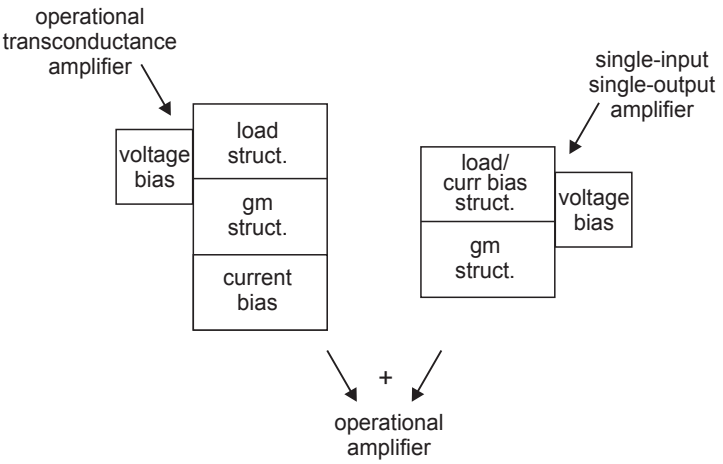


Figure 5.1 Analog voltage amplifiers

- **Operational amplifiers** - These amplifiers are defined as analog amplifiers that have a large input impedance, a small output impedance and a high gain. They amplify a differential input signal and can be used in negative feedback with resistive or resistive and capacitive load. These amplifiers are generally realized as a two-stage amplifiers composed of a transconductance amplifier followed by an output stage.

Figure 5.1 depicts the described voltage amplifiers partitioned into basic analog structures. We limit the analysis here to two-stage amplifiers for simplicity. It is however possible to have several gain stages and an interesting study of the design of such amplifiers is given in [6].

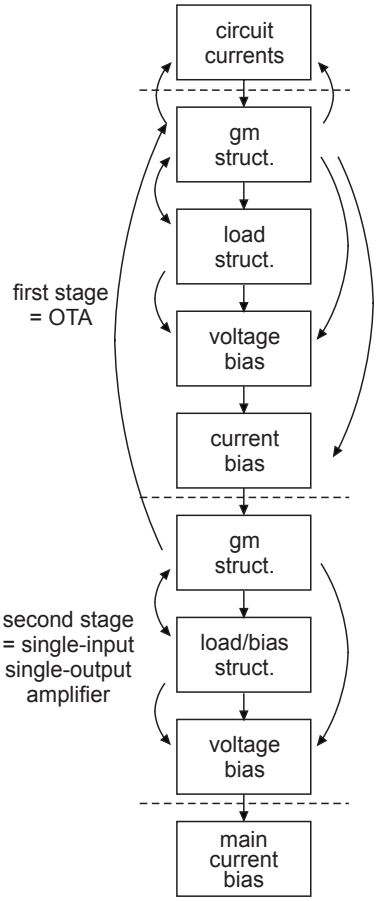


Figure 5.2 Operational amplifier procedural design sequence

Design sequence - The procedural design sequence for the design of two-stage operational amplifier is shown in Figure 5.2. The forward arrows on the right-hand side indicate the basic analog structures whose design parameters may be affected by the current design step. The backward arrows on the left-hand side indicate the possible redesign or optimization loops.

Before any basic analog structure sizing step, the circuit currents have to be determined according to total current consumption or speed requirements. The basic structures are then designed in the following order: **transconductance - load - voltage bias - current bias structure**. A transconductance amplifier is designed as the first stage, whereas a single-input single-output amplifier is designed as the second stage. The development of the procedural design sequences in the form of design scenarios will be illustrated using the examples of:

- folded cascode operational transconductance amplifier,
- fully-differential folded cascode operational transconductance amplifier, and
- Miller operational amplifier.

5.1.2 Definitions of the circuit-level design parameters

The definitions of the circuit-level design parameters that are used in the examples that follow are given hereafter.

DC gain - The DC gain of each amplifier stage is defined as a product of the equivalent transconductance of the transconductance structure and the equivalent resistance seen at the output

$$A_0 = g_m R_{OUT}. \tag{5.1}$$

Input range - The signal range at the amplifier input can be defined with respect to the differential signal and with respect to the common-mode signal (Figure 5.3).

Differential input range corresponds to the amplitude of the differential input signal $\pm\Delta v_{IN,\,DM}$ for which the differential pair at the amplifier input does not saturate (see subsection 4.4.5). For design purposes, it is often translated into maximal and minimal DC voltage level at each amplifier input that guarantees that the operating point is still in the linear region of the differential pair transfer characteristic, that is

$$V_{IN+/-,\,DM,\,MIN} \leq V_{CM,\,REF} - \frac{\Delta v_{IN,\,DM}}{2}, \text{ and} \tag{5.2}$$

$$V_{IN+/-,\,DM,\,MAX} \geq V_{CM,\,REF} + \frac{\Delta v_{IN,\,DM}}{2}. \tag{5.3}$$

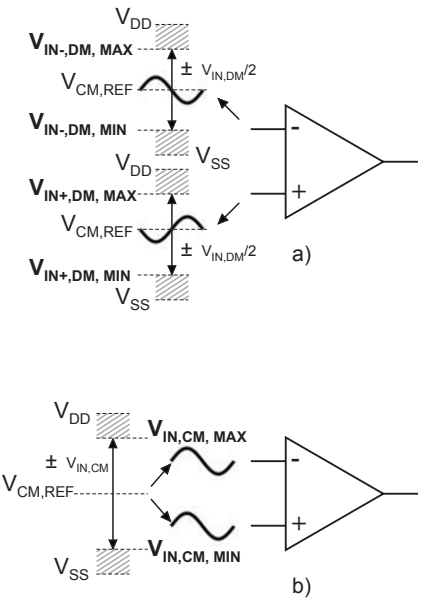


Figure 5.3 Illustration of a) differential and b) common-mode input range

Common-mode input range corresponds to the amplitude of the common-mode input signal $\pm\Delta v_{IN,\,CM}$ for which the working conditions of the differential pair are respected when both input terminal signals increase or decrease (see subsection 4.4.5). For design purposes, it is also translated into maximal and minimal common-mode DC level at the input, that is

$$V_{IN,\,CM,\,MIN} \leq V_{CM,\,REF} - \Delta v_{IN,\,CM}\text{ , and} \tag{5.4}$$

$$V_{IN,\,CM,\,MAX} \geq V_{CM,\,REF} + \Delta v_{IN,\,CM}. \tag{5.5}$$

Output range (swing) - The output voltage range, also denoted as output swing, $\pm\Delta v_{OUT}$ corresponds to the amplitude of the output or the differential output signal for which the transistors of the output stage do not move into the (transistor) linear operating region. For design purposes, this can be translated into maximal and minimal DC voltage level at the output terminals, as depicted in Figure 5.4. For the single-ended output case, we can write

$$V_{OUT,\,MIN} \leq V_{CM,\,REF} - \Delta v_{OUT}\text{ , and} \tag{5.6}$$

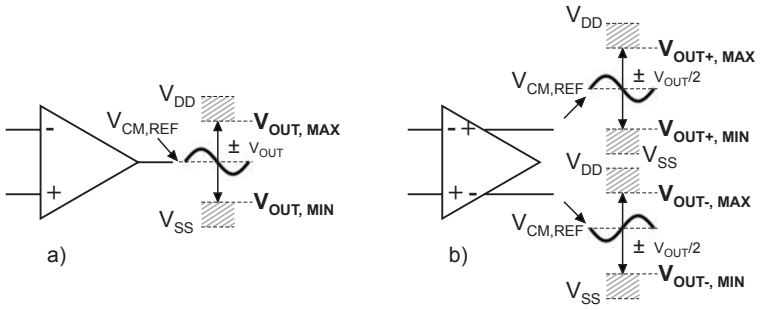


Figure 5.4 Illustration of a) single-ended and b) differential output voltage range

$$V_{OUT,\,MAX} \geq V_{CM,\,REF} + \Delta v_{OUT}, \tag{5.7}$$

while for the differential output case

$$V_{OUT+/-,\,MIN} \leq V_{CM,\,REF} - \frac{\Delta v_{OUT}}{2}, \text{ and} \tag{5.8}$$

$$V_{OUT+/-,\,MAX} \geq V_{CM,\,REF} + \frac{\Delta v_{OUT}}{2}. \tag{5.9}$$

Frequency response - When an amplifier is used in negative feedback it is mandatory to analyze its frequency response and to check the stability of all feedback loops.

The poles and zeros in the transfer function can be easily determined analyzing the gain path, i.e. the signal path from the amplifier input to the amplifier output, and using the following guidelines proposed in [2]:

- **pole** - corresponds to an equivalent resistance in parallel with an equivalent capacitance seen at each node in the gain path,
- **zero** - corresponds to an equivalent resistance in series with an equivalent capacitance in the gain path,
- **positive zero** - appears due to a possible feedback path in the gain path (for example as in the case of Miller capacitances),
- **doublet pole-zero** - appears as a consequence of two possible gain paths (for example as in the case of single-ended topology).

The amplifier in negative feedback is unconditionally stable if it behaves as a system of the first order, that is if there exists only one pole in its transfer function. However, there usually exist several nodes (poles) in the gain path, and thus the stability is ensured only if the phase shift of any feedback loop is

bigger than -135°. This implies that there is one dominant pole, and all other poles must be non-dominant, i.e. far away from the frequency where the amplifier gain becomes less than one.

In the most general case, the amplifier stability is analyzed in the most unfavorable case, that is when the negative feedback factor (usually denoted as β) is equal to one. However, in the real design situation the stability of the main feedback loop must be verified with the appropriate feedback factor value. In addition, the stability of all other feedback loops (for example the CMFB loop in the case of a fully-differential amplifier) must be checked.

The important design parameters are the following:

- **bandwidth frequency** - corresponds to the dominant pole frequency f_{dp} in the open loop transfer function,
- **gain bandwidth frequency** f_{GBW} - corresponds to the product of the DC gain and the bandwidth frequency of the open loop transfer function,

$$f_{GBW} = A_{DC} \cdot f_{dp} = A_0 \cdot \beta \cdot f_{dp}, \tag{5.10}$$

- **phase margin** - represents the difference between the phase shift at the gain bandwidth frequency and the phase shift of -180°

$$PM = 180° + \varphi(f_{GBW}). \tag{5.11}$$

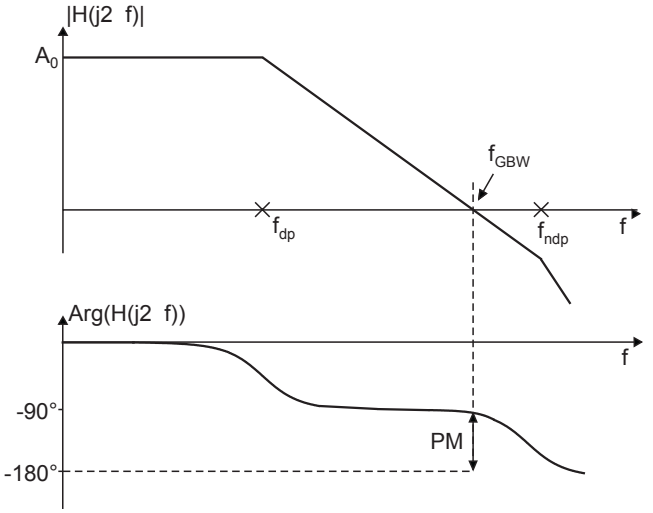


Figure 5.5 Illustration of dominant, non-dominant pole, gain bandwidth frequency and phase margin for the amplifier in open unity-gain feedback loop

Figure 5.5 illustrates the stability analysis parameters in the case of the unity-gain negative feedback. The loop gain corresponds to the DC gain of the amplifier, while the corresponding gain bandwidth frequency is often denoted as the **amplifier gain bandwidth frequency** and is given by $f_{GBW} = A_0 \cdot f_{dp}$.

Rejection ratios - The positive and negative supply voltages, as well as the common-mode input voltage are DC voltages, and thus, they are considered as constant in the ideal case. However, in real design situations there may exist perturbations of these DC voltages, that are then amplified and found in the output voltage as unwanted components. Thanks to negative feedback, these contributions are attenuated, in fact divided by the differential gain of the amplifier or by the gain of the CMFB loop, as will be shown later. An important design parameter thus becomes the ratio of the amplifier (or the CMFB loop) gain and the gain of the undesirable small-signal component denoted as:

- **positive power supply rejection ratio** $PSRR^+$,
- **negative power supply rejection ratio** $PSRR^-$,
- **common-mode rejection ratio** $CMRR$.

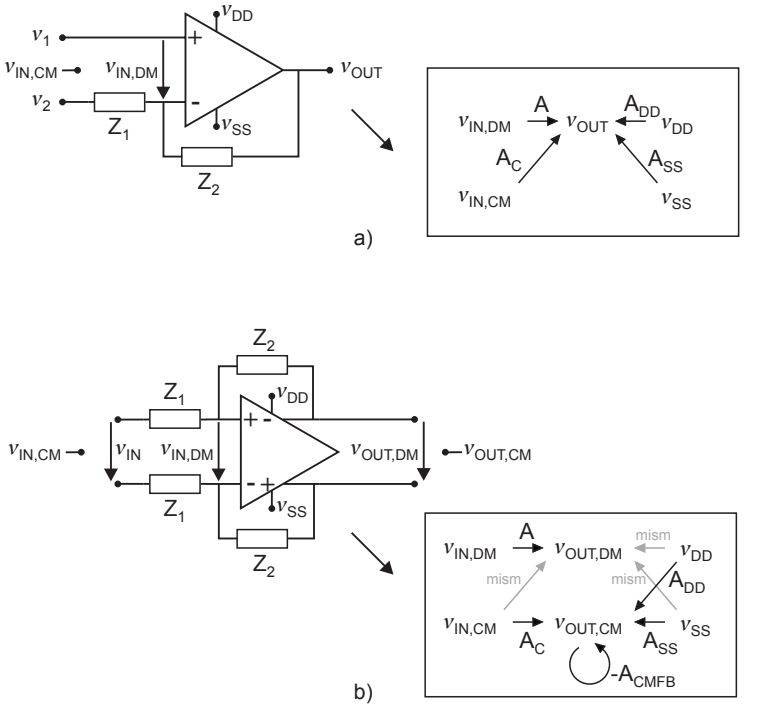


Figure 5.6 Illustration of the signal and the perturbation gain paths in the case of a) single-ended and b) fully-differential amplifier

Table 5.1 Rejection ratio definitions for the single-ended amplifier

CMRR	$\frac{A}{A_C}$
PSRR+	$\frac{A}{A_{DD}}$
PSRR-	$\frac{A}{A_{SS}}$

To determine the rejection ratios we will analyze the equivalent contribution of each voltage source in the case of a single-ended amplifier and in the case of a fully-differential amplifier both operating with a negative feedback network (Figure 5.6).

For the single-ended amplifier, shown in Figure 5.6a, the small-signal output voltage is given by

$$v_{OUT} = \left(1+\frac{Z_2}{Z_1}\right)v_1 - \frac{Z_2}{Z_1}v_2 + \left(1+\frac{Z_2}{Z_1}\right)\frac{A_{SS}}{A}v_{SS} + \left(1+\frac{Z_2}{Z_1}\right)\frac{A_{DD}}{A}v_{DD} + \left(1+\frac{Z_2}{Z_1}\right)\frac{A_C}{A}v_{IN,\,CM}, \tag{5.12}$$

where A is the gain of the amplifier, A_{DD} is the gain of the perturbation from the positive supply voltage source to the output, A_{SS} is the gain of the perturbation from the negative supply voltage source to the output and A_C is the gain of the perturbation from the common-mode input voltage equivalent source to the output. On this basis, the rejection ratios are defined with regard to the amplifier gain as shown in Table 5.1.

For the fully-differential amplifier, shown in Figure 5.6b, the gain of unwanted perturbations is the same at positive and negative amplifier output terminals. Hence, this is canceled for a perfectly balanced circuit since $v_{OUT} = v_{OUT-} - v_{OUT+}$. The unwanted contributions occur only if there exists an electrical or statistical mismatch of the transistors on the right and the left part of the circuit

$$v_{OUT,\,DM} = -\frac{Z_2}{Z_1}v_{IN,\,DM} + \frac{A_{SSmism}}{A}v_{SS} + \frac{A_{DDmism}}{A}v_{DD} + \frac{A_{Cmism}}{A}v_{IN,\,CM}. \tag{5.13}$$

Table 5.2 Rejection ratio definitions for the fully-differential amplifier

CMRR	$\frac{A_{CMFB}}{A_C}$
PSRR+	$\frac{A_{CMFB}}{A_{DD}}$
PSRR-	$\frac{A_{CMFB}}{A_{SS}}$

Since A_{SSmism}, A_{DDmism} and A_{Cmism} are very small, and divided by the amplifier gain, these contributions are usually not critical.

On the other hand, since the output common-mode voltage is determined as $v_{OUT,\,CM} = 0.5(v_{OUT+} + v_{OUT-})$, we can write that

$$v_{OUT,\,CM} = \frac{A_{DD}}{1 + A_{CMFB}} v_{DD} + \frac{A_{SS}}{1 + A_{CMFB}} v_{SS} + \frac{A_C}{1 + A_{CMFB}} v_{IN,\,CM}, \tag{5.14}$$

where A_{CMFB} is the gain of the common-mode feedback loop. These contributions are usually dominant, and thus, the rejection ratios are here defined with regard to the gain of the common-mode feedback loop as given in Table 5.2.

Noise - For each transistor in the circuit, the noise contribution is modeled as an equivalent voltage source at the transistor gate equal to its input referred noise voltage spectral density, as defined in subsection 2.2.13. The amplifier noise is expressed as the **equivalent input referred noise voltage spectral density**, calculated in the following way: for each "noise" voltage source, the gain is calculated at the output, all contributions are summed and referred to the input by dividing the sum by the amplifier gain.

Offset - For each pair of matched transistors in the circuit, the gate voltage mismatch contribution is modeled as an equivalent voltage source in the transistor gate equal to its gate voltage mismatch standard deviation, as defined in subsection 4.4.5. The amplifier offset is expressed as the **equivalent input referred offset**, calculated in the following way: for each "mismatch" voltage source, the gain is calculated at the output, all contributions are summed and referred to the input by dividing the sum by the amplifier gain.

Speed - The speed of the amplifier corresponds to the time required for the signal at the output to establish its final value or within an acceptable error. It depends on the capacitive load, the parasitic capacitances in the nodes in the gain path, as well as on the currents available to charge or discharge these capacitances. Therefore, it can be characterized by:

- **slew-rate *SR*** - defined for the unity-gain configuration and the large input voltage step as the ratio of the maximal available large signal current and the equivalent capacitance seen at the output,
- **current sourcing/sinking capability** - defined as the current that must be provided at the amplifier output for the particular working conditions (e.g. negative feedback network) and any voltage amplitude in the specified differential input voltage range.

5.2 Procedural design scenario for a folded-cascode OTA

The complete schematic of the folded cascode OTA structure together with all necessary voltage and current bias basic analog structures is shown in Figure 5.7. The same topology can be implemented using the complementary transistors, i.e. replacing all NMOS transistors by PMOS transistors and vice versa in the transconductance, load and bias structures. In both cases, the design scenario remains the same and the circuit-level design parameters are calculated in the same way. It is also important to note that the biasing circuitry, presented in this example, is just one possible way of implementation. It is decided here that all current bias sources are matched with the main current bias for circuit simplicity. Nevertheless, each current source can be implemented separately if there are additional design requirements.

5.2.1 Circuit-level design parameters

Circuit-level design parameters are calculated using the definitions presented in the previous section and summarized in Table 5.3.

It is interesting to note that the contribution of the cascode transistors M5,6 and M7,8 to the equivalent input referred noise and the equivalent input referred offset can be neglected, since the gain from the “noise” (“mismatch”) voltage source to output is much smaller than the amplifier’s gain, that is

$$\frac{A_{M5,6}}{A_0} = \frac{g_{DS3,4} \cdot R_{OUT,down}}{g_{m1} \cdot R_{OUT}} << 1, \frac{A_{M7,8}}{A_0} = \frac{g_{DS9,10} \cdot R_{OUT,up}}{g_{m1} \cdot R_{OUT}} << 1 \qquad (5.15)$$

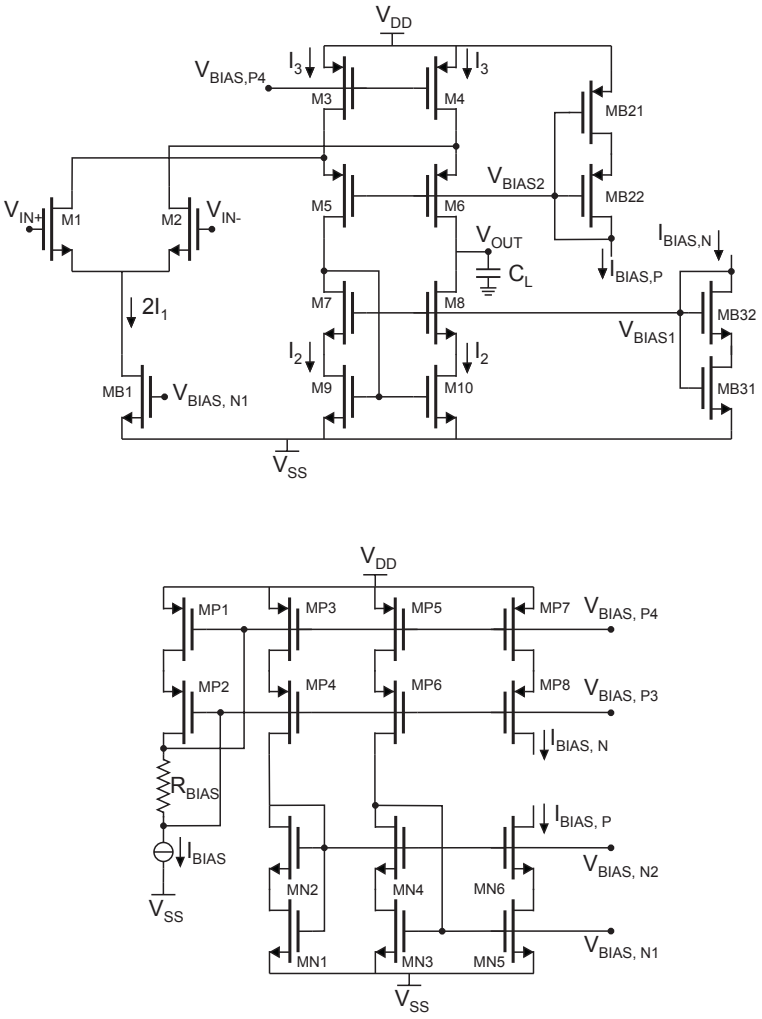


Figure 5.7 Folded cascode OTA topology with voltage and current bias structures

Table 5.3 Folded cascode OTA circuit-level design parameters

static current consumption	$\Sigma I = 2I_3 + I_{BIAS,N} + I_{BIAS,P} + 3I_{BIAS}$
gain, output resistance	$A_0 = g_{m1,2} \cdot R_{OUT},\ R_{OUT} = R_{OUT,down} \parallel R_{OUT,up}$ $R_{OUT,down} = \frac{n \cdot g_{m7,8}}{g_{DS7,8}} \cdot \frac{1}{g_{DS9,10}}$ $R_{OUT,up} = \frac{n \cdot g_{m5,6}}{g_{DS5,6}} \cdot \frac{1}{(g_{DS1,2} + g_{DS3,4})}$
common-mode input range	$V_{IN,CM,MAX} = V_{DD} - \lvert V_{DSsat3,4} \rvert - V_{DSsat1,2} + V_{GS1,2}$ $V_{IN,CM,MIN} = V_{SS} + V_{DSsatB1} + V_{GS1,2}$
output range	$V_{OUT,MAX} = V_{DD} - \lvert V_{DSsat3,4} \rvert - \lvert V_{DSsat5,6} \rvert$ $V_{OUT,MIN} = V_{SS} + V_{DSsat7,8} + V_{DSsat9,10}$
input referred equivalent noise	$v_n^2 = 2v_{n1,2}^2 + 2\left(\frac{g_{m3,4}}{g_{m1,2}}\right)^2 v_{n3,4}^2 + 2\left(\frac{g_{m9,10}}{g_{m1,2}}\right)^2 v_{n9,10}^2$
input referred equivalent offset	$v_{off}^2 = \sigma(\delta V_{G1,2})^2$ $+ \left(\frac{g_{m3,4}}{g_{m1,2}}\right)^2 \sigma(\delta V_{G3,4})^2 + \left(\frac{g_{m9,10}}{g_{m1,2}}\right)^2 \sigma(\delta V_{G9,10})^2$
slew-rate	$SR = \frac{2I_1}{C_L}$

Rejection ratios - The input common-mode signal gain is approximated by

$$A_C = \frac{1/g_{DS3,4}}{1/g_{DSB1}} \cdot g_{m6} \cdot R_{OUT}, \tag{5.16}$$

which results in

$$CMRR = \frac{g_{DS3,4}}{g_{DSB1}} \cdot \frac{g_{m1,2}}{g_{m5,6}}. \tag{5.17}$$

If the *CMRR* is not the first design priority, the transconductances of the transistors M1,2, and M5,6 and the output conductance of the transistors M3,4

are usually imposed by the gain and stability conditions. Therefore, the key design parameter is the output conductance of the current bias source MB1.

Concerning the *PSRR*, it is very difficult to determine the exact expression, since all perturbation paths including the main and the bias circuit have to be superposed. The common practice is thus to estimate the key parameters that affect the main contributions and to use them as design parameters. The hand-calculation difficulty arises from the fact that the positive and the negative supply voltage noise affects the bias voltages, and hence they can be also considered as small-signal sources. Consequently, transistors such as M3,4, M9,10 or MB1 can be considered to have a small-signal perturbation source not only in the source, but also in the gate terminal. The resulting gains are approximately the same, but with opposite signs, and thus ideally canceled. However, the calculation results never match with the simulation results, and the previous conclusion can be used only as a design guideline. On the other hand, the main contribution comes from the cascode transistors M5,6 and M7,8 since the perturbations of supply voltages result in their bias voltage small-signal changes being amplified at the output. Since the equivalent gains can be approximated by $g_{DS3,4} \cdot R_{OUT}$ and $g_{DS9,10} \cdot R_{OUT}$, it follows that the key parameters are the output conductances of transistors M3,4 and M9,10.

5.2.2 Frequency analysis

Figure 5.8 displays the equivalent half-circuit that simplifies the frequency analysis. There exist four different nodes, and two possible gain paths from the input to the output of the amplifier: the first one via the folded point FP, and the second one via the cascode point CP. Consequently, a doublet pole-zero will appear in the transfer function together with a dominant pole and two other non-dominant poles. The approximations for pole and zero frequencies are given in Table 5.4.

The gain bandwidth frequency is now calculated as

$$f_{GBW} = A_0 \cdot f_{dp} = \frac{g_{m1,2}}{2\pi(\Sigma C)_{OUT}}, \tag{5.18}$$

whereas the phase margin is given by

$$PM = 180° - atan\left(\frac{f_{dp}}{f_{GBW}}\right) - \sum_{i=1}^{3} atan\left(\frac{f_{ndpi}}{f_{GBW}}\right) + atan\left(\frac{f_z}{f_{GBW}}\right). \tag{5.19}$$

Since there are three non-dominant poles it is of interest to determine the pole that is the closest to the origin in order to define the stability condition that can be used for the design.

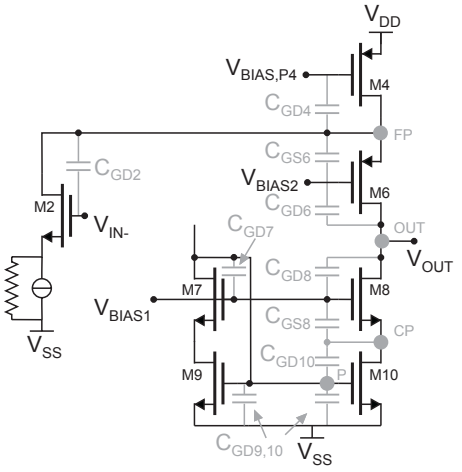


Figure 5.8 Equivalent half-circuit of the folded cascode OTA used for frequency analysis

There are two possible situations:

- $R_{OUT,up} < R_{OUT,down}$ - in this case $g_{m5,6} < g_{m7,8}$, $g_{DS5,6} > g_{DS7,8}$ and $(g_{DS1,2} + g_{DS3,4}) > g_{DS9,10}$, which implies that $IF_{7,8} < IF_{5,6}$ and $L_{7,8} > L_{5,6}$ - consequently, it can be expected that the sum of capacitances in the cascode point is larger that the sum of capacitances in the folded point, and thus it is possible to fulfil

$$\frac{g_{m7,8}}{(\Sigma C)_{CP}} < \frac{g_{m5,6}}{(\Sigma C)_{FP}}. \tag{5.20}$$

If now we impose the design condition that $IF_{7,8} > IF_{9,10}$ and $L_{7,8} = L_{9,10}$ to ensure that

$$\frac{n \cdot g_{m7,8}}{(\Sigma C)_{CP}} \leq \frac{g_{m9,10}}{(\Sigma C)_{G9,10}} \tag{5.21}$$

the stability condition can be written as

$$f_{ndp2} \geq k_{PM} \cdot f_{GBW} \tag{5.22}$$

where k_{PM} is determined by the required phase margin.

Table 5.4 Approximations for pole and zero frequencies in the transfer function of the folded cascode OTA

dominant pole f_{dp}	$\frac{1}{2\pi R_{OUT}(\Sigma C)_{OUT}}$ $(\Sigma C)_{OUT} = C_L + C_{GD5,6} + C_{GD7,8}$
non-dominant pole f_{ndp1}	$\frac{n \cdot g_{m5,6}}{2\pi(\Sigma C)_{FP}}$, $(\Sigma C)_{FP} = C_{GS5,6} + C_{GD3,4} + C_{GD1,2}$
non-dominant pole f_{ndp2}	$\frac{n \cdot g_{m7,8}}{2\pi(\Sigma C)_{CP}}$, $(\Sigma C)_{CP} = C_{GS7,8} + C_{GD9,10}$
doublet - non-dominant pole f_{ndp3}	$\frac{g_{m9,10}}{2\pi(\Sigma C)_{G9,10}}$, $(\Sigma C)_{G9,10} = 2C_{GS9,10} + g_{m10}R_{OUT}C_{GD9,10}$
doublet - zero f_z	$\frac{2g_{m9,10}}{2\pi(\Sigma C)_{G9,10}}$

- $R_{OUT,up} > R_{OUT,down}$ - in this case $g_{m5,6} > g_{m7,8}$, $g_{DS5,6} < g_{DS7,8}$ and $(g_{DS1,2} + g_{DS3,4}) < g_{DS9,10}$, which implies that $IF_{7,8} > IF_{5,6}$ and $L_{7,8} < L_{5,6}$ - consequently, it can be expected that the sum of the capacitances in the folded point is larger than the sum of the capacitances in the cascode point, and thus it is possible to achieve

$$\frac{g_{m5,6}}{(\Sigma C)_{FP}} < \frac{g_{m7,8}}{(\Sigma C)_{CP}}. \tag{5.23}$$

If now we impose the same design condition as in the previous case to ensure that the non-equality (5.21) is still valid, the stability condition can be written as

$$f_{ndp1} \geq k_{PM} \cdot f_{GBW}. \tag{5.24}$$

where k_{PM} is determined by the required phase margin.

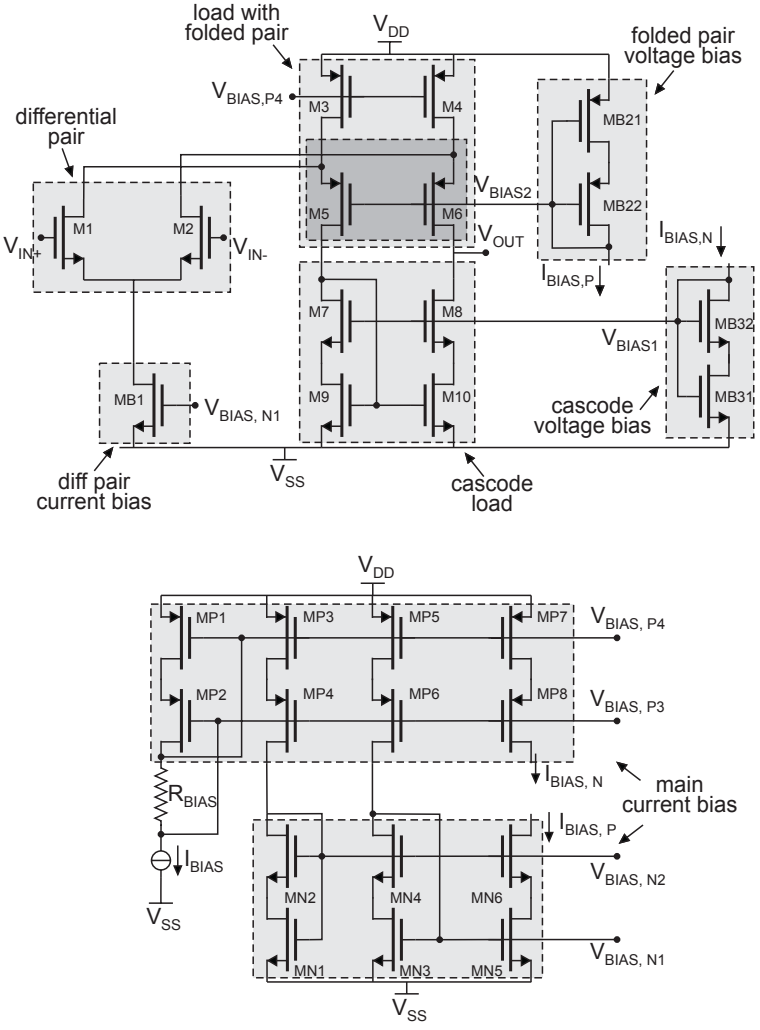


Figure 5.9 Folded cascode OTA partitioned into basic analog structures

5.2.3 Circuit partitioning

The circuit from Figure 5.7 is partitioned into the following basic analog structures:

- differential pair,
- load with folded pair = bias mirror M3,4 + folded pair,
- cascode load,
- folded pair voltage bias,
- cascode voltage bias,
- differential pair current bias,
- main current bias,

and redrawn in Figure 5.9. It must be pointed out that the folded pair belongs at the same time to the transconductance structure (folded cascode differential pair) and to the load/bias structure (on the upper side). However, because of the stability conditions, and the relations between the non-dominant pole frequencies and the resistances $R_{OUT,down}$, $R_{OUT,up}$ discussed previously, it is more convenient to design the folded pair transistors M5,6 together with the current bias transistors M3,4 to avoid the optimization loops.

5.2.4 Derivation of the specifications

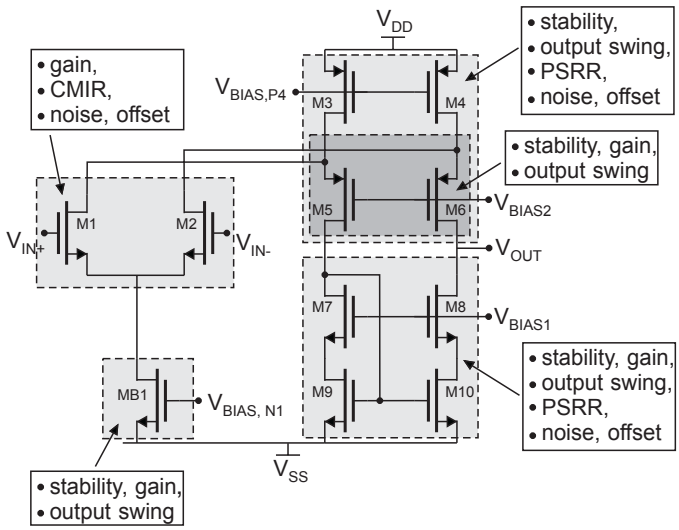


Figure 5.10 Folded cascode OTA circuit-level parameter sets that are used for the derivation of basic analog structures' specifications

The circuit-level design parameters that are used to derive the specifications for each basic analog structure are indicated in Figure 5.10. The corresponding circuit-level and basic analog structures' design parameters are listed in Table 5.5.

Table 5.5 Table for derivation of the specifications for folded cascode OTA

basic analog structure	circuit-level or other basic analog struct. parameters	basic analog structure's parameters
differential pair	• gain, • gain bandwidth, • common-mode input range, • noise, offset	• $g_{m1,2}/g_{DS1,2}$, • $g_{m1,2}$, • $V_{GS1,2}$, • g_{m1}, $v_{n1,2}$, $\sigma(\delta v_{G1,2})$
load with folded pair - folded pair	• gain, • stability, • output range	• $g_{m5,6}/g_{DS5,6}$, • $g_{m5,6}/(\Sigma C)_{FP}$, • $V_{DSsat5,6}$
load with folded pair - current source	• gain, *PSRR* • output range, • noise, offset,	• $g_{DS3,4}$, • $V_{DSsat3,4}$, • $g_{m3,4}$, $v_{n3,4}$, $\sigma(\delta v_{G3,4})$
cascode load	• gain, • stability, • output range, • noise, mismatch, • PSRR	• $g_{m7,8}/g_{DS7,8}$, $g_{DS9,10}$, • $g_{m7,8}/(\Sigma C)_{CP}$, • $V_{DSsat7,8}$, $V_{DSsat9,10}$, • $g_{m9,10}$, $v_{n9,10}$, $\sigma(\delta v_{G9,10})$ • $g_{DS9,10}$
folded pair voltage bias	• $IF_{5,6}$	• V_{BIAS1}
cascode voltage bias	• $IF_{7,8}$	• V_{BIAS2}
differential pair current mirror	• *CMRR*, • common-mode input range	• g_{DSB1}, • $V_{DSsatB1}$
main current bias	• IF_{B1}, W_{B1}, L_{B1} • $IF_{3,4}$, $W_{3,4}$, $L_{3,4}$	• NMOS transistors matched with MB1, • PMOS transistors matched with M3,4

5.2.5 Procedural design sequence

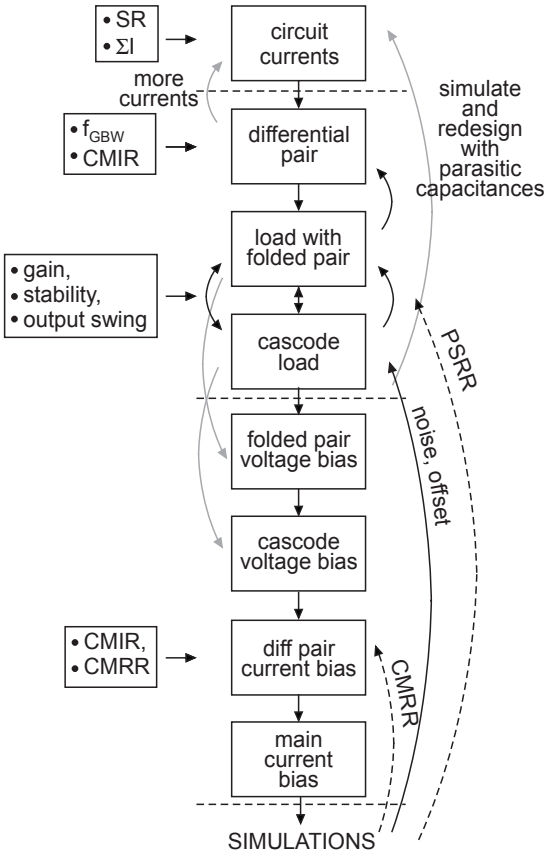


Figure 5.11 Folded cascode OTA procedural design sequence

Circuit currents - The differential pair tail current is determined according to the speed or the slew-rate specification

$$I_1 \geq \frac{SR \cdot (\Sigma C)_{OUT}}{2}. \tag{5.25}$$

The current I_2 is then chosen to be in range

$$1.2I_1 \leq I_2 \leq 2I_1 \tag{5.26}$$

which ensures that the folded-cascode pair transistors will not be turned off even when there is a large signal step at the input and all of the differential pair tail current is balanced on one side. Finally,

$$I_3 = I_1 + I_2 . \tag{5.27}$$

The other design specification that must be respected is the total current consumption. If all bias currents are equal it follows that

$$I_3 \leq \frac{\Sigma I - 5I_{BIAS}}{2}. \tag{5.28}$$

Differential pair - The inversion factor of differential pair transistors is usually determined according to the required f_{GBW} as

$$g_{m1,2} \geq 2\pi C_L \cdot f_{GBW} \rightarrow \frac{g_{m1,2}}{I_1} = \left(\frac{g_m}{I_{Dsat}}\right)_{1,2} \rightarrow IF_{1,2} . \tag{5.29}$$

However, if the common-mode input range specification has higher priority, the maximal value of the inversion factor is imposed by the acceptable gate-source voltage headroom $V_{GS1,2}$.

The transistor length $L_{1,2}$ is chosen to fulfill the condition that the output conductance of the transistors M1,2 is equal to or smaller than the output conductance of transistors M3,4, that is

$$L_{1,2} \geq \frac{I_1}{I_3} \cdot \frac{U_{a,P}}{U_{a,N}} \cdot L_{3,4} . \tag{5.30}$$

This allows us to simplify further design steps approximating the $R_{OUT,up}$ as $\frac{n \cdot g_{m5,6}}{g_{DS5,6}} \cdot \frac{1}{g_{DS3,4}}$ or $\frac{n \cdot g_{m5,6}}{g_{DS5,6}} \cdot \frac{1}{2g_{DS3,4}}$.

The design of the load with folded pair and cascode load transistors usually demands optimization loops to adjust the transistor design parameters because of the "stability - output resistance" relations. If the designer decides to set the $R_{OUT,up} > R_{OUT,down}$, then the gain is approximated as

$$A_0 \approx \frac{g_{m1,2}}{g_{DS9,10}} \cdot \frac{n \cdot g_{m7,8}}{g_{DS7,8}} \tag{5.31}$$

and the cascode load has to be designed in the first place. Otherwise, the gain is approximated as

$$A_0 \approx \frac{g_{m1,2}}{g_{DS3,4}} \cdot \frac{n \cdot g_{m5,6}}{g_{DS5,6}} \tag{5.32}$$

and the load with folded pair has to be designed before the cascode load.

Load with folded pair - If this basic analog structure is designed before the cascode load, the inversion levels of transistors are set according to the output range specification

$$\begin{aligned} &V_{DSsat3,4} \rightarrow IF_{3,4}, \\ &V_{DSsat5,6} \rightarrow IF_{5,6}. \end{aligned} \tag{5.33}$$

The transistor lengths can be determined with regard to the gain, by either choosing

$$g_{DS3,4} = g_{DS5,6} = \sqrt{\frac{g_{m1,2} \cdot n \cdot g_{m5,6}}{A_0}} \rightarrow L_{3,4} = L_{5,6}, \tag{5.34}$$

or setting $L_{5,6}$ in the range (L_{MIN} - $2L_{MIN}$) to minimize the sum of parasitic capacitances, and then calculating

$$g_{DS3,4} = \frac{g_{m1,2} \cdot n \cdot g_{m5,6}}{A_0 \cdot g_{DS5,6}} \rightarrow L_{3,4}. \tag{5.35}$$

If noise or offset reduction is a high design priority, then the minimal inversion factor of the transistors M3,4 is imposed by the condition

$$g_{m3,4} < g_{m1,2}. \tag{5.36}$$

On the other hand, if the cascode load is already designed, the inversion level of transistors M5,6 is chosen to fulfill the stability requirements, that is

$$\begin{aligned} &IF_{5,6} < IF_{7,8}, \\ &n \cdot g_{m5,6} \geq 2\pi(\Sigma C)_{FP} \cdot k_{PM} \cdot f_{GBW}, \end{aligned} \tag{5.37}$$

whereas the inversion factor of the transistors M3,4 is calculated as in the previous case according to the acceptable saturation voltage $V_{DSsat3,4}$.

The transistor lengths $L_{5,6}$ are usually set in the range (L_{MIN} - $2L_{MIN}$) to minimize the sum of parasitic capacitances, whereas the transistor lengths $L_{3,4}$ are set to ensure

$$g_{DS3,4} \cdot g_{DS5,6} < g_{DS7,8} \cdot g_{DS9,10}. \tag{5.38}$$

Cascode load - If this basic analog structure is designed in the first place, the inversion levels of transistors are determined from the output range specification

$$\begin{aligned} &V_{DSsat7,8} \rightarrow IF_{7,8}, \\ &V_{DSsat9,10} \rightarrow IF_{9,10}. \end{aligned} \tag{5.39}$$

The transistor lengths can be determined with regard to the gain by setting

$$g_{DS7,8} = g_{DS9,10} = \sqrt{\frac{g_{m1,2} \cdot n \cdot g_{m7,8}}{A_0}} \rightarrow L_{7,8} = L_{9,10}. \tag{5.40}$$

If noise or offset reduction is a high design priority, then the minimal inversion factor of the transistors M9,10 is imposed by the condition

$$g_{m9,10} < g_{m1,2}. \tag{5.41}$$

If the folded pair and the current bias sources are already designed, the inversion level of transistors M7,8 is chosen with regard to the stability conditions, that is

$$\begin{aligned} &IF_{5,6} > IF_{7,8}, \\ &n \cdot g_{m7,8} \geq 2\pi(\Sigma C)_{CP} \cdot k_{PM} \cdot f_{GBW}, \end{aligned} \tag{5.42}$$

whereas the inversion factor of the transistors M9,10 is calculated according to the acceptable saturation voltage $V_{DSsat9,10}$.

The transistor lengths $L_{7,8}$ are usually set in the range (L_{MIN} - $2L_{MIN}$) to minimize the sum of parasitic capacitances, whereas the transistor lengths $L_{9,10}$ are set to ensure

$$g_{DS3,4} \cdot g_{DS5,6} > g_{DS7,8} \cdot g_{DS9,10}. \tag{5.43}$$

After the core of the amplifier is designed, the circuit can be simulated with ideal voltage and current bias sources to confirm the obtained gain and stability versus the specifications. If it is necessary, all previously described design steps can be repeated, but now taking into account the simulated parasitic capacitances and the possible degradation of the output conductances due to short transistor lengths.

Folded cascode voltage bias - The required bias voltage is determined as

$$V_{BIAS2} = V_{DSsat3,4} + V_{GS5,6}, \tag{5.44}$$

where the gate-source voltage $V_{GS5,6}$ is calculated from the inversion factor $IF_{5,6}$. The transistors are sized as shown in subsection 4.6.1.

Cascode voltage bias - In the same way as in the previous step, the required bias voltage is determined as

$$V_{BIAS1} = V_{DSsat9,10} + V_{GS7,8}, \tag{5.45}$$

where the gate-source voltage $V_{GS7,8}$ is calculated from the inversion factor $IF_{7,8}$. The transistors are sized as shown in subsection 4.6.1.

Differential pair current bias - The inversion factor of the transistor MB1 is calculated from the acceptable saturation voltage $V_{DSsatB1}$ that is determined according to the required common-mode input range.

The transistor length is set to achieve the output conductance g_{DSB1} value that ensures obtaining the required *CMRR*.

Main current bias - The transistors MN3,5 are matched with the differential current bias transistor MB1, while the cascode transistors MN4,6 are usually placed at the limit of weak inversion to obtain the minimal saturation voltage. The transistors MN1,2 represent the voltage bias of cascode transistors MN4,6, and are designed as shown in subsection 4.6.1.

The transistors MP1,3,5,7 are matched with the transistors M3,4, while the cascode transistors MP2,4,6,8 are usually placed at the limit of weak inversion to obtain the minimal saturation voltage. The resistance R_{BIAS} is determined as

$$R_{BIAS} = \frac{V_{GP2} - V_{GP1}}{I_{BIAS}}, \tag{5.46}$$

where the gate voltages V_{G1} and V_{G2} are calculated according to chosen inversion levels.

After the complete circuit has been sized at the transistor level, it is simulated to verify the achieved performances. At this point, fine-tuning is usually required to improve noise, offset or the rejection ratios. The backward arrows in Figure 5.11 indicate which blocks need to be redesigned (or their parameters adjusted) to improve these circuit-level parameters.

5.3 Procedural design scenario for a fully-differential folded cascode OTA

Figure 5.12 depicts a fully-differential variant of the folded cascode amplifier and its **C**ommon-**M**ode **F**eed**B**ack (CMFB) amplifier. The CMFB amplifier is a part of the CMFB loop, indispensable for any fully-differential structure. It controls the DC level of the voltages at the output terminals, i.e. the common-mode output voltage of the amplifier $V_{CM,OUT} = (V_{OUT+} + V_{OUT-})/2$, since the differential mode has no control over it.

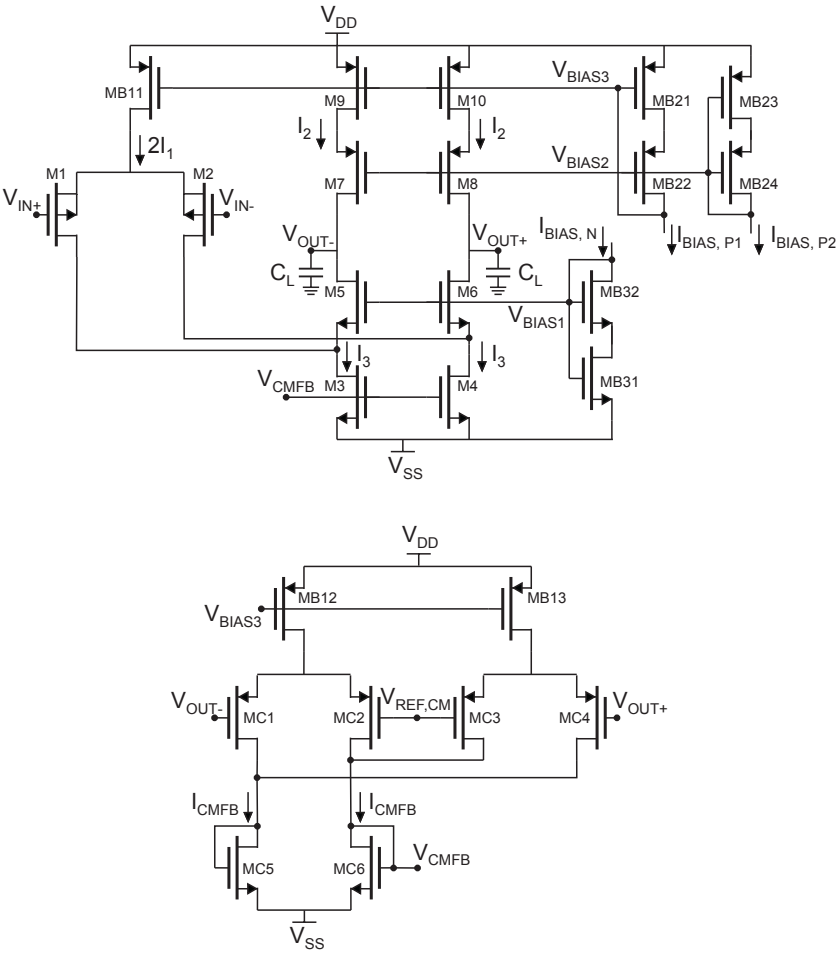


Figure 5.12 Fully-differential folded cascode OTA and its CMFB circuit

The proposed topology of the CMFB amplifier is just one possible realization, and there exist numerous other topologies described in the literature [3], [4]. It is chosen here for illustration purposes, and also because it is used in the practical design example in chapter 8. The common-mode output voltage is sensed thanks to two-differential pairs, MC1,2 and MC3,4, and compared to the reference voltage $V_{CM,REF}$. The differential pair output terminals are connected in such a way that the CMFB amplifier output voltage, denoted here as V_{CMFB}, becomes a function of the difference between the common-mode output voltage and the reference voltage. The voltage V_{CMFB} controls the current sources M3,4 of the main amplifier, and in this way closes the negative feedback loop that minimizes the difference between the common-mode output voltage and the reference voltage. If this voltage difference increases, V_{CMFB} also increases, and consequently the bias currents I_3. As a result, the voltage drop across the cascode load (M7,8 and M9,10) increases, the DC level of both output voltages decreases, and thus the common-mode output voltage.

In this example, all bias currents are matched, that is the bias current of the differential pair I_1, the branch current I_2 of the main amplifier and the CMFB bias current I_{CMFB}. In this way, if there exists any cause of mismatch (for example a temperature gradient), all circuit currents are affected in the same way. The main current bias, shown in Figure 5.13, provides only the bias currents required by the corresponding voltage bias structures. However, in some other design examples these current sources can be also implemented separately or matched with the main current bias circuit.

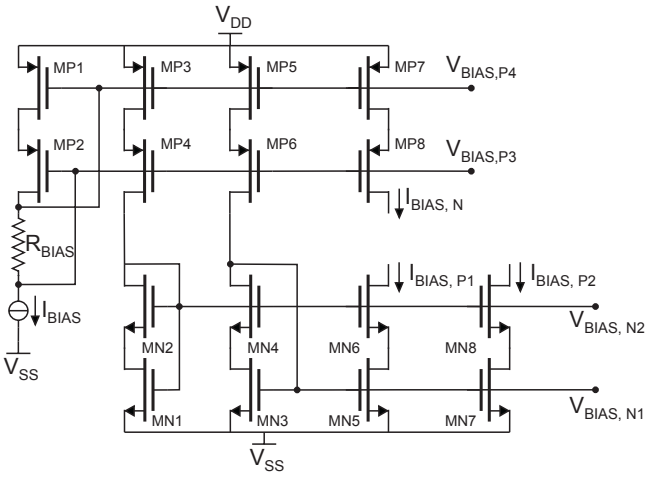


Figure 5.13 Current biasing circuit of a fully-differential folded cascode OTA

5.3.1 Circuit-level design parameters

Circuit-level design parameters of the main amplifier and the CMFB amplifier are summarized in Table 5.6.

Rejection ratios - The input common-mode signal gain is again approximated by (5.16), which results in

$$CMRR = k \cdot \frac{g_{DS3,4}}{g_{DSB11}} \cdot \frac{g_{mc1,2,3,4}}{g_{m5,6}}. \tag{5.47}$$

If the *CMRR* is not the first design priority, the transconductances of the transistors MC1,2,3,4, and M5,6 and the output conductance of the transistors M3,4 are usually imposed by the gain and stability conditions. The key design parameter is again the output conductance g_{DSB11}.

Concerning the *PSRR*, the key parameters are again the output conductances of transistors M3,4 and M9,10.

Table 5.6 Fully-differential folded cascode OTA circuit-level design parameters

static current consumption	$\Sigma I = 2I_3 + I_{BIAS,N} + I_{BIAS,P1} + I_{BIAS,P2} + 2I_{CMFB} + 3I_{BIAS}$
differential gain, output resistance	$A_0 = g_{m1,2} \cdot R_{OUT},\ R_{OUT} = R_{OUT,down} \parallel R_{OUT,up}$ $R_{OUT,up} = \frac{n \cdot g_{m7,8}}{g_{DS7,8}} \cdot \frac{1}{g_{DS9,10}}$ $R_{OUT,down} = \frac{n \cdot g_{m5,6}}{g_{DS5,6}} \cdot \frac{1}{(g_{DS1,2} + g_{DS3,4})}$
CMFB loop gain	$A_{0,CMFB} = k \cdot g_{mC1,2,3,4} \cdot R_{OUT}, k = \frac{I_3}{I_{CMFB}}$
common-mode input range of the main amplifier	$V_{IN,CM,MAX} = V_{DD} - \lvert V_{DSsatB11} \rvert - \lvert V_{GS1,2} \rvert,$ $V_{IN,CM,MIN} = V_{SS} + V_{DSsat3,4} + \lvert V_{DSsat1,2} \rvert - \lvert V_{GS1,2} \rvert$
output range	$V_{OUT,MAX} = V_{DD} - \lvert V_{DSsat7,8} \rvert - \lvert V_{DSsat9,10} \rvert$ $V_{OUT,MIN} = V_{SS} + V_{DSsat3,4} + V_{DSsat5,6}$

Table 5.6 (continued) Fully-differential folded cascode OTA circuit-level design parameters

differential input range of the CMFB amplifier	$V_{OUT,MAX} = V_{DD} - \lvert V_{DSsatB12,13}\rvert - \lvert V_{GSC1,2,3,4}\rvert$, $V_{OUT,MIN} = V_{SS} + V_{GSC6} + \lvert V_{DSsatC1,2,3,4}\rvert - \lvert V_{GSC1,2,3,4}\rvert$
input referred equivalent noise	$v_n^2 = 2v_{n1,2}^2 + 2\left(\frac{g_{m3,4}}{g_{m1,2}}\right)^2 v_{n3,4}^2 + 2\left(\frac{g_{m9,10}}{g_{m1,2}}\right)^2 v_{n9,10}^2$
input referred equivalent offset	$v_{off}^2 = \sigma(\delta V_{G1,2})^2 + \left(\frac{g_{m3,4}}{g_{m1,2}}\right)^2 \sigma(\delta V_{G3,4})^2 + \left(\frac{g_{m9,10}}{g_{m1,2}}\right)^2 \sigma(\delta V_{G9,10})^2$
slew-rate	$SR = \frac{I_1}{C_L}$

5.3.2 Frequency analysis

In the case of a fully-differential structure, the stabilities of two feedback loops have to be analyzed: the main amplifier feedback loop and the CMFB loop. The main amplifier will be treated here in the most unfavorable case, that is in the unity-gain negative configuration.

Main feedback loop - Using the equivalent half-circuit from Figure 5.14, the dominant and non-dominant pole frequencies of the main feedback loop are approximated and given in Table 5.7. The gain bandwidth frequency of the open unity-gain feedback loop is calculated as

$$f_{GBW,OTA} = A_0 \cdot f_{dp,OTA} = \frac{g_{m1,2}}{2\pi(\Sigma C)_{OUT}}, \tag{5.48}$$

and the phase margin

$$PM = 180° - atan\left(\frac{f_{GBW,OTA}}{f_{dp,OTA}}\right) - atan\left(\frac{f_{GBW,OTA}}{f_{ndp,OTA}}\right). \tag{5.49}$$

Accordingly, the stability condition is

$$f_{ndp,OTA} = k_{PM} \cdot f_{GBW,OTA}, \tag{5.50}$$

where k_{PM} is determined from the phase margin specification.

At this point, it is important to note that the non-dominant pole frequency strongly depends on the transconductance and the parasitic capacitances of transistors M5,6. On the other hand, the intrinsic gain of the same transistors determines the output resistance $R_{OUT,down}$. Therefore, it is convenient to approximate the output resistance as

$$R_{OUT} \approx R_{OUT\ down}, \tag{5.51}$$

which imposes that

$$R_{OUT,\,up} >> R_{OUT,\,down}. \tag{5.52}$$

This simplifies the design procedure since the gain-stability trade-off is found simply by sizing the transistors M5,6, and there are no additional conditions as in the case of a single-ended amplifier.

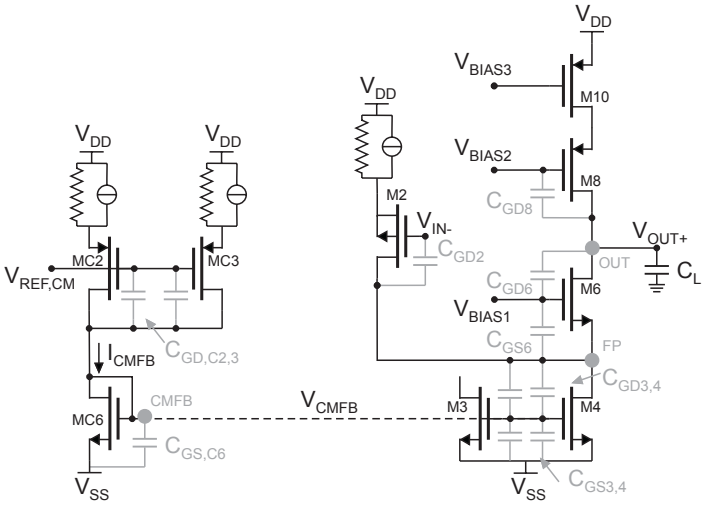


Figure 5.14 Equivalent half-circuit used for the frequency analysis of the main amplifier and the CMFB amplifier feedback loop

Table 5.7 Approximations for pole frequencies in the transfer function of the fully-differential folded cascode OTA

dominant pole $f_{dp,OTA}$	$\frac{1}{2\pi R_{OUT}(\Sigma C)_{OUT}}$, $(\Sigma C)_{OUT} = C_L + C_{GD5,6} + C_{GD7,8}$
non-dominant pole $f_{ndp,OTA}$	$\frac{n \cdot g_{m5,6}}{2\pi(\Sigma C)_{FP}}$, $(\Sigma C)_{FP} = C_{GS5,6} + C_{GD3,4} + C_{GD1,2}$

CMFB loop - The CMFB loop inherits the dominant and one non-dominant pole from the main feedback loop since it uses one part of the main amplifier gain path. However, there is an additional pole at the output node of the CMFB amplifier, indicated as CMFB in Figure 5.14. According to the approximations for the pole frequencies given in Table 5.8, it is obvious that this non-dominant pole is closer to the origin as $g_{mC6} < g_{m5,6}$ (because even if $IF_{C6} = IF_{5,6}$, the drain current $I_{DsatC6} = I_{CMFB}$ is smaller than the drain current $I_{Dsat5,6} = I_2$) and $(\Sigma C)_{CMFB} > (\Sigma C)_{FP}$.

The gain bandwidth frequency of the CMFB loop is given by

$$f_{GBW,CMFB} = A_{0,CMFB} \cdot f_{dp} = \frac{k \cdot g_{mC1,2,3,4}}{2\pi(\Sigma C)_{OUT}}, \tag{5.53}$$

and the phase margin is

$$PM_{CMFB} = 180° - atan\left(\frac{f_{GBW,CMFB}}{f_{dp,CMFB}}\right) - atan\left(\frac{f_{GGBW,CMFB}}{f_{ndp1,CMFB}}\right) - atan\left(\frac{f_{GBW,CMFB}}{f_{ndp2,CMB}}\right). \tag{5.54}$$

Accordingly, the stability condition is

$$f_{ndp1,CMFB} = k_{PM} \cdot f_{GBW,CMFB}, \tag{5.55}$$

where k_{PM} is determined from the phase margin specification.

Table 5.8 Approximations for pole frequencies in the CMFB loop transfer function

dominant pole $f_{dp,CMFB}=f_{dp,OTA}$	$\frac{1}{2\pi R_{OUT}(\Sigma C)_{OUT}}$ $(\Sigma C)_{OUT} = C_L + C_{GD5,6} + C_{GD7,8}$
non-dominant pole $f_{ndp1,CMFB}$	$\frac{g_{mC6}}{2\pi(\Sigma C)_{CMFB}}$ $(\Sigma C)_{CMFB} = = C_{GSC6} + 2C_{GS3,4} + g_{m3,4}R_{OUT}C_{GD4}$
non-dominant pole $f_{ndp2,CMFB}=f_{ndp,OTA}$	$\frac{n \cdot g_{m5,6}}{2\pi(\Sigma C)_{FP}}$, $(\Sigma C)_{FP} = C_{GS5,6} + C_{GD3,4} + C_{GD1,2}$

5.3.3 Circuit partitioning

The fully-differential folded cascode OTA, its CMFB amplifier and main current bias circuit are divided into basic analog structures in Figure 5.15 and Figure 5.16 as follows:

- differential pair,
- load with folded pair = bias mirror M3,4 + folded pair,
- cascode load,
- folded pair voltage bias,
- cascode voltage bias,
- differential pair current bias,
- CMFB differential pairs,
- CMFB diodes,
- CMFB current bias,
- main current bias.

In the same way as in the single-ended case, to facilitate gain-stability compromises the folded pair transistors are designed with the current bias transistors M3,4 as a load structure. The differential pair current bias and the common-mode feedback current bias are matched with the cascode load

transistors that behave as a load and a current bias at the same time. Finally, the current I_3 bias transistors M3,4 and the CMFB diode MC6 are also matched, since they create a simple current mirror.

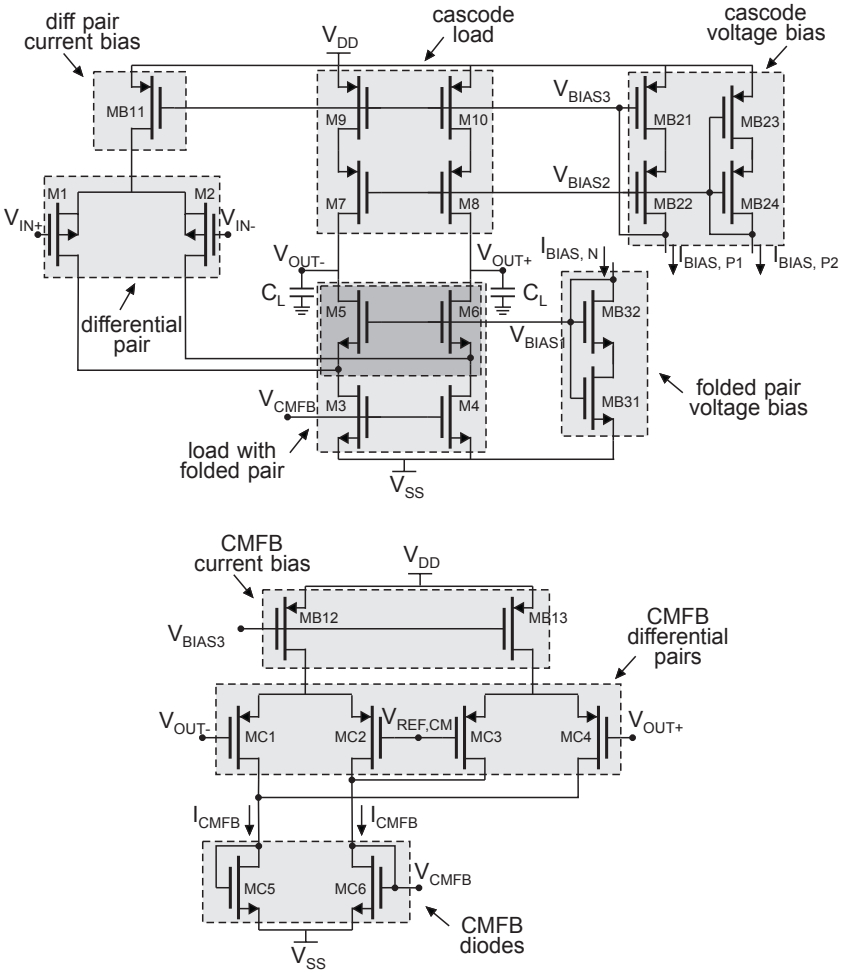


Figure 5.15 Fully-differential folded cascode OTA and its CMFB amplifier partitioned into basic analog structures

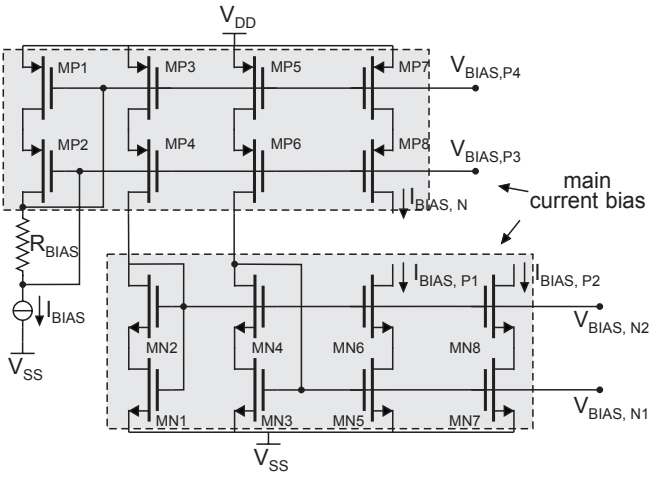


Figure 5.16 Main current bias of fully-differential folded cascode OTA partitioned into basic analog structures

5.3.4 Derivation of the specifications

The specifications for the design of basic analog blocks are derived in the same way as in the previous example. The design parameters of the basic analog structures are listed in Table 5.9, whereas the circuit-level design parameter sets that are affected by the design parameters of the corresponding basic analog structure are indicated in Figure 5.17.

Table 5.9 Table for derivation of the specifications for fully-differential folded cascode OTA

basic analog structure	circuit-level or other basic analog struct. parameters	basic analog structure parameters
differential pair	• gain, • gain bandwidth, • common-mode input range, • noise, offset	• $g_{m1,2}/g_{DS1,2}$, • $g_{m1,2}$, • $V_{GS1,2}$, • $g_{m1,}$ $v_{n1,2,}$ $\sigma(\delta v_{G1,2})$
load with folded pair - folded pair	• gain, • stability, • output range	• $g_{m5,6}/g_{DS5,6}$, • $g_{m5,6}/(\Sigma C)_{FP}$, • $V_{DSsat5,6}$,

Table 5.9 (continued) Table for derivation of the specifications for fully-differential folded cascode OTA

basic analog structure	circuit-level or other basic analog struct. parameters	basic analog structure parameters
load with folded pair - current source	• gain, *PSRR* • output range, • noise, offset, • CMFB stability	• $g_{DS3,4}$, • $V_{DSsat3,4}$, • $g_{m3,4}$, $v_{n3,4}$, $\sigma(\delta v_{G3,4})$, • $C_{GS3,4}$
cascode load	• R_{OUT}, • output range, • noise, mismatch, • PSRR	• $g_{m7,8}/g_{DS7,8}$, $g_{DS9,10}$, • $V_{DSsat7,8}$, $V_{DSsat9,10}$, • $g_{m9,10}$, $v_{n9,10}$, $\sigma(\delta v_{G9,10})$ • $g_{DS9,10}$
folded pair voltage bias	• $IF_{5,6}$	• V_{BIAS1}
cascode voltage bias	• $IF_{7,8}$	• V_{BIAS2}
differential pair current mirror	• *CMRR*, • common-mode input range	• g_{DSB11}, • $V_{DSsatB11}$, • matched with cascode load
CMFB differential pair	• CMFB gain, CMFB stability, • *CMRR*, *PSRR*, • output range (CMFB differential input range)	• $g_{mC1,2,3,4}$, • A_{CMFB}, $g_{mC1,2,3,4}$, • $V_{GSC1,2,3,4}$
CMFB diodes	• CMFB stability	• g_{mC6}, $(\Sigma C)_{CMFB}$ • matched with current sources M3,4
CMFB current bias	• output range (CMFB differential input range)	• $V_{DSsatB12,13}$ • matched with cascode load
main current bias	• no particular design requirements	

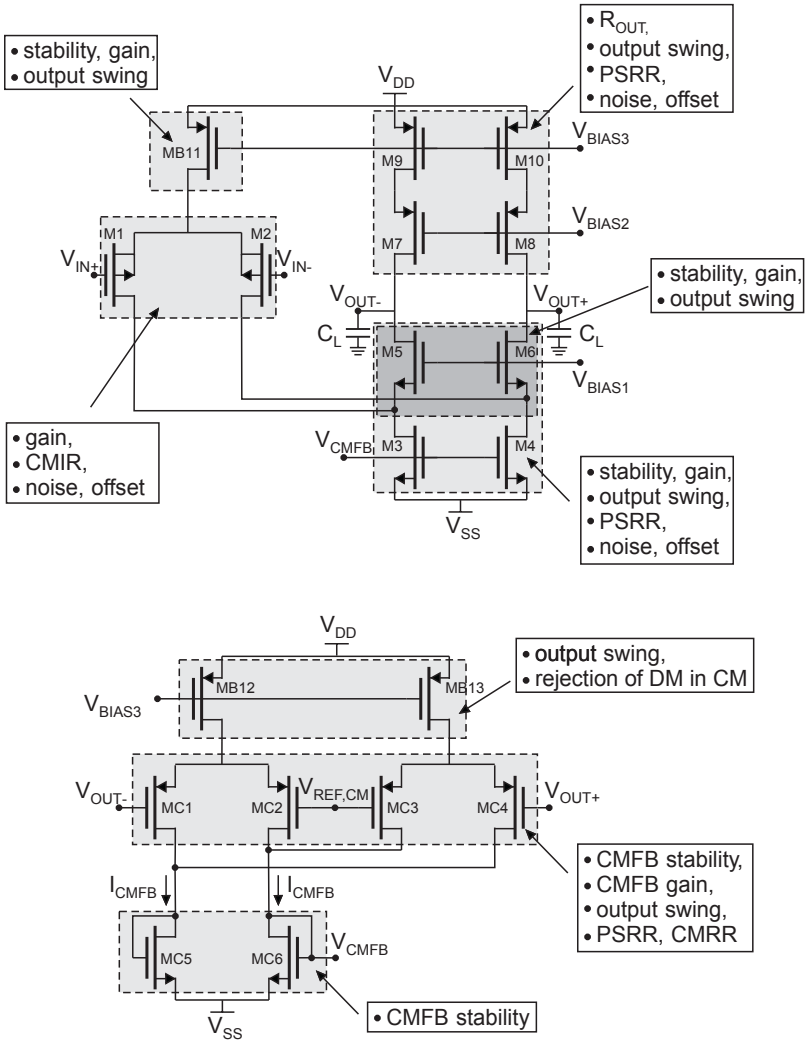


Figure 5.17 Fully-differential folded OTA circuit-level parameter sets that are used for the derivation of basic analog structures' specifications

5.3.5 Procedural design sequence

The procedural design sequence together with the indications of the possible optimization loops is shown in Figure 5.18.

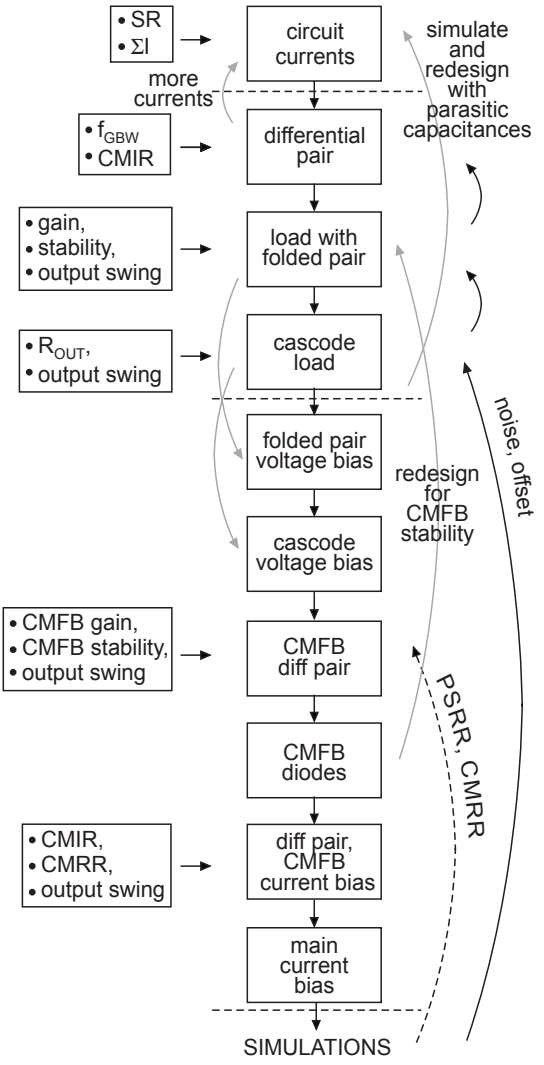


Figure 5.18 Fully-differential folded cascode OTA procedural design sequence

Circuit currents - The main amplifier currents are determined in a similar way to those in the single-ended case, that is

$$I_1 \geq SR \cdot (\Sigma C)_{OUT}, \tag{5.56}$$

$$1.2I_1 \leq I_2 \leq 2I_1, \tag{5.57}$$

$$I_3 = I_1 + I_2. \tag{5.58}$$

The CMFB bias current I_{CMFB} can be set initially in the range (I_{BIAS} - $2I_{BIAS}$). It can be adjusted later if a compromise between CMFB gain and stability needs to be found. This is usually achieved by choosing the appropriate value for the current ratio k, defined as

$$k = I_3/I_{CMFB}, \tag{5.59}$$

since the transistors M3,4 and the CMFB diode MC6 create a simple current mirror. The design specification that also must be respected is the total current consumption. If we assume that all bias currents are equal, it follows that

$$I_3 \leq \frac{\Sigma I - 5I_{BIAS}}{2(1 + 1/k)}. \tag{5.60}$$

Differential pair - Differential pair transistors are sized in exactly the same way as in subsection 5.2.5.

On the other hand, according to the frequency analysis, an important design condition is $R_{OUT,up} >> R_{OUT,down}$. To achieve this, it is usually sufficient to set $R_{OUT,up} = 10R_{OUT,down}$. This allows us to approximate the gain of the folded cascode OTA as

$$A_0 \approx \frac{g_{m1,2}}{g_{DS3,4}} \cdot \frac{n \cdot g_{m5,6}}{g_{DS5,6}}. \tag{5.61}$$

Load with folded pair - The inversion levels of transistors are set according to the output range specification

$$V_{DSsat3,4} \rightarrow IF_{3,4}, \tag{5.62}$$
$$V_{DSsat5,6} \rightarrow IF_{5,6}.$$

The transistor lengths $L_{5,6}$ can be either set in the range (L_{MIN} - $2L_{MIN}$) to minimize the sum of parasitic capacitances or calculated from the stability requirement

$$n \cdot g_{m5,6} \geq 2\pi(\Sigma C)_{FP} \cdot k_{PM} \cdot f_{GBW}, \tag{5.63}$$

where k_{PM} is determined from the phase margin specification. In the latter case, a good design practice is to overestimate the sum of parasitic capacitances in the folded point by

$$(\Sigma C)_{FP} \approx 3 \cdot COX \cdot \frac{I_2}{2 \cdot n \cdot KP \cdot IF_{5,6} \cdot V_t^2} \cdot L_{5,6}^2 \tag{5.64}$$

and determine the maximal acceptable $L_{5,6}$ using the aforementioned condition (5.63).

The transistor lengths $L_{3,4}$ are determined with regard to the gain specification from which the output conductance is calculated as

$$g_{DS3,4} = \sqrt{\frac{g_{m1,2} \cdot n \cdot g_{m5,6}}{A_0 \cdot g_{DS5,6}}} \rightarrow L_{3,4}. \tag{5.65}$$

If noise or offset reduction is a high design priority, then the minimal inversion factor of the transistors M3,4 is imposed by the condition

$$g_{m3,4} < g_{m1,2}. \tag{5.66}$$

Cascode load - The inversion levels of transistors are determined from the output range specification

$$\begin{aligned} V_{DSsat7,8} &\rightarrow IF_{7,8}, \\ V_{DSsat9,10} &\rightarrow IF_{9,10}. \end{aligned} \tag{5.67}$$

The transistor lengths can be determined with regard to the output resistance, by setting

$$g_{DS7,8} = g_{DS9,10} = \sqrt{\frac{n \cdot g_{m7,8}}{10 \cdot R_{OUT,down}}} \rightarrow L_{7,8} = L_{9,10}. \tag{5.68}$$

Again, if noise or offset reduction is a high design priority, then the minimal inversion factor of the transistors M9,10 is imposed by the condition

$$g_{m9,10} < g_{m1,2}. \tag{5.69}$$

After the core of the main amplifier is designed, the circuit can be simulated with ideal voltage and current bias sources, and an ideal (or behavioral) CMFB model to confirm the obtained gain and stability versus the given specifications. The previous design steps can be repeated, but now taking into consideration the real (not estimated or overestimated values) of parasitic capacitances and the possible degradation of the output conductances due to short transistor lengths.

Folded cascode voltage bias - The required bias voltage is determined as

$$V_{BIAS1} = V_{DSsat3,4} + V_{GS5,6}, \tag{5.70}$$

where the gate-source voltage $V_{GS5,6}$ is calculated from the inversion factor $IF_{5,6}$. The transistors are sized as shown in subsection 4.6.1.

Cascode voltage bias - The transistors M21 and M22 are matched with the transistors M9,10 and M7,8, respectively. This means that the transistor lengths are the same, while the transistors' widths are set according to the ratio of drain currents I_2/I_{BIAS}.

The required cascode bias voltage is determined as

$$V_{BIAS2} = V_{DSsat9,10} + V_{GS7,8}, \tag{5.71}$$

where the gate-source voltage $V_{GS7,8}$ is calculated from the inversion factor $IF_{7,8}$. The transistors M23 and M24 are then sized as shown in subsection 4.6.1.

Differential pair current bias - In this example, the transistor MB11 is simply matched with the transistors M9,10.

However, if the common-mode input range or *CMRR* specification cannot be achieved in this way, the differential pair current bias needs to be redesigned separately. In this case, the inversion factor of transistor MB11 is calculated from the acceptable saturation voltage $V_{DSsatB11}$ that is determined according to the required common-mode input range. The transistor length is set to achieve the output conductance g_{DSB11} that ensures obtaining the required *CMRR*.

CMFB differential pair - The best approach is to set the inversion factor of the CMFB differential pair transistors to a value determined according to the required CMFB loop gain, namely

$$g_{mC1,2,3,4} \geq \frac{A_{0,CMFB}}{k \cdot R_{OUT,OTA}} \rightarrow \frac{g_{mC1,2,3,4}}{I_{CMFB}} = \left(\frac{g_m}{I_{Dsat}}\right)_{C1,2,3,4} \rightarrow IF_{C1,2,3,4}. \tag{5.72}$$

However, if the output range of the OTA (that is the differential input range of the CMFB amplifier) has higher priority, the maximal value of the inversion factor is imposed by the acceptable gate-source voltage headroom $V_{GSC1,2,3,4}$.

CMFB diodes - The transistors MC5,6 are matched with the transistors M3,4 that have been sized previously.

If the specified phase margin is not achieved, the first step is to adjust the current ratio k (and eventually the transconductance of the CMFB differential pair) and find the "CMFB gain-stability" compromise. If this is not the case, then the transistors M3,4 have to be redesigned to minimize the sum of parasitic capacitances and move the non-dominant pole to the higher frequencies. This optimization path is indicated in Figure 5.18.

CMFB current bias - In this example, the transistors MB12,13 are matched with the transistors M9,10.

However, if the output range of the OTA (that is the differential input range of the CMFB amplifier) has higher priority, then this current bias needs to be redesigned separately. In this case, the inversion factor of the transistors MB12,13 is calculated from the acceptable saturation voltage $V_{DSsatB12,13}$.

Main current bias - Since there are no special design requirements, both cascode current mirrors are designed to minimize the saturation voltages, and thus the required voltage headroom. The resistance R_{BIAS} is determined as in the single-ended case as

$$R_{BIAS} = \frac{V_{GP2} - V_{GP1}}{I_{BIAS}}, \tag{5.73}$$

where the gate voltages V_{G1} and V_{G2} are calculated according to chosen inversion levels.

After the complete circuit has been sized at the transistor level, it is simulated to verify the achieved performances. At this point, fine-tuning is usually required to improve noise, offset or rejection ratios. The backward arrows in Figure 5.18 indicate which blocks need to be redesigned (or their parameters adjusted) to improve these circuit-level parameters.

5.4 Procedural design scenario for a Miller operational amplifier

The Miller operational amplifier is a classical, widely used analog cell. Figure 5.19 depicts a variant with the p-channel input differential pair and all bias current sources matched. Since it is a two-stage amplifier, the compensation capacitance C_C is usually mandatory for stability.

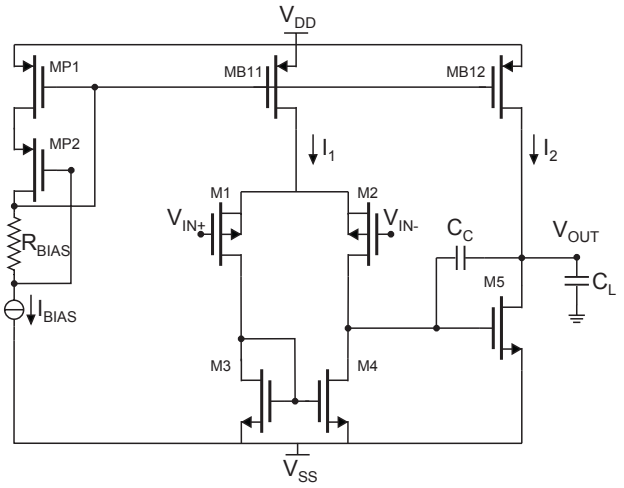


Figure 5.19 Miller operational amplifier

5.4.1 Circuit-level design parameters

Similarly to the two previous examples, the circuit-level design parameters are calculated, and then summarized in Table 5.10.

Rejection ratios - The gain of the input common-mode signal is approximated by

$$A_C = \frac{1/g_{DS3,4}}{1/g_{DSB11}} \cdot g_{m5} \cdot R_{OUT2}, \tag{5.74}$$

which gives

$$CMRR \approx \frac{1}{2} \cdot \frac{g_{m1,2}}{g_{DSB11}} , \tag{5.75}$$

if it is assumed that $g_{DS1,2} \approx g_{DS3,4}$. As discussed previously, the key design parameter is the output conductance of the current bias source MB11.

Concerning the *PSRR*, for the transistors M3,4, and MB11,12 it is considered that they have a small-signal perturbation source not only in the source, but also in the gate terminal. The resulting gain contributions are approximately the same, but with opposite sign, and thus ideally canceled. On the other hand, due to transistor M5, the noise in the negative supply is amplified to the output by a factor of $g_{m5} \cdot R_{OUT2}$, which gives

Table 5.10 Miller operational amplifier circuit-level design parameters

static current consumption	$\Sigma I = I_1 + I_2 + I_{BIAS}$
gain, output resistance	$A_0 = A_{0,1} \cdot A_{0,2}$ $A_{0,1} = g_{m1,2} \cdot R_{OUT1},\ R_{OUT1} = \frac{1}{g_{DS1,2} + g_{DS3,4}}$ $A_{0,2} = g_{m5} \cdot R_{OUT2},\ R_{OUT2} = \frac{1}{g_{DS5} + g_{DSB12}}$
common-mode input range	$V_{IN,CM,MAX} = V_{DD} - \lvert V_{DSsatB11}\rvert - \lvert V_{GS1,2}\rvert$ $V_{IN,CM,MIN} = V_{SS} + V_{GS3,4} + \lvert V_{DSsat1,2}\rvert - \lvert V_{GS1,2}\rvert$
output range	$V_{OUT,MAX} = V_{DD} - \lvert V_{DSsatB12}\rvert$ $V_{OUT,MIN} = V_{SS} + V_{DSsat5}$
input referred equivalent noise	$v_n^2 = 2v_{n1,2}^2 + 2\left(\frac{g_{m3,4}}{g_{m1,2}}\right)^2 v_{n3,4}^2 + \frac{v_{n5}^2 + \left(\frac{g_{mB11}}{g_{m5}}\right)^2 v_{nB11}^2}{A_{0,1}^2}$
input referred equivalent offset	$v_{off}^2 = \sigma(\delta V_{G1,2})^2 + \left(\frac{g_{m3,4}}{g_{m1,2}}\right)^2 \sigma(\delta V_{G3,4})^2$
slew-rate	$SR = min\left\{\frac{I_1}{C_C}, \frac{I_2}{C_L}\right\}$

$$PSRR = \frac{A_0}{g_{m5} \cdot R_{OUT2}} = A_{0,1}. \tag{5.76}$$

5.4.2 Frequency analysis

The Miller amplifier structure has three poles: in the gate of transistors M3,4, in the output node of the first stage and in the output node of the second stage. The two poles in the output nodes of the first and the second stage are split thanks to the compensation capacitance, but at the cost of the additional positive zero. The approximations are listed in Table 5.11 without entering into

calculation details, since this topic is extensively treated in subsection 7.2.1. The only important point to note is that the approximations are valid only if the compensation and the load capacitances are much larger than the parasitic capacitances of the transistors.

As in the previous examples, the gain bandwidth frequency is approximated as

$$f_{GBW} = A_{0,1} \cdot f_{dp} = \frac{g_{m1}}{2\pi C_C}, \tag{5.77}$$

and the phase margin is calculated as

$$PM = 180° - atan\left(\frac{f_{dp}}{f_{GBW}}\right) - \sum_{i=1}^{2} atan\left(\frac{f_{ndpi}}{f_{GBW}}\right) - atan\left(\frac{f_z}{f_{GBW}}\right). \tag{5.78}$$

Usually $C_C < C_L$, and then obviously we have $f_z > f_{ndp1}$. As a result, the positive zero does not need to be canceled (for example by using the nulling resistance [1]) since the negative phase shift does not degrade the phase margin in this case.

Table 5.11 Approximations for pole and zero frequencies in the Miller operational amplifier open unity-gain transfer function (valid only if C_C and C_L are much larger than the parasitic capacitances of the transistors)

dominant pole f_{dp}	$\frac{1}{2\pi R_{OUT1} \cdot A_{0,2} C_C}$
non-dominant pole f_{ndp1}	$\frac{g_{m5}}{2\pi C_L}$
positive zero f_z	$\frac{g_{m5}}{2\pi C_C}$
non-dominant pole f_{ndp2}	$\frac{g_{m3,4}}{2\pi(\Sigma C)_{G3,4}}$ $(\Sigma C)_{G3,4} = C_{GD3,4} + g_{m3,4} R_{OUT1} C_{GD4}$

On the other hand, since the transistors M5 and M3,4 usually need to be matched, that is $IF_{3,4} = IF_5$, to avoid electrical mismatch at the output of the first stage, and since the bias current of the second stage is several times lager than the differential pair tail current (that is $I_2 = mI_1$), we can write

$$\frac{g_{m3,4}}{I_1/2} = \frac{g_{m5}}{I_2} \rightarrow g_{m5} = 2m \cdot g_{m3,4}. \tag{5.79}$$

Hence, taking into account the assumption that the load capacitance is much larger than the parasitic capacitances, it follows that

$$f_{ndp1} = \frac{g_{m5}}{2\pi C_L} = \frac{2m \cdot g_{m3,4}}{2\pi C_L} < \frac{g_{m3,4}}{2\pi(\Sigma C)_{G3,4}} = f_{ndp2}. \tag{5.80}$$

The stability condition is thus defined with regard to f_{ndp1} as

$$f_{np1} > k_{PM} \cdot f_{GBW}, \tag{5.81}$$

where k_{PM} is determined from the phase margin specification. Further, it can be rewritten in the following way

$$\frac{g_{m5}/I_2}{2\pi C_L} > k_{PM} \cdot \frac{g_{m1,2}/mI_1}{2\pi C_C}. \tag{5.82}$$

This gives the relation between the inversion factors of the "transconductance" transistors, the compensation capacitance and the load capacitance and can be used if it is required to trade off these design parameters for the stability.

5.4.3 Circuit partitioning

The Miller operational amplifier is partitioned into basic analog structures in Figure 5.20. It consists of the following building blocks:

- differential pair,
- active load,
- differential pair current bias,
- common source,
- common source load/current bias,
- main (current/voltage) bias.

In this example all current bias transistors are matched, and they create a simple current mirror with the transistors MP1,2.

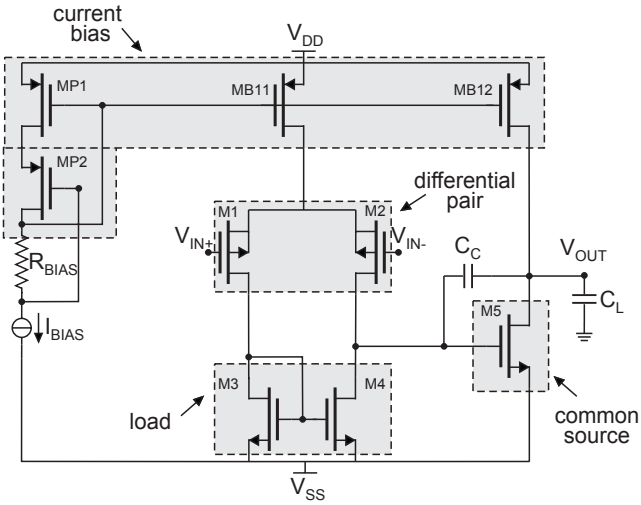


Figure 5.20 Miller operational amplifier partitioned into basic analog structures

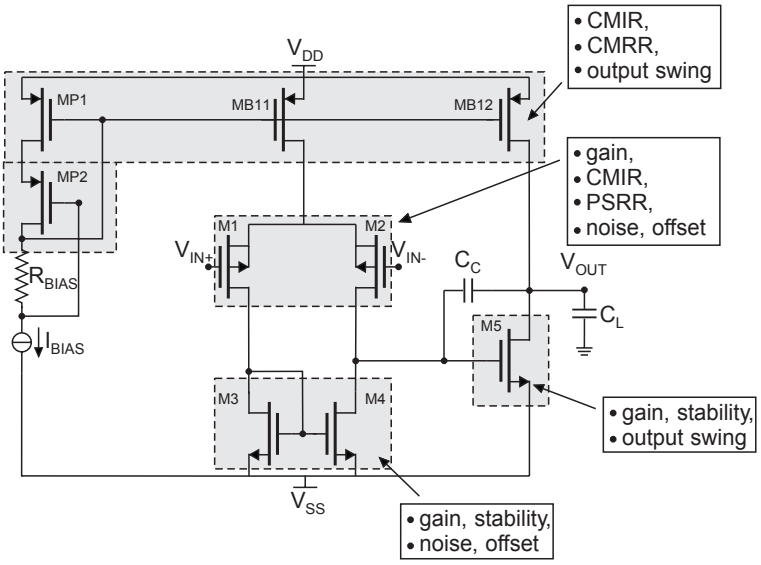


Figure 5.21 Miller operational amplifier circuit-level parameter sets that are used for the derivation of the basic analog structures' specifications

5.4.4 Derivation of the specifications

The partitioned amplifier is redrawn in Figure 5.21 with the appropriate circuit-level design parameter sets indicated. The basic analog structures specifications are derived in Table 5.12.

The additional design conditions are:

- the transistors M3,4 and M5 are matched to avoid electrical offset at the output of the first stage,
- the current bias transistors M11, M12 are matched with the transistor MP1,
- $g_{DS1,2} \geq g_{DS3,4}$, to simplify the gain of the first stage as g_{m1}/g_{DS1},
- $g_{DS5} \geq g_{DSB12}$, to simplify the gain of the second stage as g_{m5}/g_{DS5}.

Table 5.12 Table for derivation of the specifications of the Miller operational amplifier

basic analog structure	circuit-level or other basic analog struct. parameters	basic analog structure parameters
differential pair	• gain, *PSRR* • gain bandwidth, • common-mode input range, • noise, offset	• $g_{m1,2}/g_{DS1,2}$, • $g_{m1,2}$, • $V_{GS1,2}$, • g_{m1}, $v_{n1,2}$, $\sigma(\delta v_{G1,2})$
load	• stability, • noise, offset, • (output range)	• $g_{m3,4}/(\Sigma C)_{G3,4}$, • $g_{m3,4}$, $v_{n3,4}$, $\sigma(\delta v_{G3,4})$, • since $IF_{3,4} = IF_5$
common source	• gain, • stability, • output range, • (noise)	• g_{m5}/g_{DS5}, • g_{m5}, • V_{DSsat5}, • the contribution can be neglected
current bias	• *CMIR*, • *CMRR*, • output range	• $V_{DSsatB11}$, • g_{DSB11}, • $V_{DSsatB12}$

5.4.5 Procedural design sequence

The procedural design steps and the possible optimization loops are illustrated in Figure 5.22.

Circuit currents (and compensation capacitance initial value) - The circuit currents are chosen to respect the total current consumption specification on one hand, and the speed or slew-rate requirements on the other hand. If the bias current of the first stage is calculated with regard to the specified slew-rate, then the initial value of the compensation capacitance also has to be specified in this step.

Since there is no general rule of how to chose the compensation capacitance value, we propose here to start with a value that is in the range of 1pF to 5pF or with half of the value that corresponds to the maximal allowed capacitance surface in the circuit layout. The final value, required to fulfill the stability condition, is determined later when the second stage is designed. This will however require repetition of some or all of the design steps, since some of the circuit-level design parameters can be affected.

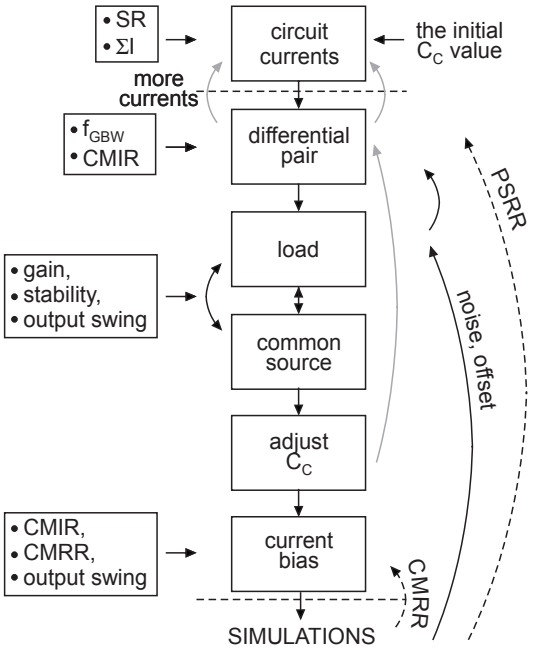


Figure 5.22 Miller operational amplifier procedural design sequence

The bias current of the second stage is generally chosen to be 8 to 12 times larger than the bias current of the first stage so that the required gain and stability are achieved.

Differential pair - The inversion factor of differential pair transistors is usually determined according to the required f_{GBW} as

$$g_{m1,2} \geq 2\pi C_C \cdot f_{GBW} \rightarrow \frac{g_{m1,2}}{I_1/2} = \left(\frac{g_m}{I_{Dsat}}\right)_{1,2} \rightarrow IF_{1,2}, \tag{5.83}$$

where the initial value of the compensation capacitance is used. Later, when the final value is chosen, this calculation has to be repeated.

On the other hand, if the common-mode input range specification has higher priority, the maximal value of the inversion factor is imposed by the acceptable gate-source voltage headroom $V_{GS1,2}$.

The transistor lengths $L_{1,2}$ are determined with regard to the specified gain of the first stage, since the output resistance of the first stage is approximated by $1/g_{DS1,2}$ as discussed previously.

Load - As the transistors M3,4 and M5 have the same inversion factor, the initial value can be determined from the output range specification

$$V_{DSsat3,4} = V_{DSsat5} \rightarrow IF_{3,4,5}. \tag{5.84}$$

However, if noise or offset reduction is a high design priority, then the minimal inversion factor of the transistors M3,4 can be imposed by the condition

$$g_{m3,4} < g_{m1,2}. \tag{5.85}$$

Common source - Since the inversion factor is set in the previous step, the transconductance of the second stage is determined by the bias current I_2. Therefore, the minimal acceptable transconductance is calculated from the stability condition as

$$g_{m5} > k_{PM} \cdot f_{GBW} \cdot 2\pi C_L, \tag{5.86}$$

in order to determine the minimal acceptable bias current as

$$I_2 > g_{m5} \cdot nV_t \cdot \left(\frac{1}{2} + \sqrt{\frac{1}{4} + IF_5}\right). \tag{5.87}$$

The transistor length L_5 is calculated from the specified gain of the second stage, since the output resistance of the second stage is simplified to $1/g_{DS5}$. Since the transistor lengths $L_{3,4}$ are equal to L_5, the additional condition to be fulfilled is that the output conductance of the transistors M1,2 is equal to or larger than the output conductance of transistors M3,4, that is

$$L_5 = L_{3,4} \geq \frac{U_{a,P}}{U_{a,N}} \cdot L_{1,2}\,. \tag{5.88}$$

Compensation capacitance - After the gain stages have been designed, the compensation capacitance value can be adjusted, using the condition defined previously

$$\frac{g_{m5}/I_2}{2\pi C_L} > k_{PM} \cdot \frac{g_{m1,2}/mI_1}{2\pi C_C} \rightarrow \frac{(g_m/I_{Dsat})_5}{2\pi C_L} > \frac{k_{PM}}{2m}\frac{(g_m/I_{Dsat})_{1,2}}{2\pi C_C}. \tag{5.89}$$

Obviously, if the capacitance value is changed, this may require adjustment of the bias current of the second stage or of the inversion level of the transistors M1,2 or the transistor M5. Therefore, a good design practice is to combine this step with the simulations of the amplifier biased with the ideal current sources. It is important to note that it is not sufficient to increase the compensation capacitance value to achieve stability, since there is a kind of saturation level as shown in Figure 5.23. Hence, it is mandatory to find the compromise between the design parameters in equation (5.89).

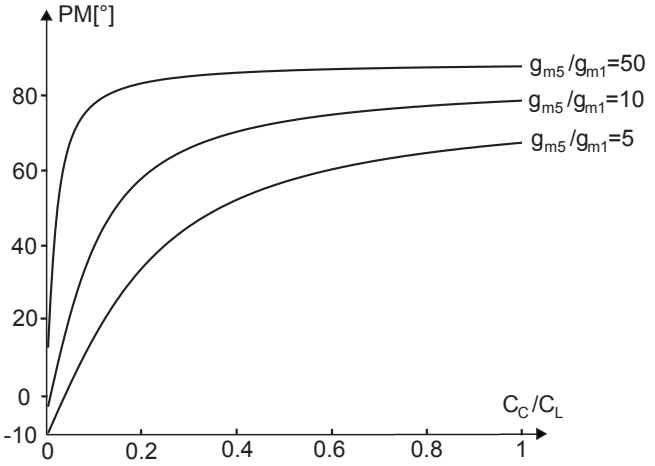


Figure 5.23 Phase margin of the open unity-gain loop as a function of the compensation capacitance for different ratios of transconductances of the first and the second gain stages

Current bias - The inversion level of transistors MB11, MB12 (and MP1) is determined according to the output range of the amplifier, i.e. the acceptable saturation voltage $V_{DSsatB12}$. However, if the common-mode input range specification is not achieved in this way, the inversion level is calculated from the acceptable saturation voltage $V_{DSsatB11}$.

The transistor lengths are determined respecting the additional design condition, i.e. that the output conductance of transistor MB12 is smaller than the output conductance of transistor M5

$$L_{B11,\,12} \geq \frac{U_{a,\,N}}{U_{a,\,P}} \cdot L_5. \tag{5.90}$$

At the same time, if *CMRR* is given as a design specification, the transistor length has to be set to achieve the output conductance g_{DSB11} that ensures meeting the required *CMRR*.

Since the transistor MP1 is matched with the transistors MB11,12, the transistor MP2 is sized to keep it at the limit of the saturation region. The resistance R_{BIAS} is determined as

$$R_{BIAS} = \frac{V_{GP2} - V_{GP1}}{I_{BIAS}}, \tag{5.91}$$

where the gate voltages V_{G1} and V_{G2} are calculated according to the chosen inversion levels of MP1 and MP2.

After the complete circuit has been sized at transistor level, it is simulated to verify the achieved performances. At this point, fine-tuning is usually required to improve noise, offset or rejection ratios. The backward arrows in Figure 5.22 indicate which blocks need to be redesigned (or their parameters adjusted) to improve these circuit-level parameters.

5.5 Conclusion

The procedural design scenarios presented in this chapter are illustrations of the design procedure that can be applied to any analog cell. It is based on the structured design approach and consists in: circuit partitioning, derivation of the specifications for each basic analog structure, and step-by-step design sequence in transconductance - load - bias structure order.

The important advantage of the presented design procedure is that it can be easily implemented in a CAD tool that can be used for analog design assistance in interaction with a simulator. In this way, it becomes possible to encapsulate the basic analog design rules without limiting the user's interaction and decision making at the important design points.

Bibliography

[1] P. Allen, D. Holberg, “CMOS Analog Circuit Design”, Oxford University Press, second edition, January 2002, ISBN10: 019511644

[2] B. Razavi, “Design of Analog CMOS Integrated Circuits”, McGraw-Hill, August 2000, ISBN10: 0072380322

[3] D. Johns, K. Martin, “Analog Integrated Circuit Design”, Wiley, November 1996, ISBN10: 0471144487,

[4] P. Gray, P. Hurst, S. Lewis, R. Meyer, “Analysis and Design of Analog Integrated Circuits”, Wiley, February 2001, ISBN10: 0471321680

[5] K. Laker, W. Sansen, “Design of Analog Integrated Circuits and Systems”, McGraw-Hill Companies, January 1994, ISBN10: 007036060X

[6] W. Sansen, “Analog Design Essentials”, Springer, November 2006, ISBN10: 0387257462

CHAPTER 6

PAD tool

The Procedural Analog Design tool called PAD, is a chart-based design tool that allows step-by-step design of analog cells. Its interactive interface enables instantaneous visualization of design trade-offs. At each step, the user modifies interactively one subset of design parameters and observes the effects on other circuit parameters. At the end, an optimized design is ready for simulation (verification and fine-tuning). The present version of PAD covers the design of basic analog structures and the procedural design of several single-stage and two-stage amplifier topologies.

6.1 Introduction

Many CAD tools and environments have been proposed to facilitate the design task and increase the circuit quality. At the system-level, the design flow for mixed-mode circuits uses different methodologies to design and layout analog and digital parts (Figure 6.1).

For digital circuitry, the design procedure can be strongly automated. It consists of: circuit specification, behavioral modeling, circuit synthesis, layout generation, layout extraction and postlayout simulation. This design flow is capable of automating digital functionalities from behavioral modeling to physical implementation without any transistor-level manipulation, therefore facilitating cell retargeting and technology migration.

However, for analog functions, CAD tools available on the market today do not provide analog cell synthesis from the behavioral description. The problem of topology selection and circuit synthesis leads, in a classical analog design approach, directly to transistor-level sizing. This makes the analog design procedure very complicated and results in attempts at analog design automation primarily based on numerical computation. This automation without user optimization can be very restrictive, and moreover, retargeting and reuse of circuit topologies becomes practically impossible.

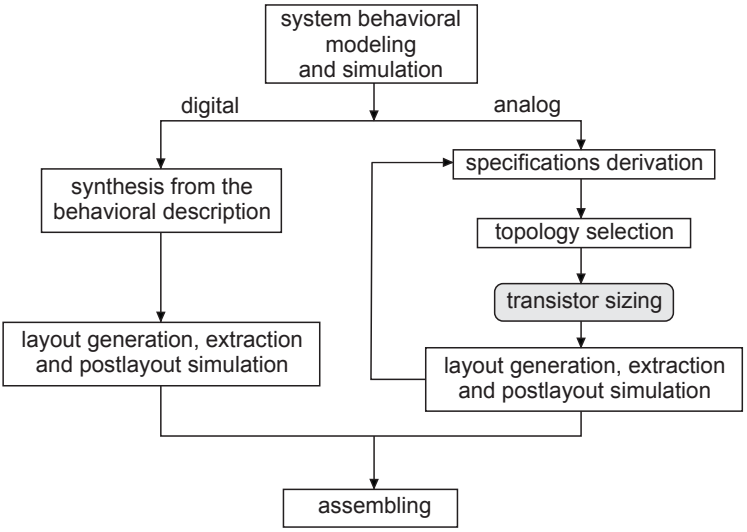


Figure 6.1 Top down mixed-mode system design flow

6.1.1 Overview of analog CAD tools

By the end of 1980s and in the early 1990s, a huge amount of academic research was accomplished in the domain of analog design automation and analog CAD tools development [1]-[3]. The principal goal was to try to replace the analog designer in some or all analog design tasks. As a result, several analog design environments based on top-down hierarchical design flow covering the full design path from topology selection, via transistor sizing and design optimization to layout generation, were proposed. The most important representatives, described in [4]-[10], are:

- ACACIA (OASYS, ANAGRAM), Carnegi-Mellon University, 1989,
- OPASYN, University of California, Berkeley, 1990,
- ADAM (IDAC, ILAC, SYNAP), CSEM, 1990,
- ASAIC (HECTOR, OPTIMAN, AUTOLAC, ISAAC, DONALD), UKL, 1990 and its successor AMGIE, UKL, 2001,
- CHIPAIDE, Imperial College, University of London, 1990.

Even though all these tools use different design strategies, the problem of transistor sizing and design optimization is usually solved as a mathematical problem using either knowledge-based or optimization-based methods. The main drawback is that the design procedure is strongly automated with very limited possibility of user intervention or insight into design decisions.

Therefore, in the case when a solution is not found, the designer must return to the simulation tools and try to find the answer usually by the "trial and error".

By the end of the 1990s, the GPCAD tool [11] which uses convex optimization techniques and geometrical programming for transistor sizing was proposed at Stanford University. Although, these methods can solve large problems, with thousand of variables and constraints, the circuit design is again treated as a strictly mathematical problem, and again it is possible to find solutions in a limited number of cases only.

Recent scientific research, however, shows a tendency towards visualization of the design parameter interdependences [12] and design tools that allow the user to find answers to design problems [13]. In addition, several design and optimization methods based on either graphical MOS models [14] or g_m/I_D methodology derived from the EKV MOS model [15], [16] have been proposed. In this way, the emphasis is rather given to assistance in the analog design procedure than to its complete automation. The main goal is to develop methods and tools that enable the exploration of design limits imposed by the technology, that minimize the computational efforts, and that provide an insight into relations between design parameters and circuit performances, as well as an understanding of analog devices behavior.

6.1.2 Procedural Analog Design (PAD) tool

In this book, we present the concept of structured analog design followed by procedural design flow that can be used for the design of a large number of analog cells. The use of structured design provides a systematic vision of analog amplifier design within the framework of a complex system. On the one hand, an amplifier is partitioned into simple analog building blocks that affect only few circuit-level parameters and require less design effort. On the other hand, the resulting behavioral model allows to extract the circuit-level requirements more precisely, and to confirm the basic block specifications that have to be respected, such as transconductance, output current, or output resistance.

Nevertheless, the realization of such a design procedure is not possible without a suitable transistor-level design methodology, which simplifies the task of transistor sizing and enables to find the design optimum. The concept of the inversion factor and the definitions of the transistor parameters as a function of it, derived from the EKV modeling approach, provide a missing link and make it possible to successfully accomplish the design procedure from system-level specifications to transistor-level implementation.

The proposed design concept is implemented as a chart-based **P**rocedural **A**nalog **D**esign tool, called PAD [17]-[18], dedicated to structured transistor-

level design of analog circuits. It allows step-by-step design of analog cells by using guidelines for each analog topology. Furthermore, its interactive interface provides instantaneous visualization of design trade-offs.

The two main purposes of the PAD tool are to reduce expert analog designers' development time and help to hand on analog knowledge to system level designers. The procedural design can be combined with the behavioral simulations (Figure 6.2) in order to find the optimal solution. Finally, the implemented basic analog structures library also facilitates migration of analog blocks from one technology to another.

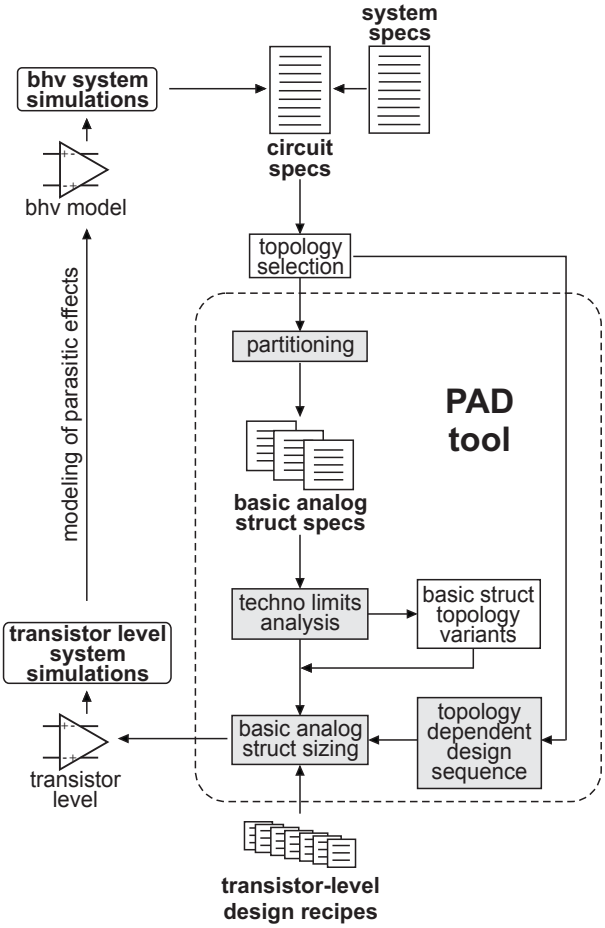


Figure 6.2 Analog design flow based on the structured design approach and the use of the PAD tool

6.2 PAD structure

The PAD tool introduces a new interactive knowledge-based design methodology for sizing of analog cells. It assists the user by presenting the necessary theoretical knowledge and tips from an experienced designer. Its interactive graphical interface provides visual feedback and the possibility of exploring parameter relations in order to find good trade-offs. The tool structure is illustrated in Figure 6.3.

A transistor-level calculator, based on the complete set of equations of the EKV MOS model [19], is capable of exploring complex relations and displays the results in charts, with which the user can interact. The EKV model is chosen because it is strongly based on device physics and dedicated to the design of analog circuits. It links the equations for weak and strong inversions in a continuous way, has a small number of parameters and a very good accuracy. Moreover, it is possible to find solutions for different input parameter sets without using complex numerical methods. A large number of transistor parameters, which are important for analog design, can be extracted from the model, such as: inversion factor, saturation voltage, Spice-like threshold voltage, Early voltage, small-signal parameters, parasitic capacitances, g_m/I_{Dsat} ratio, transconductance efficiency factor. Finally, the transistor-level design methodology proposed in chapter 2 is also based on the EKV modeling concept.

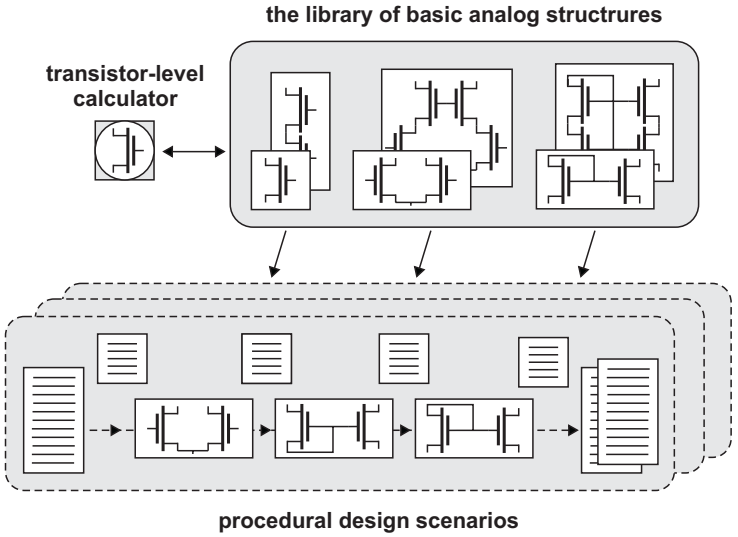


Figure 6.3 PAD structure

The library of basic analog structures relies on a transistor-level calculator and includes the most common basic blocks for the design of analog circuits. For each basic analog structure, the user-friendly graphical interface has the same appearance. All important design parameters are calculated and displayed. Furthermore, the guidelines and the basic design rules are presented to the user in a separate explanation window. The designer can analyze the parameter dependences of each basic structure, acquiring in this way an intuitive understanding of the structure behavior.

The procedural design flow of complex analog structures is based on the partitioning of each complex structure into basic analog structures. It covers the systematic design of different operational amplifier topologies, as well as the design of OTA followed by a common source/drain stage for power management applications. The design sequence for each analog topology is implemented in the form of a design scenario. The designer can navigate from one point to another through the scenario in order to modify the design at the transistor level and reach the optimum cell performance.

6.2.1 Chart-based graphical interface

The charts enable the transcription of mathematical relations into appropriate interactive graphical representations. The designer can change different parameters and observe simultaneously the behavior of the other parameters, since all dependences and characteristics of an analog structure are presented at one glance. As a result, the design trade-offs are easily found and the circuits can be optimized efficiently.

An example of an NMOS transistor design chart is shown in Figure 6.4. The DC operating point voltages and drain current, as well as transistor design variables are shown on the left-hand side of the window. The transistor design parameters and their appropriate graphical representations are organized in tabs on the right-hand side of the design window.

6.3 Basic analog structures library

The analog structures library, embedded in PAD, includes the following basic analog structures:

- differential pair,
- cascoded differential pair,
- folded cascode differential pair,
- common source,
- common drain,
- cascode,

- simple current mirror,
- and cascode current mirror.

This tool also allows the analysis of a single transistor (NMOS, PMOS) and a diode-connected transistor as a basic analog structure.

For each structure a set of general parameters (such as: small-signal model parameters, DC biasing values, parasitic capacitances, equivalent noise, speed) is displayed. Some specific parameters are also shown (such as: maximum DC offset for differential pair, current mismatch for current mirror, input capacitance, output resistance, sum of parasitic capacitances in some specific nodes). This enables analysis of the basic structure behavior in the environment of a given circuit, and observation of the parameters that are important for the design trade-offs.

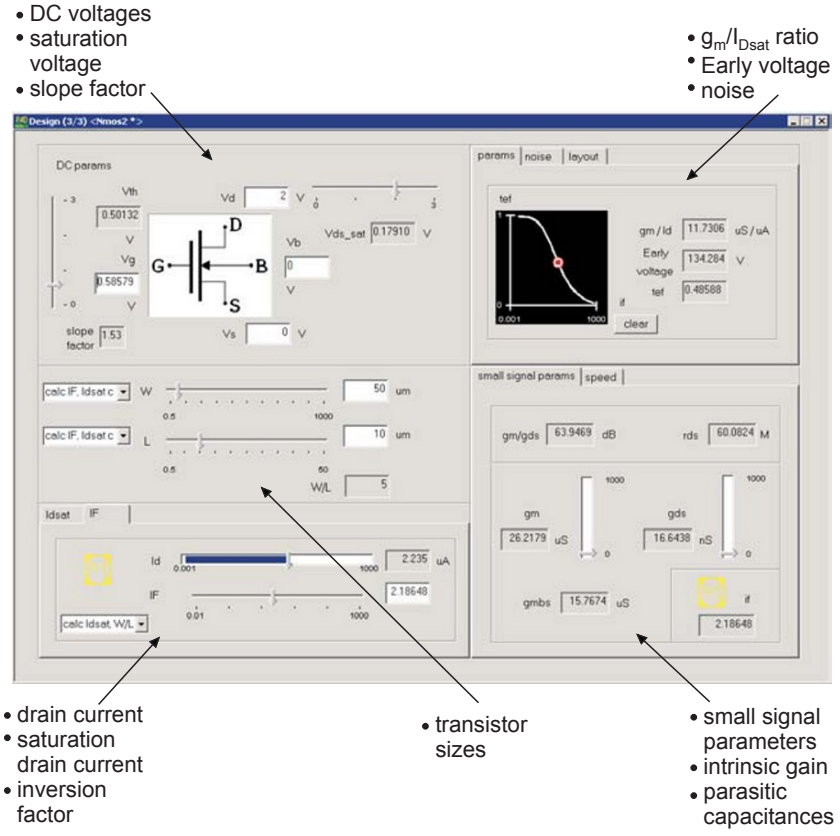


Figure 6.4 Example of PAD's graphical interface: NMOS transistor design chart

The snapshots of design charts for sizing of some basic analog structures are shown in Figure 6.5. The methodology that can be used for sizing, as well as for optimization and resizing of these blocks is presented in chapter 4. The basic analog structures library can be used for the design of a wide range of analog cells. When an analog cell is partitioned into basic blocks, as explained in chapter 4, the circuit can be designed using the design charts for sizing of each basic block, and using the simulator to confirm the circuit-level performances.

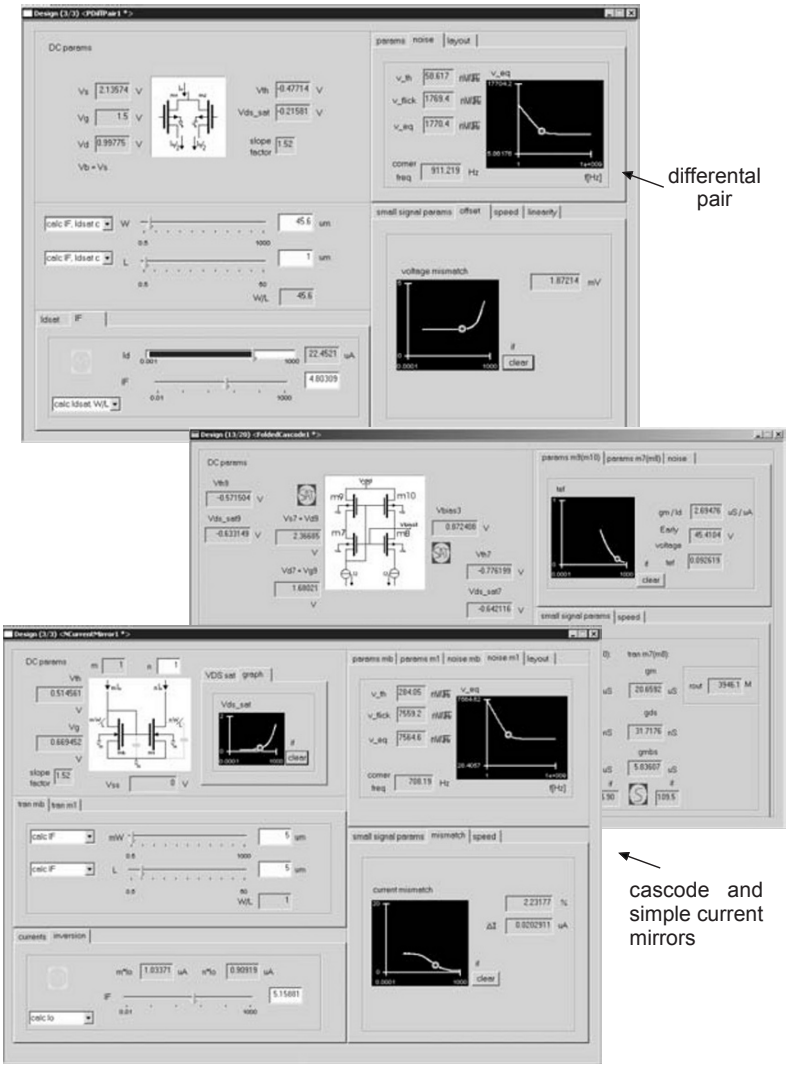


Figure 6.5 Design charts for basic analog structures

6.4 Procedural design scenarios

The current PAD tool version enables systematic design of single-stage amplifiers, such as:

- simple OTA,
- folded cascode OTA,
- and two-stage amplifier structures, such as:
- Miller operational amplifier,
- OTA followed by common source (for power management applications), and
- OTA followed by common drain (for power management applications).

The procedural analog design flow consists of: circuit partitioning into basic analog structures, chart-based sizing of basic analog structures and circuit level design. The proposed design sequence guides the user through the design revealing the parameter dependences and the role of every basic analog structure in the circuit. During and after the design of every basic analog structure, the circuit performance parameters can be checked. The final circuit-level summary provides a complete documentation of the electrical behavior of the designed cell. At the end, an optimized design is ready for simulation. The procedural analog design steps, together with examples of user interface snapshots, are illustrated in the following subsections on the example of the folded cascode OTA and the example of the Miller operational amplifier topology.

The design initialization consists of the EKV model parameters, initial design requirements, supply voltages, bias currents and compensation and/or load capacitance input. On the basis of input parameters, the circuit currents and the input pair transconductance are proposed. The minimum or maximum values of some parameters are also determined. However, there are no imposed values, and the design of every analog structure depends on user decisions.

At the circuit level, the design sequence and the partitioning into basic analog blocks are predefined. The design scenario guides the user presenting, at each step, the necessary theoretical guidelines and tips from an experienced designer in a separate window. The basic analog structures are sized one after another in a specific order, and all dependent circuit parameters are calculated or approximated in order to find trade-offs. The user has a great degree of freedom, which enables exploration of the design space and the limits of the basic structure's parameters. Furthermore, the user can move backward/forward through the scenario to reconsider the taken decisions or to check the proposed guidelines and the design rules. In addition, the dependence analysis of parameters at the circuit level is implemented as a separate design step. It also includes the frequency analysis, the equivalent noise and input offset

calculations. As a penultimate step, all circuit parameters are calculated and presented in a circuit performance summary. Finally, several simulation runs can be performed for the verification and fine-tuning of the design parameters.

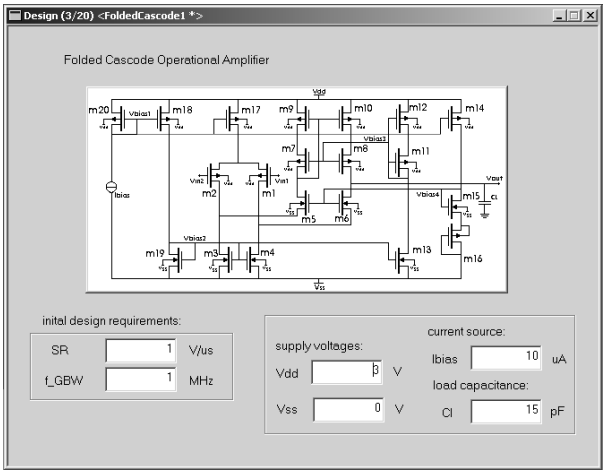
6.4.1 Example of the procedural design for a folded cascode OTA

The procedural design flow is based on analysis presented in chapter 5, section 5.2. The user interface snapshots are illustrated in Figure 6.6 and Figure 6.7. For the folded cascode OTA complex analog structure, the PAD design sequence is as follows: initialization, circuit partitioning, determination of the circuit currents, sizing of the basic analog structures, frequency analysis, circuit summary, noise and offset analysis, and simulation runs for fine-tuning. For each block the designer can modify the set of independent parameters and observe the impact of his decisions on circuit parameters. The most important point is to determine the circuit currents according to the imposed slew-rate value, as well as the transconductances for every transistor according to circuit performances. During the sizing of each block, the circuit performance parameters can be checked. After all building blocks have been designed, circuit-level behavior is summarized and an interface to a simulator is proposed. The simulation runs are needed here only for verification and fine-tuning. This design procedure can be repeated several times in interaction with the simulator if necessary. Not only is the number of simulation runs reduced to a minimum, but transistor sizing and (at the same time) circuit optimization are also achieved.

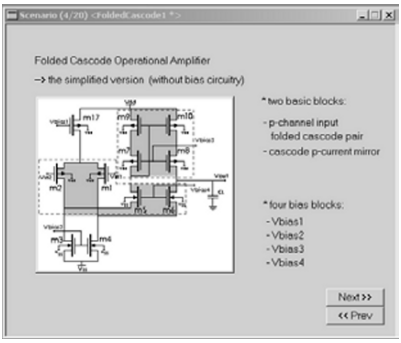
6.4.2 Example of the procedural design for a Miller operational amplifier

The procedural design flow is based on analysis presented in chapter 5, section 5.4. The user interface snapshots are illustrated in Figure 6.8 and Figure 6.9. For the Miller operational amplifier complex analog structure, the PAD design sequence is as follows: initialization, circuit partitioning, determination of first stage tail current, sizing of the basic analog structures of the first stage, frequency analysis, circuit summary, noise and offset analysis of the first stage, determination of second stage current, sizing of the second stage, frequency analysis, circuit summary, and noise analysis of the amplifier, simulation runs for fine tuning. The value of the compensation capacitance is initialized at the beginning of the design procedure, and has to be adjusted after the second stage has been designed. The circuit currents are determined according to the imposed slew-rate value, and the stability condition, i.e the required phase margin. In the same way as in the previous example, during and after the sizing of each block, the circuit performance parameters can be checked and after all building blocks have been designed, an interface to a simulator is proposed.

initialization



circuit partitioning and circuit currents' determination



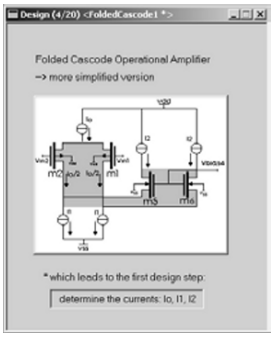


Figure 6.6 Folded cascode OTA design flow implemented in PAD (part 1)

performance evaluation after the sizing of basic analog structures

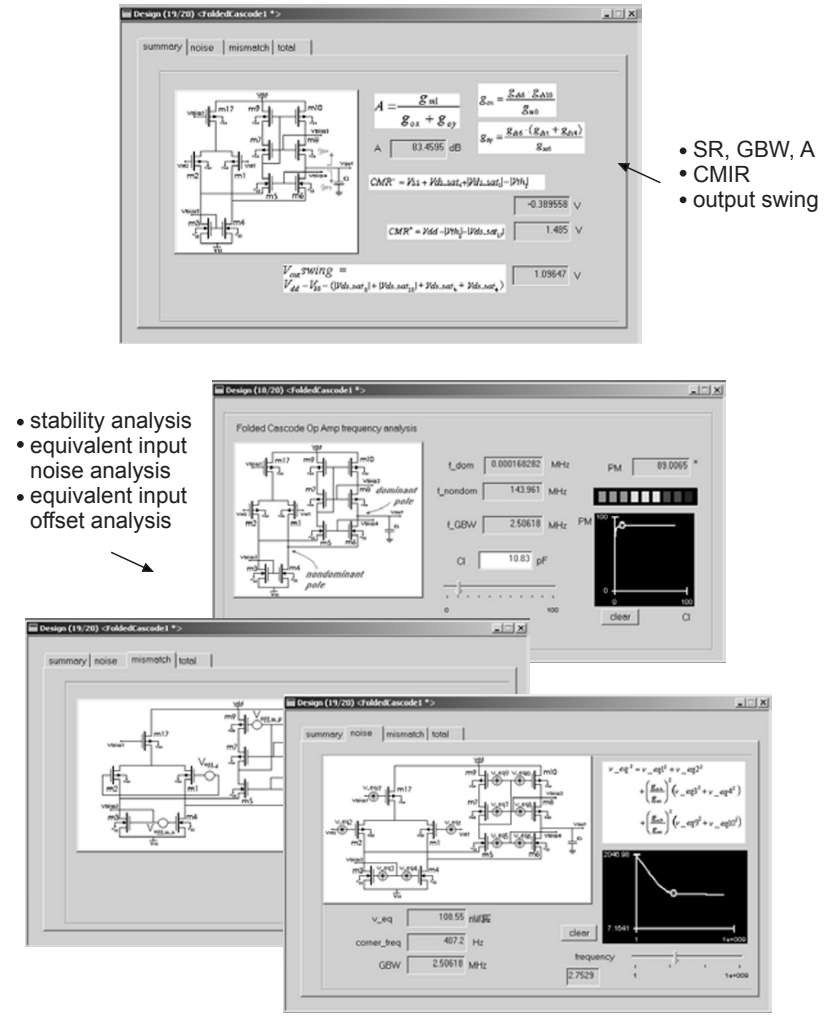
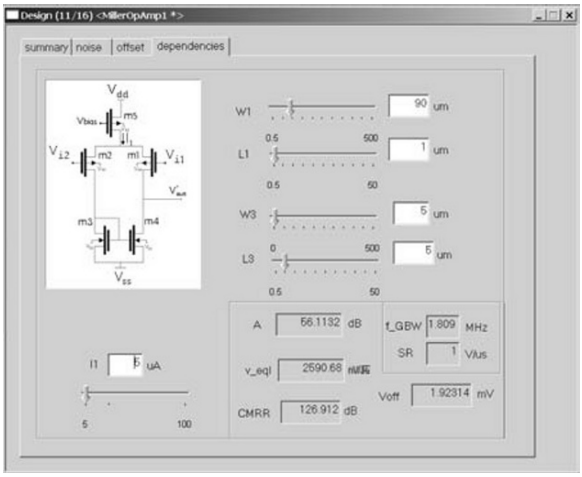


Figure 6.7 Folded cascode OTA design flow implemented in PAD (part 2)

performance analysis after the first stage design



the second stage sizing

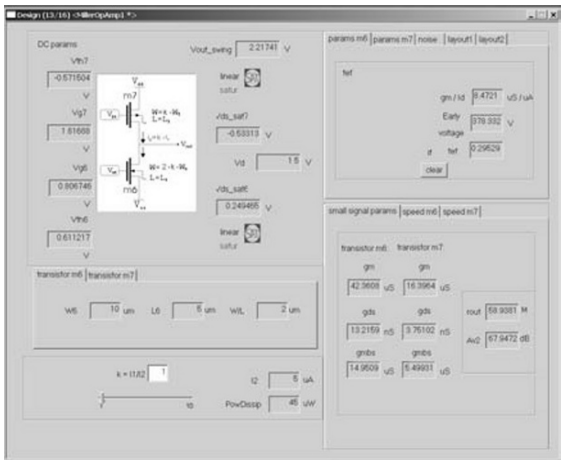


Figure 6.8 Miller operational amplifier design flow implemented in PAD (part 1)

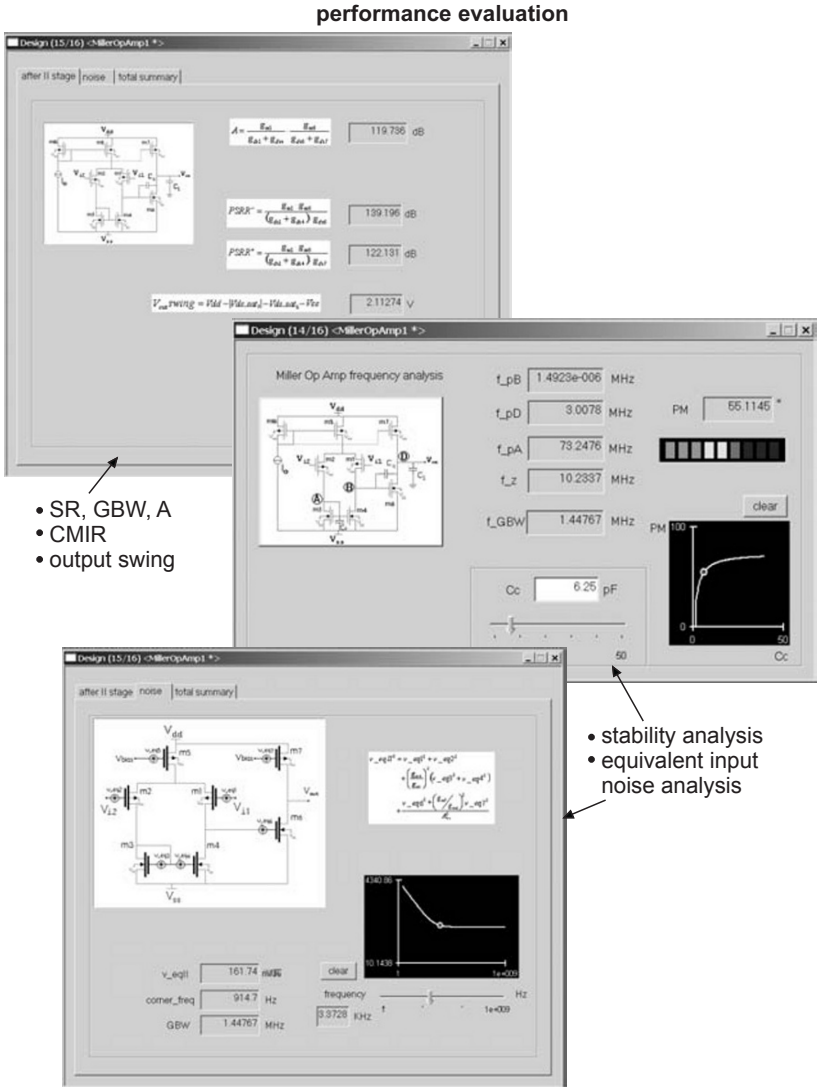


Figure 6.9 Miller operational amplifier design flow implemented in PAD (part 2)

6.5 BSIM and EKV model library file input

As described previously, the transistor level calculator implemented in PAD is based on the EKV MOS model. This implies that PAD requires the EKV model parameters (or the EKV model library file) as input. On the other hand, because the BSIM family of models have been widely used for the design of analog circuits during the past few years and BSIM version 3.3 has become an industrial standard for submicron technologies, there are some foundries that do not provide EKV model library files. Nevertheless, the EKV model is indispensable for transistor-level design of analog cells implemented in PAD. In order to make it possible to use the PAD tool with both BSIM and EKV input library files, a BSIM2EKV converter [20], [21] has been developed as a separate module (Figure 6.10). It is described in detail in chapter 3 and it generates the EKV model library file version 2.6 from the BSIM model library file version 3.3.

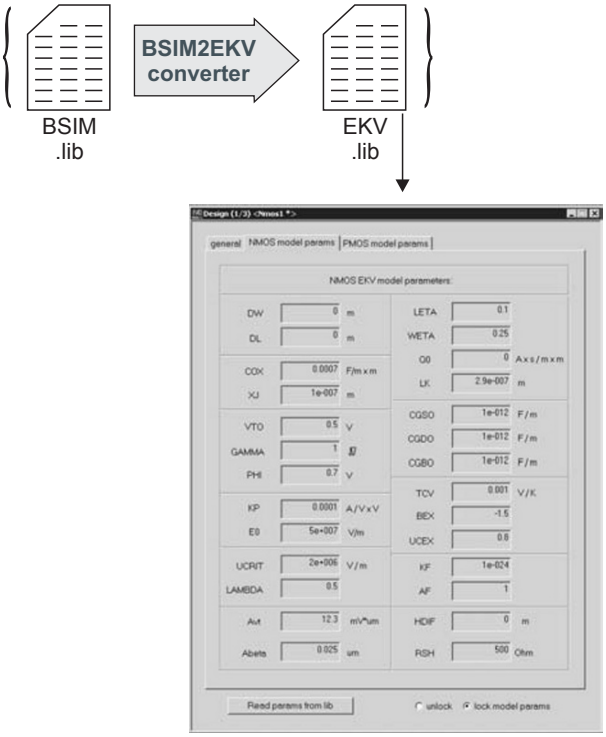


Figure 6.10 BSIM/EKV model library file as PAD input and the PAD input dialog snapshot for the EKV model parameters

6.6 Conclusion

This chapter presented a new chart-based tool based on structured design methodology. It makes it possible to encapsulate design knowledge and experience into a CAD tool and to hand them on to other designers. The proposed procedural design enables transistor sizing and circuit optimization at the same time. Simulation runs are required only for verification and fine-tuning and therefore reduced to a minimum.

The PAD tool enables design and re-design of a wide range of circuits. The present version covers the most commonly used analog topologies. For the design of other (more specific or more complicated) circuit topologies, the basic analog structures library can be used for sizing separate blocks. The simulator is then used interactively with PAD after the sizing of each block. In this way, a great improvement in the quality of the analog design can be achieved.

6.7 Download

PAD tool is developed in the frame of the Ph.D. project that was performed at Electronics Labs, EPFL, Lausanne, Switzerland. The trial version of the PAD tool and supporting documentation can be downloaded from the Electronics Labs web site: **http://analog.epfl.ch**.

Bibliography

[1] B. Antao, “Trends in CAD of analog ICs”, IEEE Circuits and Devices Magazine, Volume 12, pp. 31-41, September 1996

[2] L. R. Carley, R. Rutenbar, “How to automate analog IC designs”, IEEE Spectrum, Volume 25, pp. 26-30, August 1988

[3] L. R. Carley, G. Gielen, R. Rutenbar, W. Sansen, “Synthesis tools for mixed signal ICs: Progress on frontend and backend strategies”, 33rd Design Automation Conference, pp. :298-303, June 1996

[4] L. R. Carley, D. Garrod, R. Hadrjani, J. Kelly, T. Lim, E. Ochotta, R. Rutenbar, “ACACIA: The CMU analog design system”, Custom Integrated Circuits Conference, pp. 4.3.1-4.3.5, May 1989

[5] P. Maulik, L. Carley, D. Allstot, “Sizing of cell-level analog circuits using constrained optimization techniques”, Journal of Solid-State Circuits, Volume 28, pp. 233-241, March 1993

[6] H. Koh, H. Sequin, P. Gray, “OPASYN: a compiler for CMOS operational amplifiers”, Transactions on Computer-Aided Design of Integrated Circuits and Systems, Volume 9, February 1990

[7] M. Degrauwe, B. Goffart, B. Joss, J. Rijmenants, C. Meixenberger, T. Schwarz, T. Horstein, and P. Deck, “The ADAM analog design automation system”, International Symposium on Circuits and Systems, Volume 2, pp. 820-822, May 1990

[8] G. Gielen, W. Sansen, “Symbolic Analysis for Automated Design of Analog Integrated Circuits”, Springer, May 1991, ISBN 0792391616

[9] G. Van der Plas, G. Debyser, F. Leyn, K. Lampaert, J. Vandenbussche, G. Gielen, W. Sansen, P. Veselinovic, D. Leenarts, “AMGIE-A synthesis environment for CMOS analog integrated circuits”, Transactions on Computer-Aided Design of Integrated Circuits and Systems, Volume 20, pp. 1037-1058, September 2001

[10] C. Makris, C. Berrah, X. Xiao, M. Singha, A. Ilumoka, J. Stone, C. Toumazou, P. Cheung, R. Spence, “CHIPAIDE: a new approach to analogue integrated circuit design”, Colloquium on Analogue VLSI, pp. 1/1 –111, May 1990

[11] M. del Mar Hershenson, S. P. Boyd, T. H. Lee, “GPCAD: a tool for CMOS op-amp synthesis”, International Conference on Computer-Aided Design, pp. 296-303, November 1998

[12] R. Spence, “The facilitation of insight for analog design”, Transactions on Circuits and Systems II: Analog and Digital Signal Processing, Volume 46, pp. 540-548, May 1999

[13] R. Ionita, A. Vladimirescu, P. Jespers, “Automated design methodology for CMOS analog circuit blocks in complex systems”, International Workshop on Integrated Circuit Design, Power and Timing Modeling, Optimization and Simulation, Lecture Notes In Computer Science, Volume 2451, pp. 487-494, 2002

[14] H. Wallinga, K. Bult, “Design and analysis of CMOS analog signal processing circuits by means of a graphical MOST model”, Journal of Solid-State Circuits, Volume 24, pp. 672-680, June 1989

[15] F. Silveira, D. Flandre, P. G. A. Jespers, “A gm/ID based methodology for the design of CMOS analog circuits and its application to the synthesis of a silicon-on-insulator micropower OTA”, IEEE Journal of Solid-State Circuits, Volume 31, pp. 1314-1319, September 1996

[16] D. Binkley, M. Bucher, D. Foty, “Design-oriented characterization of CMOS over the continuum of inversion level and channel length”, IEEE International Conference on Electronics, Circuits and Systems, pp. 161-164, December 2000

[17] D. Stefanović, M. Kayal, M. Pastre, V. Litovski, “Procedural analog design (PAD) tool”, 4th International Symposium on Quality Electronic Design, pp. 313-318, March 2003

[18] D. Stefanović, M. Kayal, M. Pastre, ”PAD: A new interactive knowledge-based analog design approach”, Analog Integrated Circuits and Signal Processing Journal, Volume 42, pp. 291-299, March 2005

[19] Ch. C. Enz, E. A. Vitoz, “Charge-based MOS Transistor Modeling: The EKV Model for Low-power and RF IC Design”, John Wiley & Sons, first edition, 2006, ISBN 9780470855416

[20] D. Stefanović, F. Krummenacher, M. Pastre, M. Kayal, “BSIM2EKV: un outil pour la conversion automatique des paramètres du modèle BSIM aux paramètres du modèle EKV”, TAISA’04 5ème Colloque sur le

Traitement Analogique de l'Information, du Signal et ses Applications, pp. 85-88, September 2004

[21] D. Stefanović, F. Krummenacher, M. Paste, M. Kayal, "BSIM2EKV: BSIM3.3 to EKV 2.6 Model Library File Automatic Conversion", EKV Users' Meeting/Workshop, November 2004

CHAPTER 7

Topology variants

In this chapter, we discuss the topology variants at the basic analog structures level. In the case when certain design requirements cannot be achieved because of technology limits or the design trade-offs cannot be found, a possible solution may be to replace the basic analog structure that affects the design parameter in question by its "more advanced" version. The concept is illustrated here on the example of a fully-differential folded cascode OTA.

7.1 Basic concept

When an analog cell is partitioned into its building blocks and the design specifications are derived for all basic analog structures, it may occur that a certain design requirement cannot be achieved because of technology limits or that the necessary design trade-offs cannot be found. In this case, a possible solution is to replace the basic analog structure that affects the design parameter in question by one of its topology variants.

The topology variant represents the basic analog structure modified in such a way to make it possible to achieve the required design parameter without any sizing effort (thanks to this topology modification). Moreover, the design parameters of other basic analog structures are not affected. In this way, the designer does not have to deal with the parameter technology bounds and the design trade-offs can be relaxed.

7.1.1 Design example: fully-differential folded cascode OTA and its CMFB amplifier

The concept of topology variants will be illustrated here on the example of a fully-differential folded cascode OTA, since it is used in the practical design example presented in chapter 8. The main amplifier and its CMFB amplifier, partitioned into basic analog structures, are depicted in Figure 7.1 and Figure 7.2, respectively. For each basic analog structure, possible topology variants are indicated in the same figures.

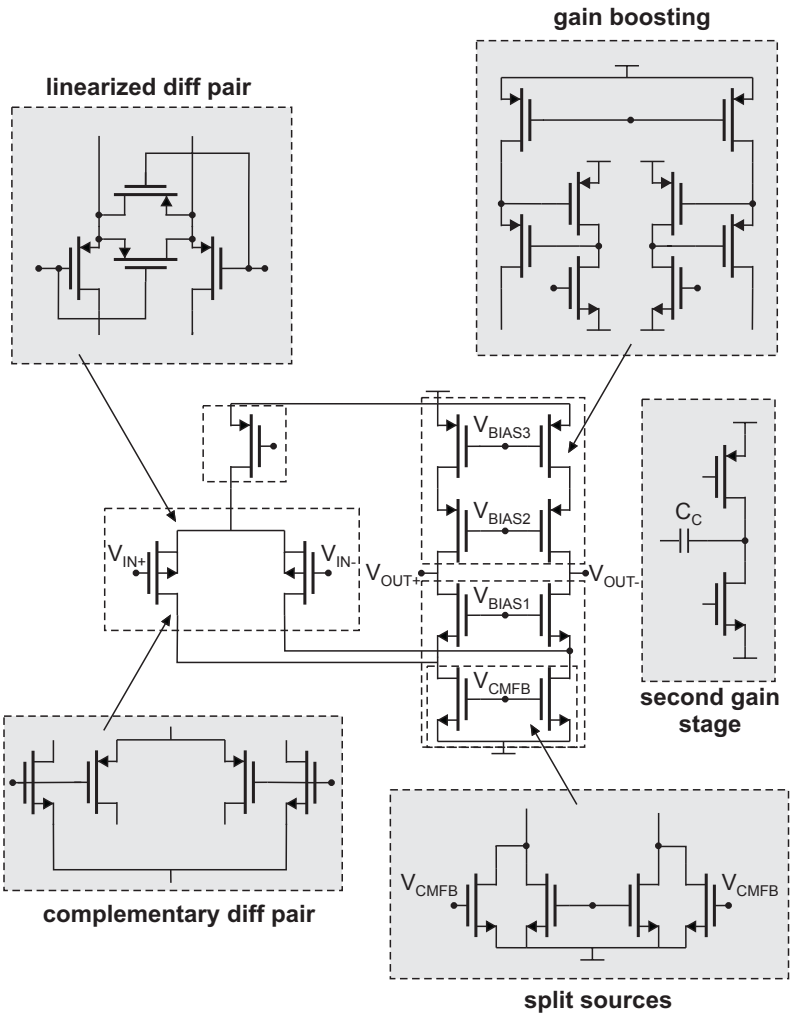


Figure 7.1 Fully-differential folded cascode OTA: the main amplifier and possible topology variants

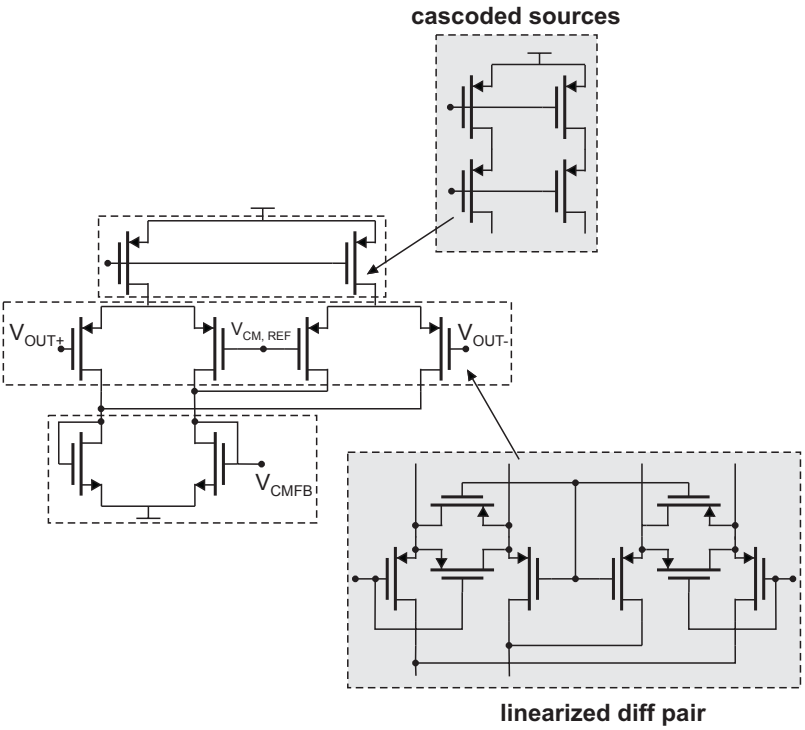


Figure 7.2 Fully-differential folded cascode OTA: the CMFB amplifier and possible topology variants

The design problems that can be resolved are the following:

- **large differential-mode input range** - using a linearized differential pair at the input of the main amplifier,
- **large common-mode input range** - using a complementary differential pair at the input of the main amplifier,
- **large differential output swing** - using linearized differential pairs at the input of the CMFB amplifier,
- **improved rejection of the differential signal in the common mode signal** - cascoding the current bias sources in the CMFB amplifier,
- **high gain** - adding a gain stage as either the second gain stage or the gain-boosting amplifier.

The proposed topology variants and their important characteristics are analyzed in detail in the following sections. The design guidelines are presented together with the possible design drawbacks in each case.

7.2 Gain enhancement - two stages

Figure 7.3a depicts the behavioral model of a single-stage amplifier ($g_{m1}R_{OUT1}$) followed by a second gain stage ($g_{m2}R_{OUT2}$). The overall gain is a product of the gain of the first stage and the gain of the second stage

$$A_0 = A_1 \cdot A_2 = g_{m1}R_{OUT1} \cdot g_{m2}R_{OUT2}. \tag{7.1}$$

Thus, the high gain specification is easily achieved, particularly if the first stage is implemented as a folded cascode structure.

Nevertheless, stability becomes an issue when such an amplifier is used with negative feedback. If only the dominant pole of each amplifier is taken into consideration, then, since the time constants $R_{OUT1}C_{OUT1}$ and $R_{OUT2}C_{OUT2}$ are often of the same order of magnitude, there are two poles in the transfer function (Figure 7.3b)

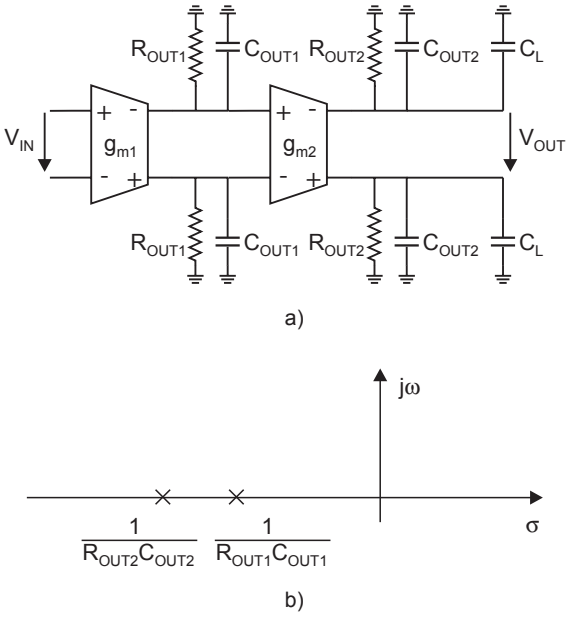


Figure 7.3 Two-stage fully-differential amplifier: a) behavioral model b) poles in the s-plane

$$f_{p1} = \frac{1}{2\pi} \cdot \frac{1}{R_{OUT1} C_{OUT1}} , \tag{7.2}$$

$$f_{p2} = \frac{1}{2\pi} \cdot \frac{1}{R_{OUT2} C_{OUT2}} \tag{7.3}$$

that are near to each other and close to the origin. Each pole contributes a phase shift of -90°, and consequently, an amplifier with negative feedback may easily be unstable. In fact, the stability depends on the negative feedback factor β that should be much smaller than one in this case. However, this is not possible, since the negative feedback network is determined by system topology and it is not permitted to change β for the purpose of stability.

7.2.1 Miller compensation

Stability can be achieved by splitting the two dominant poles in the transfer function. The equivalent circuit, shown in Figure 7.4a, contains an additional capacitance connected between the input and the output of the second stage and is known as the Miller compensation capacitance. The poles are split due to the Miller effect, but there is an additional zero in the right-half plane (see Figure 7.4b) also introducing a -90° phase shift.

The transfer function is expressed as

$$\frac{V_{OUT}}{V_{IN}} = \frac{A_0 \cdot (1 - sC_C / g_{m2})}{D(s)} \tag{7.4}$$

where the DC gain A_0 is given by (7.1) and the denominator is

$$\begin{aligned} D(s) = & \\ & 1 + s((C_C + C_{OUT1})R_{OUT1} + (C_C + C_L)R_{OUT2} + g_{m2} C_C R_{OUT1} R_{OUT1}) \\ & + s^2 (C_C (C_{OUT1} + C_L) + C_{OUT1} C_L) R_{OUT1} R_{OUT2}. \end{aligned} \tag{7.5}$$

The dominant and non-dominant pole frequencies are estimated using the assumption that $f_{ndp} \gg f_{dp}$, as proposed in [1]. Table 7.1 gives the approximations in two different cases:

- $C_L \gg C_C \gg C_{OUT1}, C_{OUT2}$,
- $C_C \geq C_L \geq C_{OUT1}, C_{OUT2}$.

The second case is more realistic for an amplifier designed in the frame of a system using modern submicron CMOS technologies. The load capacitance represents the input parasitic capacitance of the next block, and therefore, it is of the same order of magnitude as the output capacitances of the first and the

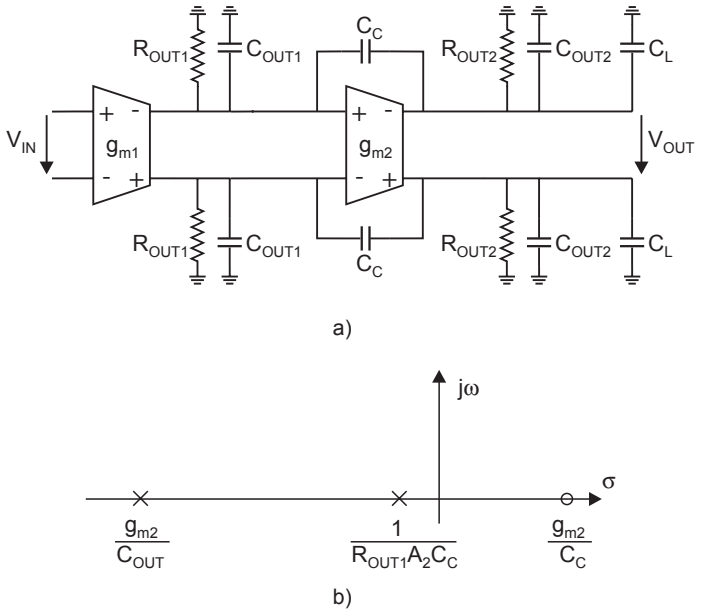


Figure 7.4 Two-stage fully-differential amplifier with Miller compensation: a) equivalent circuit b) poles and zero in the s-plane

second gain stages. Because of this, stability may again become an issue, as is shown further on.

The gain bandwidth frequency is calculated as a product of the DC gain and the dominant pole frequency (i.e. bandwidth)

$$f_{GBW} = (g_{m1}R_{OUT1} \cdot g_{m2}R_{OUT2}) \cdot \frac{1}{2\pi} \cdot \frac{1}{R_{OUT1}g_{m2}R_{OUT2}C_C} \tag{7.6}$$
$$= \frac{1}{2\pi} \cdot \frac{g_{m1}}{C_C}$$

In the first case, when $C_L \gg C_C \ll C_{OUT1}, C_{OUT2}$, it is obvious that $f_{ndp} \ll f_z$ and the stability condition is

$$f_{ndp} \geq k_{PM} \cdot f_{GBW} \tag{7.7}$$

where k_{PM} is determined by the phase margin specification. The non-equality is re-written as

$$\frac{g_{m2}}{C_L} \geq k_{PM} \cdot \frac{g_{m1}}{C_C}, \tag{7.8}$$

Table 7.1 The approximations for zero and pole frequencies of a two-stage amplifier with Miller compensation

	$C_L \gg C_C \gg C_{OUT(1,2)}$	$C_C \geq C_L \geq C_{OUT(1,2)}$
zero f_z	$\frac{1}{2\pi} \cdot \frac{g_{m2}}{C_C}$	$\frac{1}{2\pi} \cdot \frac{g_{m2}}{C_C}$
dominant pole f_{dp}	$\frac{1}{2\pi} \cdot \frac{1}{R_{OUT1} A_2 C_C}$	$\frac{1}{2\pi} \cdot \frac{1}{R_{OUT1} A_2 C_C}$
non-dominant pole f_{ndp}	$\frac{1}{2\pi} \cdot \frac{g_{m2}}{C_L}$	$\frac{1}{2\pi} \cdot \frac{g_{m2}}{C_L + C_{OUT1} + C_{OUT2}}$

revealing that there are two independent variables: g_{m2} and C_C. In other words, the stability is improved (to a certain extent, as shown in Figure 5.23), if either the transconductance of the second stage or the compensation capacitance size is augmented. Unfortunately, this is true only when C_L and C_C are large compared to the parasitic capacitances, which is rarely the case with modern submicron technologies.

On the other hand, when $C_C \geq C_L \geq C_{OUT1}, C_{OUT2}$, the positive zero is closer to the origin than the non-dominant pole, $f_z \leq f_{ndp}$, and the stability condition is

$$f_z \geq k_{PM} \cdot f_{GBW}, \tag{7.9}$$

where k_{PM} is again determined by the phase margin specification. Hence, it follows that

$$\frac{g_{m2}}{C_C} \geq k_{PM} \cdot \frac{g_{m1}}{C_C}, \tag{7.10}$$

and clearly there is only one independent variable: g_{m2}. Now, when the transconductance of the second stage is increased, the output capacitance of the second stage also increases, and the non-dominant pole moves closer to the gain bandwidth frequency. Therefore, it may be a very tedious task to achieve the required stability, and if so, probably at the cost of the current consumption. Moreover, the compensation capacitance is no more an independent variable, and if it is increased, the only consequence is the gain bandwidth degradation, which imposes re-design of the gain stages.

7.2.2 Cascode Miller compensation

When the first stage is implemented as a telescopic or folded cascode OTA, the Miller compensation capacitance can be connected between the output of the second stage and the cascode or the folded point of the OTA as shown in Figure 7.5. For this kind of topology, also discussed in [2], there are no stability problems in the case when the load, the parasitic and the compensation capacitances are of the same order of magnitude.

However, a closer look at the Bode plot of the transfer function of such an amplifier reveals the presence of two complex conjugate poles (Figure 7.6), besides the dominant pole and the positive zero. The frequencies of these poles and zeros are not easily approximated from the exact expression of the transfer function. For this purpose, a simplified half-circuit as depicted in Figure 7.7a and an "intuitive" approach are used.

The positive zero appears if there is a return path of the signal from the output to the input of the amplifier usually through a compensation or a feedback capacitance. The Miller compensation capacitance is connected between the output and the folded point, and thus, does not introduce a positive zero in this case. However, the parasitic gate-drain capacitance of transistors M13,14 in the second stage behaves exactly in the described way. Hence,

$$f_z \approx \frac{1}{2\pi} \cdot \frac{g_{m13,14}}{C_{GD13,14}}. \tag{7.11}$$

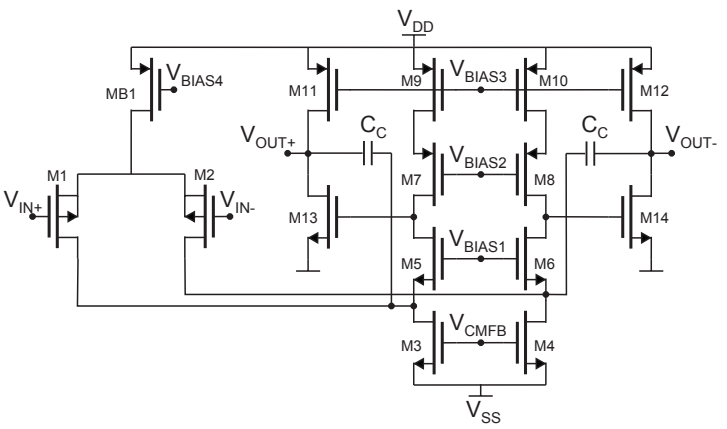


Figure 7.5 Two-stage fully-differential amplifier: folded cascode OTA followed by a common-source stage with a cascode Miller compensation

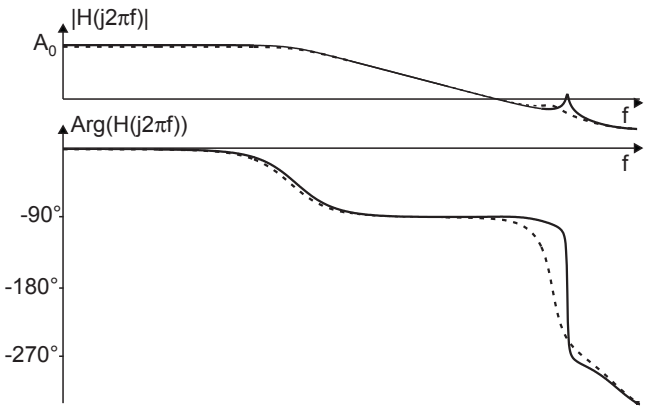


Figure 7.6 Transfer function module and phase of a two-stage amplifier with cascode Miller compensation in the case of over-damped (plain line) and under-damped (dotted line) complex conjugate poles

This zero is usually far away from the origin as $g_{m13,14}$ is high (being the transconductance of the second gain stage) and $C_{GD13,14}$ is small in the saturation region (that is the working condition of the transistors M13,14).

If now the transistors M5,6 and M13,14 are considered as an "equivalent" gain stage, then the Miller compensation capacitor is connected between its input (the folded point) and its output (the amplifier output). Consequently, the dominant pole is determined by the equivalent capacitance and the equivalent resistance seen in the folded point.

The dominant pole is approximated as

$$\begin{aligned} f_{dp} &\approx \frac{1}{2\pi} \cdot \frac{n \cdot g_{m5,6}}{(\Sigma C)_{FP} + g_{m5,6} R_{OUT1} g_{m13,14} R_{OUT2} C_C} \\ &\approx \frac{1}{2\pi} \cdot \frac{1}{R_{OUT1} \cdot A_2 C_C} \end{aligned} \tag{7.12}$$

since the sum of parasitic capacitances in the folded point can be neglected when compared to the compensation capacitance multiplied by the equivalent gain $g_{m5,6} R_{OUT1} \cdot g_{m13,14} R_{OUT2}$.

The gain bandwidth frequency is thus

$$\begin{aligned} f_{GBW} &= (g_{m1,2} R_{OUT1} \cdot g_{m13,14} R_{OUT2}) \cdot \frac{1}{2\pi} \cdot \frac{1}{R_{OUT1} g_{m13,14} R_{OUT2} C_C} \\ &= \frac{1}{2\pi} \cdot \frac{g_{m1}}{C_C}. \end{aligned} \tag{7.13}$$

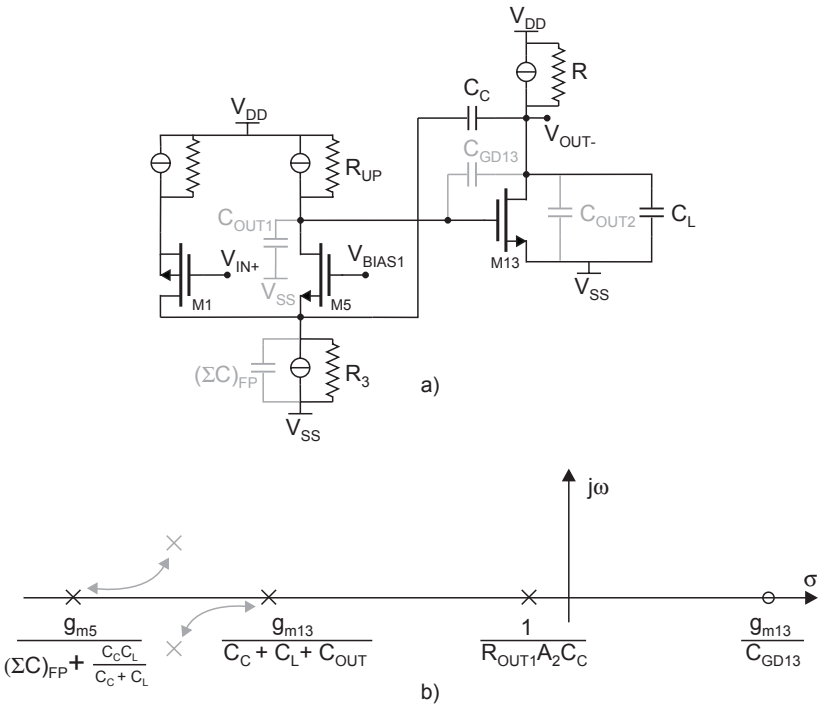


Figure 7.7 Two-stage fully-differential amplifier with cascode Miller compensation: a) equivalent half-circuit, b) zero and poles in the s-plane

The two non-dominant complex conjugate poles are recognized by a peak in the module of the transfer function and a phase shift of -180° (Figure 7.6). If the imaginary part is equal to zero, then the non-dominant poles become real and split (Figure 7.7b).

The first non-dominant pole appears at a frequency for which the impedance of the parasitic capacitance $C_{GD13,14}$ becomes close to zero, that is the output of the first stage is almost shorted to the output of the second stage. Since the transistors M13,14 act in this case as diode-connected, the equivalent resistance in this node is equal to $(1/g_{m13,14}) \parallel R_{OUT1} \approx 1/g_{m13,14}$. Therefore,

$$f_{npd1} \approx \frac{1}{2\pi} \cdot \frac{g_{m13,14}}{C_{OUT1} + C_C + C_{OUT2} + C_L}. \tag{7.14}$$

In the same way, the second non-dominant pole appears at a frequency for which the impedance of the compensation capacitance becomes close to zero, that is the output of the second stage is almost shorted to the folded-point. It follows that

$$f_{npd2} \approx \frac{1}{2\pi} \cdot \frac{n \cdot g_{m5,6}}{(\Sigma C)_{FP} + \dfrac{C_C(C_{OUT2} + C_L)}{C_C + C_{OUT2} + C_L}} . \tag{7.15}$$

To achieve stability, the non-dominant poles have to be placed at a k_{PM} times higher frequency than the f_{GBW}, where k_{PM} is defined by the specified phase margin. However, it must be pointed out that the rough approximations of non-dominant complex-conjugate poles from (7.14) and (7.15) only give the limit of their real parts (see Figure 7.7b). Therefore, the compensation capacitance value is not an independent free variable, and it is imposed by the gain bandwidth specification. The stability condition is ensured by simply respecting $g_{m13,14} > g_{m1,2}$.

In addition, there is one more important condition that must be taken into account. The complex conjugate poles produce the exponentially shaped attenuated oscillation in the time-domain response. When the settling time is a critical design requirement, it is necessary to diminish the amplitude of this oscillation. Figure 7.8 illustrates the consequences of complex conjugate poles in transient behavior in the case of over-damped and under-damped poles. A key design parameter is the ratio $g_{m13,14}/g_{m5,6}$ that has to be decreased, either by the decrease of $g_{m13,14}$ or by the increase of $g_{m5,6}$ (where the latter one is less critical for stability). When the oscillations are under-damped, it can also be observed in the transfer function module as depicted by the dotted line in Figure 7.6.

The cascode Miller compensation is often realized in a symmetrical way (Figure 7.9). The advantage is that, for the same behavior in the frequency domain, the sum of compensation capacitances, and thus their surface in the

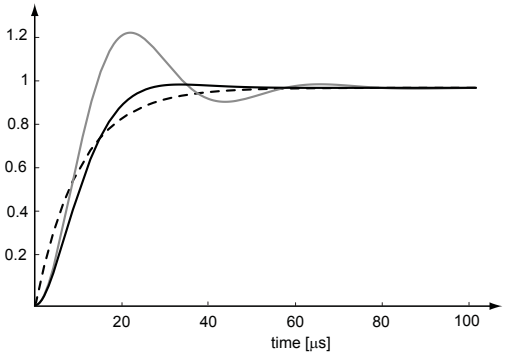


Figure 7.8 Transient response in the case of over-damped (gray line) and under-damped poles (black line) compared to the case without complex conjugate poles (dotted line)

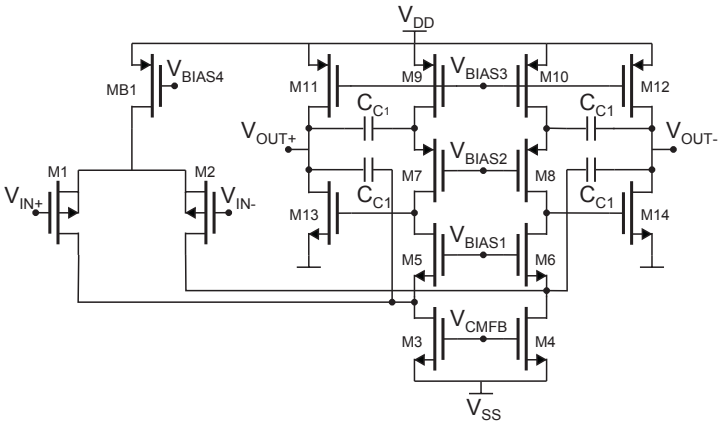


Figure 7.9 Two-stage fully-differential amplifier: folded cascode OTA followed by a common source stage with a symmetrical Miller compensation

circuit layout, is smaller than in the case when only one capacitance is used on each side (that is $C_C > C_{C1}$).

7.3 Gain enhancement - gain boosting

Gain boosting, also analyzed in [3], [4], is possible only in the case of a cascode or folded cascode structure. The additional gain stage is added between the gate and the drain of the cascode (folded-cascode) transistor. As a result, the total gain corresponds to the equivalent gain of three intrinsic stages. Figure 7.10 shows the most common way to realize a fully-differential folded cascode OTA with gain boosting.

The gain-boosting amplifier is again a fully-differential folded cascode amplifier, that requires its own CMFB amplifier and a biasing circuitry that provides the common-mode reference voltages at each side. Not only does the circuit complexity become really important, but there are also additional stability requirements. Besides the main feedback loop, there are three CMFB loops that must be checked for stability. The design of such a structure is described in [3].

A less complex circuit is proposed in Figure 7.11. The gain-boosting amplifier is a simple common-source stage, described in chapter 4, added to each cascode structure. There is no need for either the additional CMFB loop or the reference voltages. The only additional design requirement is the extremely careful matching of the common-source structures connected to the left and the right half of the amplifier.

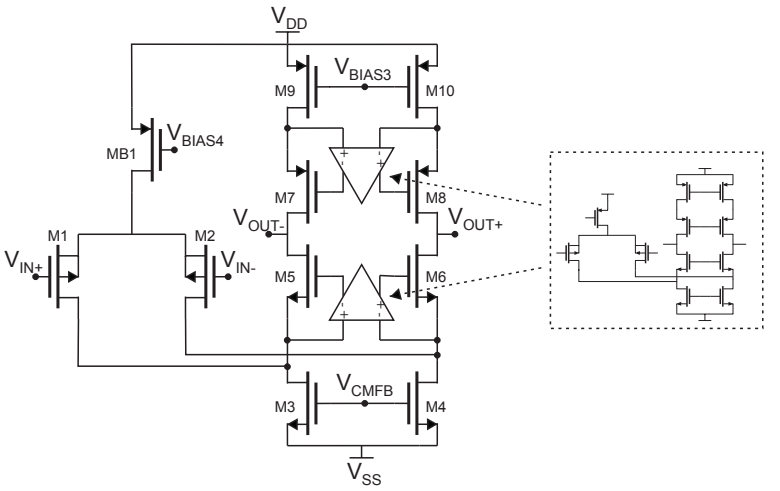


Figure 7.10 Fully-differential folded cascode OTA with gain boosting

The DC gain of the amplifier is

$$A_0 = g_{m1,2} \cdot (A_{0,GB} R_{OUT,OTA}) \approx \frac{g_{m1,2}}{g_{ds1,2}} \cdot \frac{g_{m5,6}}{g_{ds5,6}} \cdot \frac{g_{mD1,2}}{g_{dsD1,2}}. \tag{7.16}$$

Each gain-boosting stage creates a gain loop that minimizes the DC voltage variation in the source of the cascode transistor. The output resistance is, therefore, multiplied by the gain of this loop, that is $1 + A_{0,GB} \approx A_{0,GB}$. Here $A_{0,GB}$ is approximated by the gain of the common-source stage on the lower side, since $R_{OUT,down} \ll R_{OUT,up}$ (see subsection 5.3.2 and subsection 5.3.5).

The amplifier transfer function is a product of the transfer function of the OTA and the transfer function of the gain-boosting amplifier. The module and phase are depicted in Figure 7.12. The dominant pole corresponds to the dominant pole of the gain-boosting amplifier. Because of two possible gain paths, a positive zero appears at the gain bandwidth frequency of the gain-boosting amplifier and creates a zero-pole doublet with the dominant pole of the OTA. The gain bandwidth frequency of such a structure corresponds to the gain bandwidth frequency of the OTA, and thus, the overall stability depends on the non-dominant pole frequency of the OTA.

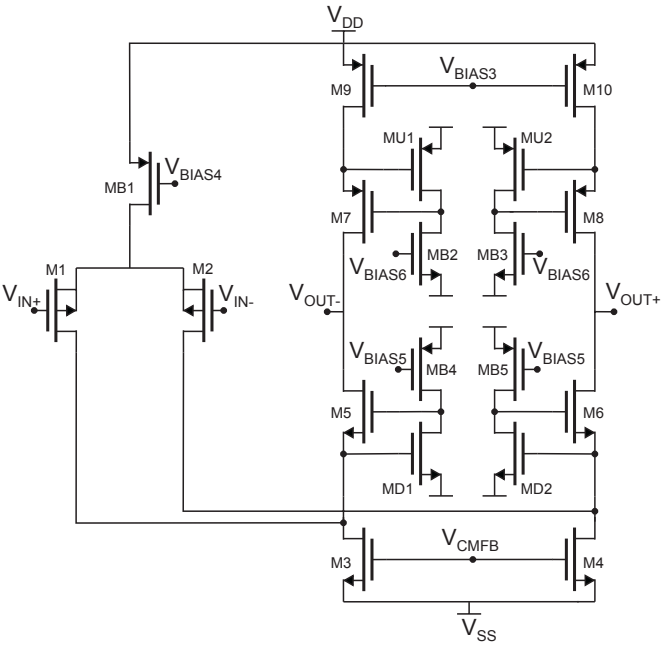


Figure 7.11 Fully-differential folded cascode amplifier with a common-source stage as a gain-boosting amplifier

The presence of a doublet in the transfer function, observed as a bump in the phase characteristics, has important consequences on the time-domain response. Figure 7.13 analyzes the open-loop step response, normalized with DC gain, in three possible situations:

- no doublet in the transfer function is present, only a dominant and a non-dominant pole (dotted line),
- a doublet is present and the zero frequency is lower than the pole frequency (plain black line),
- a doublet is present, but the pole frequency is lower than the zero frequency (plain grey line).

It is obvious that the zero introduces a fast settling component, whereas the pole introduces a slow settling component. Accordingly, the situation where a zero is followed by a pole is more favorable. This imposes an important design condition

$$f_{GBW,GB} < f_{dp,OTA}, \tag{7.17}$$

and also means that the slow settling component influence must be minimized.

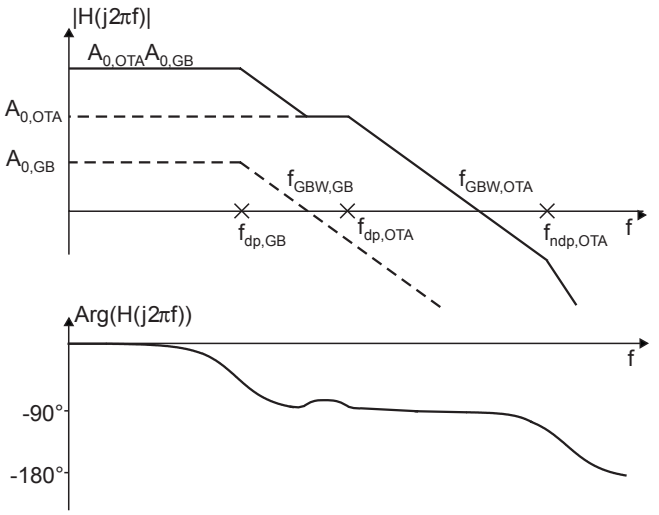


Figure 7.12 Transfer function module and phase of an amplifier with gain boosting

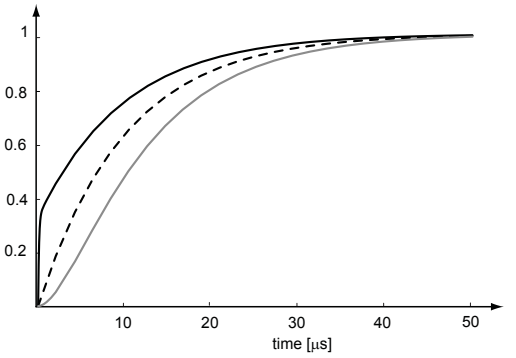


Figure 7.13 Transient response without doublet in the transfer function (dotted line) compared with the transient response when a zero-pole (black line) or a pole-zero (gray line) doublet is present in the transfer function

It is important to notice that $f_{dp,GB} \ll f_{GBW,GB}$ is a poor design choice, because in this case the transient response is determined by the dominant pole of the gain-boosting amplifier. It introduces a large time constant that becomes dominant compared to the other contributions, and this seriously degrades the response delay of the amplifier. The distance between the gain bandwidth and the dominant pole frequency is imposed by the DC gain value, and thus a simple gain stage, such as a common-source stage, with a modest $A_{0,GB}$ is an advantage in terms of time-domain response.

The equivalent half-circuit from Figure 7.14a is used to determine the equivalent resistances and the equivalent capacitances in nodes on signal paths. The approximated frequencies of the dominant pole f_{dp}, the non-dominant pole f_{ndp} and the doublet pole f_p (i.e. dominant pole of the OTA) are given by

$$f_{dp} = \frac{1}{2\pi} \cdot \frac{1}{(R \parallel 1/g_{dsD1,2}) \cdot (\Sigma C)_{GB}} = \frac{1}{2\pi} \cdot \frac{1}{R_{GB} \cdot (\Sigma C)_{GB}}, \tag{7.18}$$

$$f_p = \frac{1}{2\pi} \cdot \frac{1}{R_{OUT,OTA} \cdot C_{OUT,OTA}}, \tag{7.19}$$

$$f_{ndp} = \frac{1}{2\pi} \cdot \frac{n \cdot g_{m5,6}}{(\Sigma C)_{FP}}. \tag{7.20}$$

The sum of the parasitic capacitances in the output node of the gain-boosting amplifier $(\Sigma C)_{GB}$ is increased by the equivalent capacitance seen in the gate of transistor M5, moving the dominant pole slightly forward to the origin. On the other hand, the equivalent capacitance seen in the gate of transistor MD1 contributes to the sum of the parasitic capacitances in the folded point.

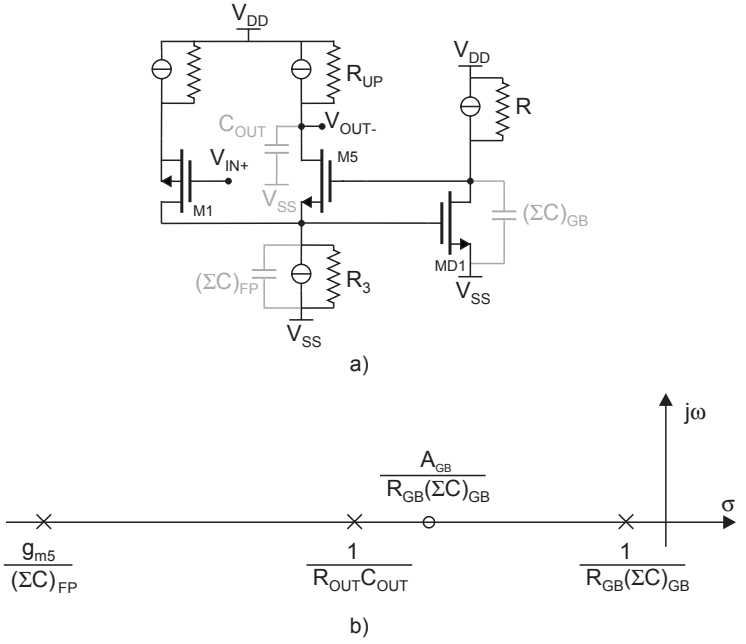


Figure 7.14 Fully-differential folded cascode OTA with common-source gain-boosting stage: a) equivalent half-circuit, b) zero and poles in the s-plane

The latter influence is more critical, since it affects the non-dominant pole position, and thus the overall stability. This finally imposes a requirement to minimize the sizes of the gain-boosting transistors and ensure $f_{ndp} < k_{PM} \cdot f_{GBW}$, where k_{PM} is determined by the specified phase margin, and the gain bandwidth frequency is, according to the previous analysis,

$$f_{GBW} = \frac{1}{2\pi} \cdot \frac{g_{m1,2}}{C_{OUT,OTA}}. \tag{7.21}$$

Finally, the frequency of the zero is obtained when two possible gain paths, i.e. with and without a gain-boosting loop, are superposed in the following way:

$$\begin{aligned} &\frac{g_{m1,2}R_{OUT}}{(1+j\omega/\omega_p)(1+j\omega/\omega_{ndp})} + \frac{g_{m1,2}R_{OUT} \cdot g_{mD1,2}R_{GB}}{(1+j\omega/\omega_p)(1+j\omega/\omega_{ndp})(1+j\omega/\omega_{dp})} \\ &= g_{m1,2}R_{OUT} \cdot g_{mD1,2}R_{GB} \cdot \frac{\left(1+\frac{1}{g_{mD1}R_{GB}}+j\omega/g_{mD1}R_{GB}\omega_{dp}\right)}{(1+j\omega/\omega_p)(1+j\omega/\omega_{ndp})(1+j\omega/\omega_{dp})} \\ &= A_0 \frac{1+j\omega/\omega_z}{(1+j\omega/\omega_p)(1+j\omega/\omega_{ndp})(1+j\omega/\omega_{dp})} \end{aligned} \tag{7.22}$$

Here, the $g_{mD1,2}R_{GB}$ is sufficiently large to consider $1+\frac{1}{g_{mD1,2}R_{GB}} \approx 1$. Thus, it follows that

$$f_z = g_{mD1,2}R_{GB} \cdot f_p = \frac{1}{2\pi} \cdot \frac{g_{mD1,2}R_{GB}}{R_{OUT,OTA} \cdot C_{OUT,OTA}}. \tag{7.23}$$

Another point that must be taken into account, when the gain-boosting stage is added, is the limitation of the differential output swing. As mentioned previously, the DC voltage in the source of a cascode transistor is fixed by the gain-boosting amplifier. In the case of common-source stage, this voltage headroom cannot be lower than one threshold voltage of the NMOS or the PMOS transistor, unless the transistor is placed in "very" weak inversion. However, weak inversion is not favorable because of the large parasitic capacitances that can degrade stability. The swing on each amplifier side is thus

$$\begin{aligned} V_{OUT+,-MIN} &= VTO_N + V_{DSsat5,6} \\ V_{OUT+,-MAX} &= VTO_P + V_{DSsat7,8} \end{aligned} \tag{7.24}$$

and may be a serious limitation when a rail-to-rail output is required.

7.4 Input common-mode range enhancement - complementary differential pair

A large input common-mode range is a typical design requirement for a voltage follower stage. When the differential amplifier, either single ended or fully differential, operates in a unity-gain configuration, the output signal is brought back directly to one of the input terminals, as illustrated in Figure 7.15. Thanks to the negative feedback, the difference between the signals at the amplifier input is minimized. In this way the most faithful copy of the signal coming from the previous stage is provided at the output, and at the same time, this previous stage is isolated from the rest of the system.

The described behavior implies that the voltage-follower amplifier must have a sufficiently large input common-mode range. Even though the differential input signal is small, the input common-mode varies with the amplitude of the signal from the previous stage since both input signals increase or decrease at the same time. The input common-mode range is defined (repeated from subsection 4.4.5) as a common-mode voltage range where the working conditions of the differential pair at the input:

- both transistors operate in saturation, $V_{DS} > V_{DSsat}$,
- the current provided by the bias current mirror is constant,

are respected. Consequently, the gain of the amplifier is constant or within the specified range, and the error between its output and input signal is efficiently minimized.

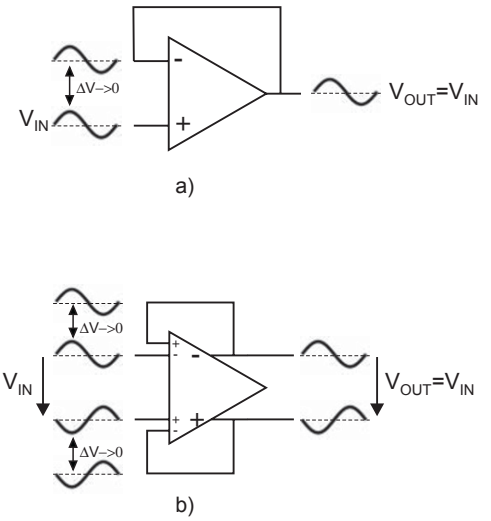


Figure 7.15 Input and output waveforms in the case of: a) single-ended and b) fully-differential voltage follower i

Figure 7.16 shows a fully-differential folded cascode OTA with a PMOS differential pair at the input. When the two input signals increase or decrease, the desired amplifier behavior is ensured while the transconductance is constant, i.e both differential pair transistors are in saturation and have the same inversion factor. On one hand, if the source voltage of the differential pair, imposed by the input common-mode voltage, starts to reduce the voltage headroom of the bias transistor MB1, then the differential pair tail current and the equivalent transconductance start to decrease. On the other hand, when the input common-mode voltage decreases, there is a moment when the transistors MP1,2 constituting the differential pair enter the linear region, and then they no longer behave as a differential pair. Therefore, the maximal input common mode voltage that ensures the saturation voltage headroom of the bias mirror and the desired inversion factor of the transistors of the differential pair is

$$V_{IN,CM,MAX} = V_{DD} - |V_{DSatB1}| - |V_{GS1,2}|\,, \tag{7.25}$$

while the minimal input common mode voltage determined to ensure that the transistors operate in saturation is

$$V_{IN,CM,MIN} = V_{SS} + V_{DSsat3,4} + |V_{DSsat1,2}| - |V_{GS1,2}|, \tag{7.26}$$

and it is approximately equal to that of the negative supply rail. Hence, the amplifier can follow the input signal without degradation only when its amplitude is smaller than ($V_{IN,CM,MAX}$ - V_{CM}).

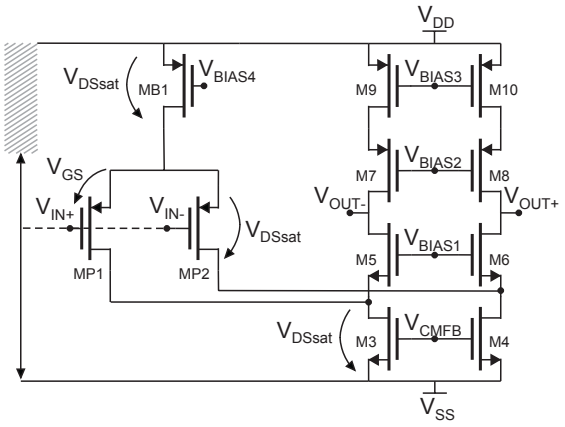


Figure 7.16 Fully-differential folded cascode OTA with a PMOS input pair

An opposite type of situation occurs when the NMOS differential pair is at the input, as depicted in Figure 7.17. The maximal and minimal input common-mode voltages are, in this case, given by

$$V_{IN,CM,MAX} = V_{DD} - |V_{DSsat3,4}| - V_{DSsat1,2} + V_{GS1,2} \qquad (7.27)$$
$$V_{IN,CM,MIN} = V_{SS} + V_{DSatB1} + V_{GS1,2}$$

The limitation is now on the lower side, as on the upper side the signal can reach the positive supply rail voltage. In fact, if the common-mode signal is lower than $V_{IN,\,CM,\,MIN}$ the equivalent transconductance decreases, and the gain of the amplifier becomes too small to follow the input signal without error.

To overcome the input common-mode limitation, a logical solution is to realize a circuit having NMOS and PMOS input pairs connected in parallel, and perform in this way the addition of the equivalent transconductances. Now, when the common-mode voltage is low and the NMOS input pair starts to change the operating region, the PMOS input pair continues to operate correctly, and vice versa. Although the total transconductance changes, it is never lower than the one-pair equivalent transconductance. This means that the gain of the amplifier is sufficiently high to minimize the error and follow the signal coming from the previous stage. The equivalent transconductances of an amplifier with an NMOS input stage, a PMOS input stage, and an NMOS and a PMOS input stage in parallel are compared in Figure 7.18.

The amplifier with an NMOS pair and a PMOS input pair, i.e. complementary input pairs, is displayed in Figure 7.19. For the input common-mode signal in the range

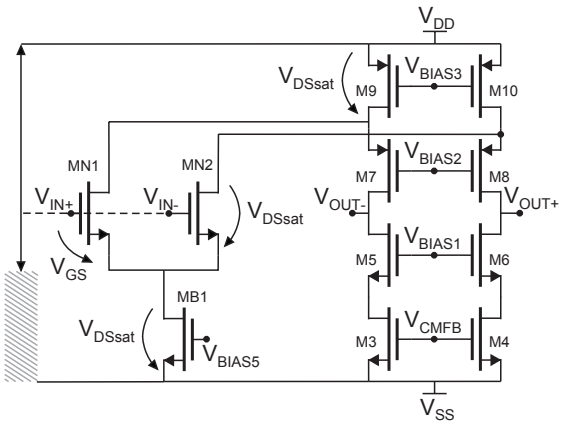


Figure 7.17 Fully-differential folded cascode OTA with an NMOS input pair

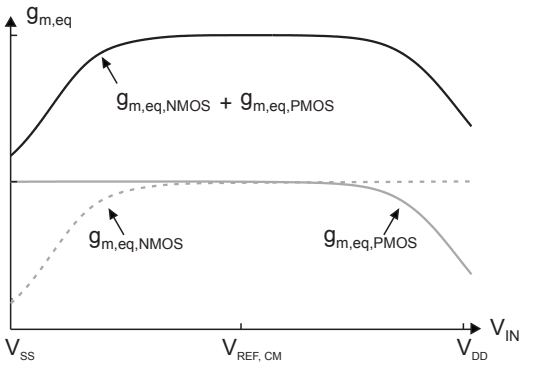


Figure 7.18 Equivalent transconductance of an amplifier having NMOS (dotted gray line), PMOS (full gray line) and complementary (full black line) input pairs

$$V_{SS} + V_{DSsatB2} + V_{GSN} < V_{IN,CM} < V_{DD} - |V_{DSsatB1}| - |V_{GSP}| \tag{7.28}$$

the equivalent transconductance is equal to the sum of the transconductances of the two differential pairs, and its maximal value is $g_{m,eq,NMOS} + g_{m,eq,PMOS}$ ($=2g_m$). On the other hand, in the rail-to-rail input common-mode voltage range, its minimal value is $g_{m,eq,NMOS} = g_{m,eq,PMOS}$ ($= g_m$).
There are two important consequences of such behavior:

- the gain of the amplifier changes from $2g_mR_{OUT}$ to g_mR_{OUT},
- the gain bandwidth frequency changes from $2g_m/(2\pi C_L)$ to $g_m/(2\pi C_L)$,

if it is assumed that the output resistance does not change when the signal swings from negative to positive supply rail. However, the output resistance can not be always considered constant and in some applications this may be an important drawback.

The fact that the gain is not constant is not critical if the unity gain error

$$\varepsilon = \left| 1 - \frac{1}{1 + 1/A} \right|, \tag{7.29}$$

that is the error between the amplifier output signal and the signal to follow, is within the specifications when calculated for the lower gain value, i.e. g_mR_{OUT}. Concerning the stability, it must be ensured for the "maximal" gain bandwidth $2g_m/(2\pi C_L)$ in order to avoid degradation of the phase margin. If this is respected, the complementary input pair makes it possible to achieve the rail-to-rail input common-mode range without degradation of other circuit

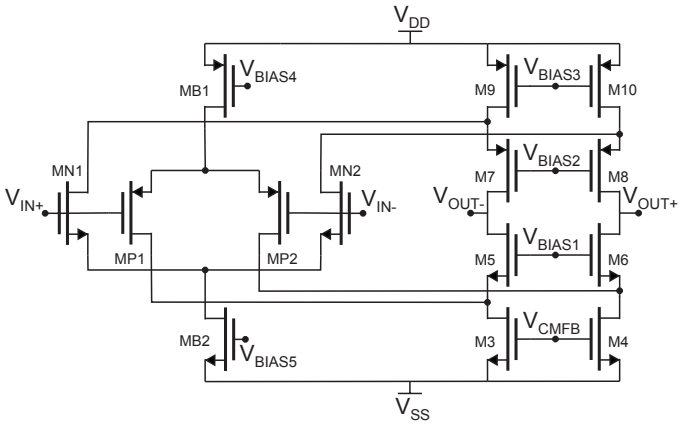


Figure 7.19 Fully-differential folded cascode OTA with NMOS and PMOS input pairs

parameters. However, there exist techniques to achieve a constant g_m behavior, when it is required, and they are presented in [2].

7.5 Differential input range enhancement - linearized differential pair

When a negative feedback network is connected between the input and the output of the amplifier, the differential signal at its input is small due to this negative feedback. On the contrary, the signal driven to the input of a CMFB amplifier is the output signal of the main amplifier, and hence, has a large swing. The common-mode variations can also be present (and are minimized by a CMFB loop), but in the discussion that follows the DC common-mode level is considered constant. Figure 7.20 depicts the typical situation of a differential pair at the CMFB amplifier input.

The important design parameter here is the differential input range. It represents a voltage range where the equivalent transconductance is constant or within a specified error range, and the working conditions of the differential pair are respected. A useful parameter for estimating the differential input range is the saturation voltage of the differential pair, defined in chapter 4 as its design parameter. It depends on the inversion factor, and increases when the transistor operating region is moved toward strong inversion. Nevertheless, sometimes even in strong inversion the required differential input range is not achieved, or the inversion factor is imposed by another specification of higher priority, such as gain. In this case a design trade-off has to be found, since weak

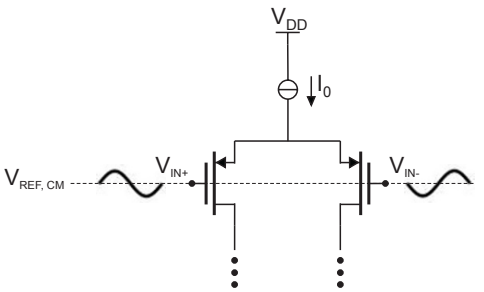


Figure 7.20 Differential pair with a large differential signal swing at the input

inversion ensures maximal gain, whereas strong inversion ensures maximal differential input range.

The efficient solution is a topology variant consisting in a differential pair linearization proposed in [5] and shown in Figure 7.21. The transistors MR1 and MR2, connected in the source of the differential pair, act as an equivalent resistance. When the input voltages change, these transistors first operate in the linear region and then, for large input differences, in the saturation region. The current flowing through MR1 and MR2 is proportional to the difference of the currents flowing through M1 and M2

$$I_{R1} + I_{R2} = \frac{I_{D1} - I_{D2}}{2} = \frac{I_{OUT,DM}}{2}. \tag{7.30}$$

While transistors MR1, MR2 are in the linear region, the equivalent resistance changes with the differential voltage at the input. The voltage headroom between the sources of transistors M1 and M2 is

$$V_{S1} - V_{S2} = R_{eq} \cdot \frac{I_{OUT,DM}}{2}, \tag{7.31}$$

and the source voltages V_{S1} and V_{S2} are a function of the differential voltage at the input and the differential current at the output.

In the case of the classical differential pair, when the voltage difference at the input is small, the voltage in the source of the differential pair is constant. When this small differential voltage at the input increases, the inversion factor of one transistor increases, and the inversion factor of the other transistor decreases by the same amount. The equivalent transconductance

$$g_{m,eq} = \frac{d(I_{D1} - I_{D2})}{dV_{IN,DM}} = I_S \cdot \frac{d(IF_1 - IF_2)}{dV_{IN,DM}} \tag{7.32}$$

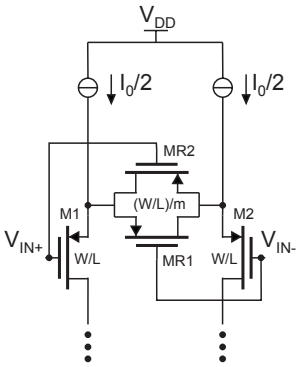


Figure 7.21 The linearization of the differential pair

is constant. However, for a large input difference, the transistor that conducts more current imposes the voltage in the source, and the slope d$(IF_1\text{-}IF_2)$/ d$V_{IN,DM}$ is no longer constant.

If the linearization transistors are present, there is a kind of negative feedback. When the input differential voltage increases, the differential current at the output increases. On the other hand, the equivalent resistance diminishes, and the source voltages V_{S1} and V_{S2} change slightly. Consequently, the slope d$(IF_1\text{-}IF_2)$/d$V_{IN,DM}$ stays constant as long as the transistors MR1 and MR2 are in the linear region and behave as the equivalent resistances. When they start to operate in saturation, the equivalent resistance becomes independent of the differential voltage at the input, and the equivalent transconductance is not constant.

Figure 7.22 shows the equivalent transconductance of a linearized differential pair operating in moderate inversion for different sizes of the transistors MR1,2 compared to the equivalent transconductance of the same differential pair without linearization. The scaling factor m is usually in the range 1-10, and the linearization transistors are sized as

$$\left(\frac{W}{L}\right)_{R1,2,3,4} = \left(\frac{W/m}{L}\right)_{1,2,3,4} \tag{7.33}$$

that is with regard to the sizes of the differential pair transistors. The larger the value of m, the larger the differential input range. However, the optimal value of the scaling factor m is chosen for a variation of the equivalent transconductance that is within the specifications. (In the example from Figure 7.22, an optimal value would be $m = 3$.)

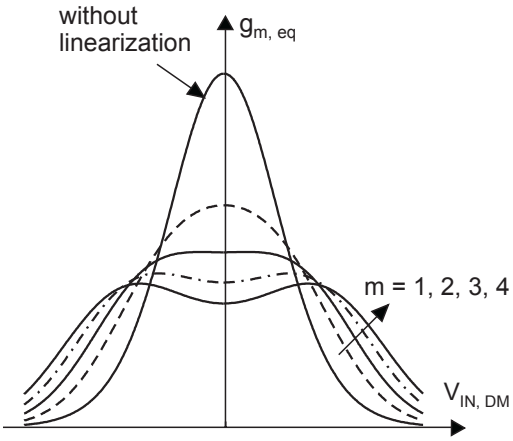


Figure 7.22 Differential input range without/with the differential pair linearization

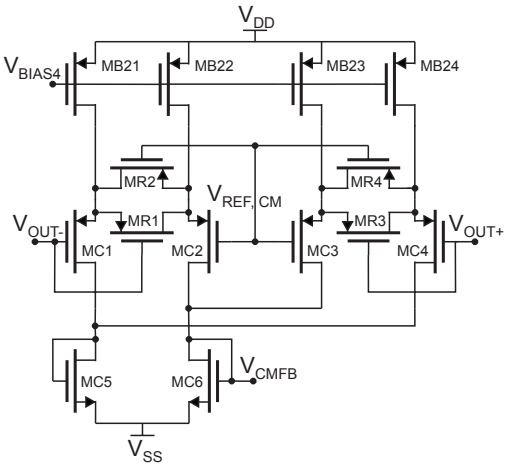


Figure 7.23 CMFB amplifier with linearized differential pairs

An example of a CMFB amplifier with linearized differential pairs is shown in Figure 7.23. The linearization can be applied to a differential pair operating in weak, moderate or strong inversion and always results in an improvement of the differential input range. However, an optimal scaling factor must be found in each case. In addition, there is also the following implementation requirement: all transistors (differential pair transistors and linearization transistors) must have the bulk connected to the same voltage, i.e the positive supply voltage in the case of PMOS transistors and the negative supply voltage in the case of NMOS transistors.

7.6 Rejection of the differential signal in a common-mode signal - cascoded current bias sources

The input stage of a CMFB amplifier consisting of two differential pairs is displayed in Figure 7.24. Each differential pair sees the common-mode reference voltage on one side and one output signal of the main amplifier on the other side, as illustrated in the same figure. Consequently, the input common-mode signal of each differential pair is

$$V_{CM,IN,CMFB+,-} = \frac{V_{REF,CM} + V_{OUT+,-}}{2}. \tag{7.34}$$

Since the main amplifier output signals have a large swing, the input common-mode signals $V_{CM,IN,CMFB+,-}$ follow these variations. The source voltages V_{S1} and V_{S2}, determined by a common-mode level, also change. Since the amplifier output signals are opposite in phase, the input common-mode signals, as well as the source voltages V_{S1} and V_{S2}, change in the opposite way. Because of these voltage variations, the bias current of one differential pair increases, while the bias current of the other differential pair decreases, in the case when the bias mirrors are realized as simple current mirrors (Figure 7.24). Therefore, the differential signal is "copied" into the bias currents which now change instead of being constant.

These bias currents are partitioned and summed by two differential pairs and determine the CMFB amplifier output voltage, which controls the common-mode level at the output of the main amplifier

$$V_{CM,OUT} = \frac{V_{OUT+} + V_{OUT-}}{2}. \tag{7.35}$$

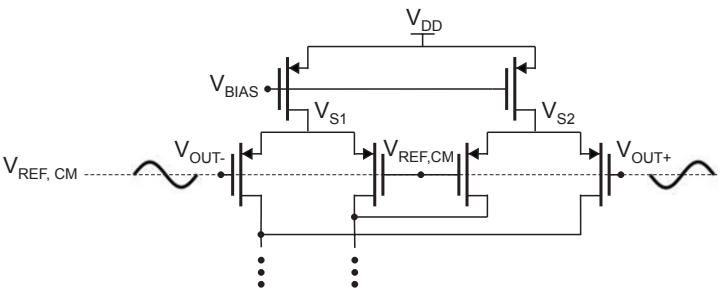


Figure 7.24 Two differential pairs at the CMFB amplifier input and typical input waveforms

Thus, the common-mode regulation loop is closed with the following result: the output common-mode signal is proportional to the differential output signal, and it becomes an AC instead of DC signal.

The solution is to cut the perturbation propagation chain at the most critical point by cascoding the bias current mirrors. Now, the variations of voltages V_{S1} and V_{S2} do not affect the bias currents and, as long as these currents stay constant, the differential signal is not copied into the common-mode output signal of the main amplifier. The modified topology is depicted in Figure 7.25.

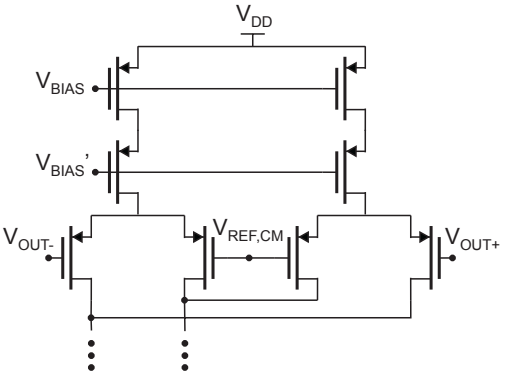


Figure 7.25 CMFB input stage with cascoded current bias sources

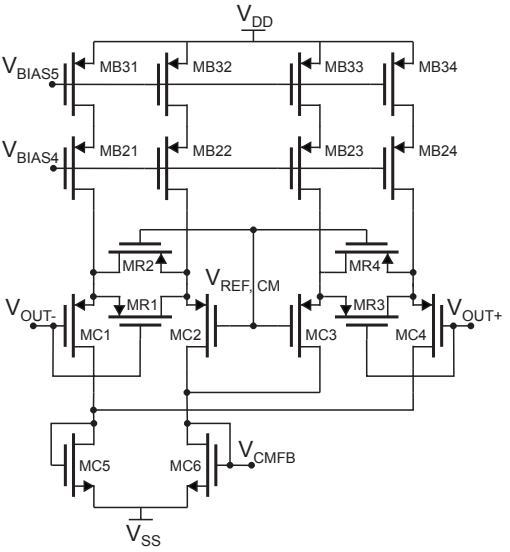


Figure 7.26 Complete schematic of the CMFB amplifier with a linearized differential pair and cascoded current sources

The complete schematic of the CMFB amplifier including the linearized differential pair for the differential input range enhancement and the cascoded bias sources for the rejection of the differential signal in the common-mode signal is shown in Figure 7.26.

7.7 CMFB stability improvement - split bias sources

As discussed in chapter 5, the CMFB loop transfer function inherits the dominant and non-dominant poles of the main amplifier because it uses one part of its gain path. Moreover, there exists an additional non-dominant pole due to the equivalent resistance and the equivalent capacitance seen in the output node of the CMFB amplifier. The module of the main feedback loop transfer function and the module of the open CMFB loop transfer function for the basic fully-differential folded cascode OTA topology are shown in Figure 7.27. The pole approximations are repeated from chapter 5 in Table 7.2.

The sum of the capacitances denoted as $(\Sigma C)_{CMFB}$ represents the equivalent capacitance seen in the output node of the CMFB amplifier and includes the capacitance seen in the gate of the transistor MC6 and in the gates of the transistors M3 and M4. Since the transistor MC6 is often at the limit of strong inversion, and the transistor M5 is at the limit of weak inversion, it follows that $g_{m5} > g_{mC6}$. Consequently, the second non-dominant pole in the CMFB loop is usually closer to the origin than the non-dominant pole of the OTA, as depicted in Figure 7.27. Now, if the gain of the CMFB loop is high, a stability problem may occur.

Table 7.2 Approximations of the poles in the transfer function of the main feedback loop and the CMFB loop for the basic fully-differential folded cascode OTA topology

	main amplifier (OTA)	CMFB loop
dominant pole $f_{nd,OTA}$	$\frac{1}{2\pi} \cdot \frac{1}{R_{OUT} \cdot C_L}$	$\frac{1}{2\pi} \cdot \frac{1}{R_{OUT} \cdot C_L}$
non-dominant pole $f_{nd,OTA} = f_{nd2,\ CMFB}$	$\frac{1}{2\pi} \cdot \frac{g_{m5,6}}{(\Sigma C)_{FP}}$	$\frac{1}{2\pi} \cdot \frac{g_{m5,6}}{(\Sigma C)_{FP}}$
non-dominant pole $f_{nd1,\ CMFB}$	-	$\frac{1}{2\pi} \cdot \frac{g_{mC6}}{(\Sigma C)_{CMFB}}$

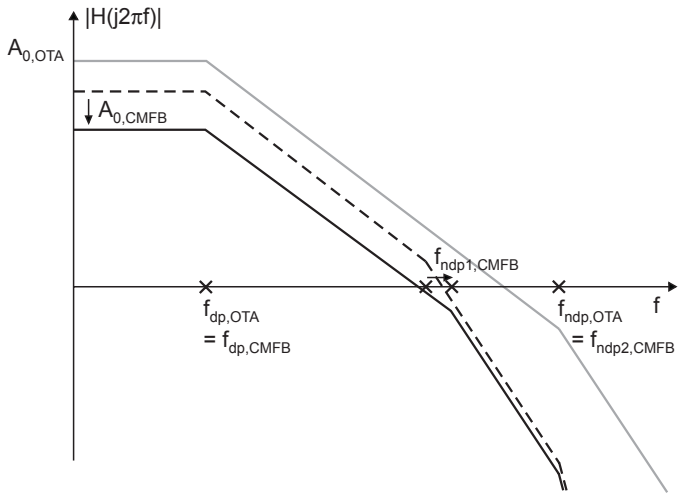


Figure 7.27 Transfer function module of the main amplifier (gray line) and the CMFB loop with (black line) and without (black dotted line) topology modification

The gain of the CMFB loop is approximated as

$$A_{o,CMFB} = \frac{g_{mc1,2,3,4}}{g_{mc6}} \cdot \frac{k \cdot g_{mc6}}{g_{m5,6}} \cdot g_{m5,6} \cdot R_{OUT} = k \cdot g_{mc1,2,3,4} R_{OUT} \tag{7.36}$$

where k is the ratio of the bias currents, that is

$$I_3 = I_{Dsat3,4} = kI_{DsatC6} = kI_{CMFB}. \tag{7.37}$$

The ratio of the transconductances is the same as the ratio of the currents, that is $g_{m3,4} = k \cdot g_{mc6}$, since the transistors MC6 and M3,4 have the same bias voltage, and thus have the same inversion level.

One way to improve the stability of the CMFB loop is to reduce the loop gain. This can be done by moving transistor MC6 to operate in strong inversion, thus decreasing the transconductance g_{mc6}. Nevertheless, the consequences are the following: the differential input range is improved, but the rejection ratios (CMRR, PSRR) are reduced.

The other possibility is to diminish the sum of the parasitic capacitances $(\Sigma C)_{CMFB}$ and to move the second non-dominant pole to higher frequencies. In this case, the transistor MC6, and hence the transistors M3 and M4, must be re-designed to have smaller sizes. As a result, the output resistance and the differential output swing are degraded, and, in most cases, this is not acceptable.

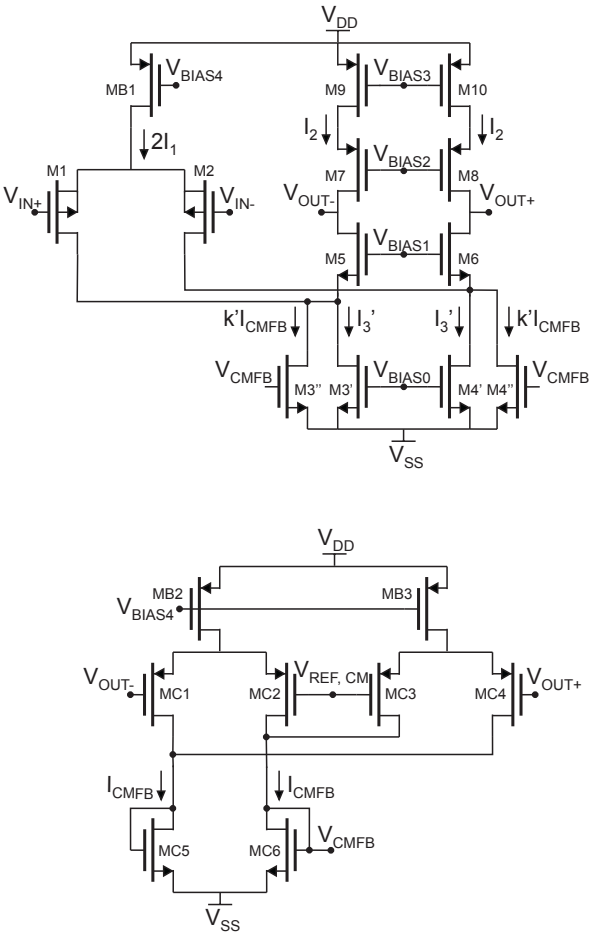


Figure 7.28 Fully-differential folded cascode OTA with split bias sources and its CMFB circuit

The topology modification can be the best "CMFB loop gain versus parasitic capacitances" compromise. It consists in decreasing the current bias ratio by splitting the bias current source controlled by the CMFB control voltage V_{CMFB} into two sources (Figure 7.28). The first source is controlled by a fixed bias voltage V_{BIAS0} and provides a large current I_3', while the second source is biased by V_{CMFB} controlling the common-mode voltage at the OTA output and provides the minimal required current $k'I_{CMFB}$. Since now $I_3 = I_3' + k'I_{CMFB}$, the transistors M3 and M4 are, in fact, split in two, and there is no need to re-design any transistor. The reduction of the bias current ratio reduces the gain and moves the additional non-dominant pole to higher

frequencies at the same time, as illustrated in Figure 7.27. The sum of the parasitic capacitances $(\Sigma C)_{CMFB}$ is minimized since only the gates of the transistors M3'' and M4'' are connected to the CMFB amplifier output. On the other hand, the gain of the OTA is unchanged, as well as its non-dominant pole.

An important drawback of the proposed topology variation is the slower time response of the OTA, as there is less current to control the common-mode voltage at the output. If this is a critical design requirement, then the minimal bias current ratio k' has to be chosen with regard to the response delay specification.

7.8 Conclusion

This chapter presents the concept of topology variants in the example of a fully-differential folded cascode OTA and its CMFB amplifier. In each case, the design parameter that has been improved due to the topology modification is identified, and the design guidelines, as well as the design drawbacks, are presented. However, the presented approach can be used as a solution of design problems in the case of any other analog amplifier.

Bibliography

[1] P. Allen, D. Holberg, “CMOS Analog Circuit Design”, Oxford University Press, second edition, January 2002, ISBN10: 0195116445

[2] R. Hoge, R. Hogervorst, J. Huijsing, “Design of Low-voltage Low Power Operational Amplifier Cells”, Springer, first edition, April 2002, ISBN10: 0792397819

[3] K. Bult, G. Geelen, “A fast settling CMOS op amp for SC circuits with 90-dB DC gain”, Journal of Solid-State Circuits, Volume 25, pp. 1379-1384, December 1990

[4] B. Razavi, “Design of Analog CMOS Integrated Circuits”, McGraw-Hill Science/Engineering/Math, first edition, August 2000, ISBN: 0072380322

[5] F. Krummenacher, N. Joehl, “A 4-MHz CMOS continuous-time filter with on-chip automatic tuning”, Journal of Solid-State Circuits, Volume 23, pp. 750-758, June 1988

CHAPTER 8

Practical example: the design of analog amplifiers in the Delta-Sigma modulator system

In this chapter, three analog amplifiers that are parts of a hybrid multi-bit Delta-Sigma modulator are designed starting from the system-level specifications. It is demonstrated how to take into account the environment of an amplifier in the system, determine an appropriate design testbench, use the design flow described in the previous chapters, and evaluate the results. In each case, the fully differential folded cascode OTA is chosen as the most suitable topology to fulfill high gain and high gain bandwidth design requirements. However, topology variants were necessary in order to achieve the additional design requirements, whereas the procedural design scenario was the same for all three amplifiers.

8.1 Delta-Sigma modulator system

The design flow from system description to transistor level is illustrated here using as example three fully differential amplifiers. The amplifiers are designed as a part of a low-pass second-order hybrid multi-bit Delta-Sigma modulator depicted in Figure 8.1. The design of such a Delta-Sigma modulator has been described at the system level by S. Pesenti [1]. The approach used to design its analog parts represents a practical demonstration of the structured analog design approach discussed in the previous chapters.

The main strategy for the design of the aforementioned second-order multi-bit Delta-Sigma modulator is to lower total power consumption. Therefore, the system is realized as a hybrid structure. The first integrator is designed as a continuous-time stage that reduces significantly power consumption compared to the usual switched-capacitor implementation. The second integrator is realized as a switched-capacitor stage, where the transition from continuous to discrete-time domain is realized by a full-clock-cycle-sampling scheme.

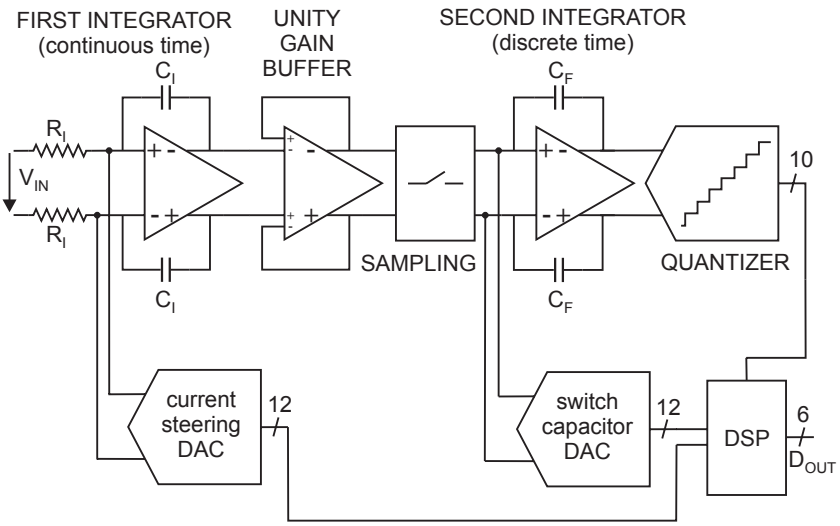


Figure 8.1 Low-pass second-order hybrid multi-bit Delta-Sigma modulator topology

The sampling stage has two sets of capacitors that alternately charge and discharge providing in this way twice as much time for the signal to settle. This obviously lowers the current sourcing/sinking capability requirement of the first stage.

With the use of the signal prediction algorithm proposed by S. Pesenti in [1], the number of comparators is reduced from 32 to 10. Not only the total power consumption and quantizer complexity are reduced, but also the equivalent capacitive load of the second integrator is decreased, and consequently its current sourcing/sinking capability requirement.

The use of a multi-bit technique [2] increases the number of internal levels and reduces the voltage steps. The main benefits are reduced jitter sensitivity of the first stage and reduced sampling frequency. Moreover, the specifications for the design of analog circuitry are relaxed in terms of response delay.

System-level specifications - Table 8.1 gives an overview of the most important system-level specifications.

System metric - The parameter used to evaluate the performance of such a system is the signal to noise-and-distortion ratio (*SNDR*). The target value, in this practical example, is 83 dB where the following contributions are taken into account:

- Delta-Sigma modulator quantization noise,
- first integrator thermal noise,
- second integrator switch capacitor noise, and

Table 8.1 System specifications for the Delta-Sigma modulator design

CMOS technology	0.18 μm
Voltage supply	1.8 V
Current consumption	< 3.5 mA
Differential input amplitude	0.5 V
Signal bandwidth	0.5 MHz
Sampling frequency	32 MHz
Internal quantizer level number	11
Emulated quantizer level number	32

- all non-linearity contributions.

However, for simulation purposes the target *SNDR* value is 93 dB, calculated with regard to the quantization noise and the non-linearity contributions only.

Passive elements in the system - According to the system-level specifications the passive elements are determined in [1] as follows:

- first integrator capacitance: $C_I = 10.8$ pF and resistance: $R_I = 6$ KΩ,
- sampling capacitance: $C_S = 1.05$ pF,
- second integrator capacitance: $C_F = 0.2$ pF,
- and switch-capacitor DAC element capacitance: $C_D = 25$ fF.

8.1.1 Design flow

The three different analog amplifiers implement the following analog functions of the Delta-Sigma modulator:

- continuous-time integrator,
- discrete-time integrator, and
- voltage follower.

According to the system schematic in (Figure 8.1) each amplifier is a fully-differential structure. The unity-gain buffer as a voltage follower minimizes the difference between the differential signal at the output and the differential signal at the input, and thus, in the fully-differential case is realized as a difference differential amplifier [3].

Extraction of the specifications - The specification sets are derived from the system specifications in Table 8.1 and from the time-domain simulations. Figure 8.2 shows the behavioral model of the amplifiers in both the first and the second integrators. The fully-differential difference amplifier is modeled as

in Figure 8.3. The CMFB circuit, necessary for fully-differential structures, is modeled as ideal and added at the output of each model.

Since the main design concern is the current consumption, the current sourcing/sinking capability at the output becomes an important design parameter and is determined carefully. The use of the g_m-cell facilitates this task and makes it possible to extract the required value accurately.

If all other blocks in the system are modeled as ideal, then the extracted specifications are too optimistic. The simulation results, when this information is used as a design starting point, are not satisfactory. The reason for the degradation is that some interaction problems are impossible to predict if the amplifier is surrounded by blocks with ideal behavior. Therefore, the specifications for the design of each amplifier are derived correctly only after all surrounding blocks are designed at the transistor level. This also implies a certain system-level design sequence that must be respected. The final specifications are presented together with the testbench circuits in section 8.2.

System-level design sequence - The design sequence for the Delta-Sigma modulator analyzed in this example is: current-steering DAC, sampling stage, unity-gain buffer, first integrator, switch-capacitor DAC, quantizer, and second integrator.

Derivation of the testbench - To determine a design testbench it is mandatory to:

- identify the negative feedback network,
- estimate the equivalent load at the output,
- identify the characteristics of differential and common-mode signals at the input.

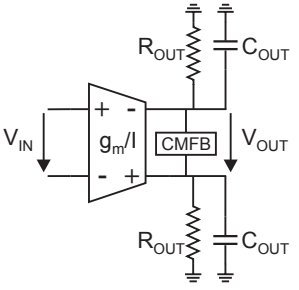


Figure 8.2 Behavioral model of the fully-differential transconductance amplifier

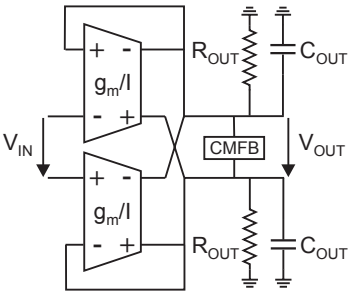


Figure 8.3 Behavioral model of the fully-differential difference amplifier

At this point, the analysis types and simulation sets are also defined. Further discussion on this follows in section 8.2.

Amplifier design - Topology selection and topology variants at the level of basic analog structures are discussed in section 8.3. The procedural design flow and optimization steps are discussed separately for each amplifier in sections 8.4, 8.5, 8.6, respectively.

8.1.2 Verification of the performance

The performance of each block designed at the transistor level is confirmed by time-domain system simulation. The *SNDR* value is computed from the simulation results for at least 20 periods and then compared to the target value of 93 dB. The acceptable variation of this value due to simulator precision and calculation tolerance is fixed at ±2 dB. In addition, the interactions between two or more successive blocks designed at the transistor level are checked. The most important simulation chains are:

- first integrator and unity gain buffer,
- complete first stage: current steering DAC, first integrator and unity-gain buffer,
- unity-gain buffer and sampling circuit,
- second integrator and quantizer,
- sampling circuit, second integrator and switch-capacitor DAC.

Simulation results - In spite of satisfactory *SNDR* values when each amplifier is tested alone, the simulations of several blocks at the transistor level always require design iterations. The main issues are discussed hereafter.

The problem in the interaction between the first integrator and the unity-gain buffer is the presence of a differential signal in the common-mode signal at the output of first integrator. This non-ideality gravely perturbs the buffer stage if the amplitude of the common-mode variation is too high or the *CMRR* of the buffer stage is too low. Therefore, two additional specifications have to be extracted:

- maximal acceptable variation of the common-mode signal at the output of the integrator stage,
- and minimal acceptable *CMRR* of the buffer stage.

To resolve this issue, the topology of the common-mode feedback circuit, which is the main contributor, is modified as presented in section 8.3. On the other hand, the current bias circuit of the input pair of the buffer stage is re-designed. The parameters of both circuits are fine-tuned in several iterations.

The main difficulty in the interaction of the amplifier in the second integrator with its environment is to determine precisely the maximum current needed at the output. The use of a behavioral model allows us to re-extract the

correct specification for the output current only if the quantizer, sampling circuit and switch-capacitor DAC are completed at the transistor level. The amplifier re-design, however, requires some optimization steps, which are analyzed in section 8.4.

The simulation results for the analog chain consisting of the first integrator, unity-gain buffer, current steering DAC, sampling stage, and the second integrator (all with the final transistor sizes) are presented in Table 8.2. The system is simulated in two cases: with schematics at the transistor level and with schematics extracted from the layout. The calculated *SNDR* values are within the acceptable range.

Figure 8.4 displays the compete layout of the Delta-Sigma modulator realized in 0.18 μm CMOS technology.

Table 8.2 Simulation results for the analog chain in the Delta-Sigma modulator

analog chain	*SNDR* [dB]
schematic at the transistor level	91.1
schematic extracted from the layout	90.4

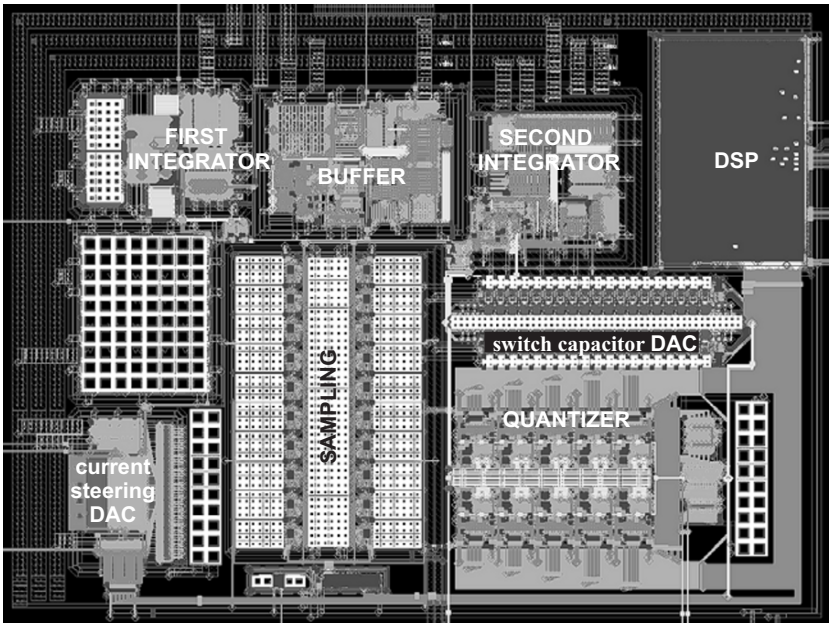


Figure 8.4 Delta-Sigma modulator layout (size: 1 mm x 0.75 mm)

8.2 Derivation of the testbench

A Delta-Sigma modulator is a complex system consisting of analog, mixed-mode and digital blocks. Besides the system specifications, the non-idealities of each block determine the required behavior of neighboring blocks. Accordingly, the testbench derivation for each block is mandatory.

The design of an analog amplifier is possible only in the frame of its testbench. After the initial transistor sizes have been calculated, all required simulations in the AC and DC domains are performed using the testbench circuits. This also holds for all optimization or fine-tuning steps. Time-domain system simulation is only a final verification step.

The design testbench is correctly defined if all surrounding blocks are implemented at the transistor level. Hence, the system-level design sequence has to be respected absolutely. The testbench derivation for the amplifiers in the Delta-Sigma modulator follows the order of the design sequence.

8.2.1 Fully-differential difference amplifier

The unity-gain buffer acts as a kind of isolation stage between the continuous-time integrator and the sampling stage which converts the signal to the discrete-time domain. Thus, the design requirements for the amplifier in the first integrator are relaxed in terms of response delay and current sourcing/ sinking capability. The continuous-time signal is copied to the output of the buffer stage that must provide sufficient output current to charge/discharge the sampling capacitor within a certain delay.

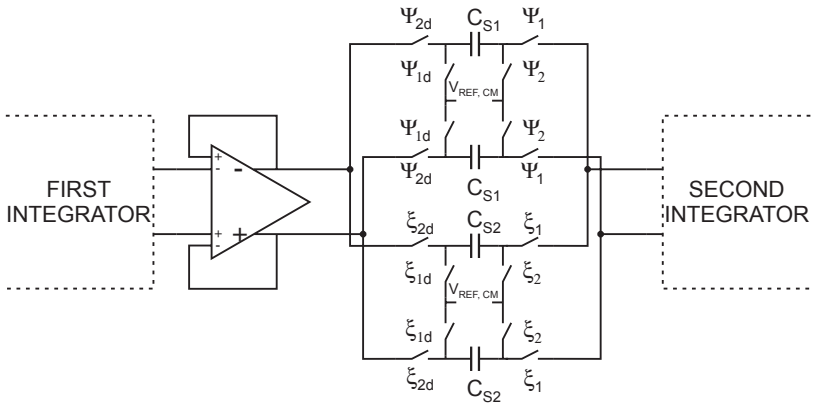
The sampling stage is realized with two capacitance pairs as depicted in Figure 8.5. When control voltages ψ_2 and ψ_{2d} are high, the first pair of capacitances C_{S1} is connected to the buffer output and the integration result from the first stage is sampled. At the same time, when the control voltages ξ_1 and ξ_{1d} are high, and the second pair of the capacitances C_{S2}, which are charged previously, is connected to the second integrator input. In the next half of the clock cycle, the capacitances switch roles. As a result, the capacitive charge at the buffer output is either the sampling capacitance C_{S1} or the sampling capacitance C_{S2}, except for a very short lapse of time when the sampling switches change state. The design testbech is, thus, as simple as in Figure 8.6, with an additional design condition: the stability must be ensured with and without a load capacitor.

The main design requirement is linearity; the differential output signal must track the differential input signal in a defined ± 0.5V signal range within an error fixed at 0.1%. The output current and the gain are determined from the behavioral simulation, whereas the unity-gain bandwidth is chosen to be at

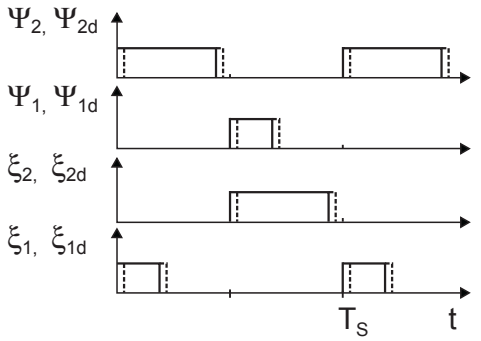
least two times greater than the sampling frequency f_S. All specifications are listed in Table 8.3. The indicated minimal input common-mode rejection ratio is determined after the first integrator stage has been designed.

The simulation set is the following:

- AC analysis (closed-loop transfer function, open-loop stability, CMFB loop stability, *CMRR* in signal range),
- DC analysis (equivalent transconductance, gain error, differential output swing),
- and transient analysis in the system (*SNDR*).



a)



b)

Figure 8.5 a) Unity-gain buffer followed by the sampling stage and b) time diagram of the control voltages of sampling switches

Table 8.3 Specification list for the fully-differential difference amplifier

current consumption	$\Sigma I < 450\ \mu A$
sourcing/sinking capability at the output	$I_{OUT} > 100\ \mu A$
gain	$A_0 > 70$ dB
unity- gain bandwidth frequency	$f_{GBW} > 70$ MHz
open loop stability (with/without load)	$PM > 70°$
CMFB loop stability	$PM > 70°$
CMRR	> 40 dB
DC gain error	$\varepsilon < 0.1\%$
differential output swing	$\Delta V_{OUT} > \pm 0.5$ V

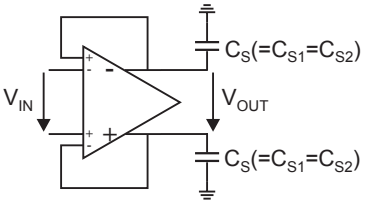


Figure 8.6 Design testbench for the difference differential amplifier

8.2.2 Fully-differential amplifier in the first integrator

The first stage integrates the difference between the input signal (converted to current through resistance R_I) and the signal coming from the modulator feedback loop (sum of the currents sourced/sunk by the current steering DAC elements). The design testbench is drawn in Figure 8.7. The influence of the current-steering DAC is neglected, because it behaves as a current source/sink. The load capacitance $C_{IN,\ BUFFER}$ represents the total parasitic capacitance at the buffer input and is equal to 0.1 pF.

The design specifications, summarized in Table 8.4, are mainly calculated from the system-level specifications. The highest frequency, for which the integration function has to be ensured, is chosen to be at least two times greater than the sampling frequency f_S. The acceptable variation of the output common-mode signal is determined only when the interaction with the buffer stage is tested.

Table 8.4 Specification list for the design of the fully-differential amplifier in the first integrator

integrator frequency	f_I = 2.6 MHz
integration end frequency	$f_{I,\,end}$ > 70 MHz
current consumption	ΣI < 200 μA
gain	A_0 > 70 dB
main feedback loop stability	PM > 70°
CMFB feedback loop stability	PM > 70°
differential output swing	ΔV_{OUT} > ±0.5 V
output common-mode variation	$\Delta V_{OUT,\,CM}$ > ±5 mV

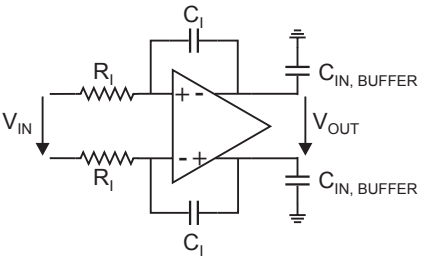


Figure 8.7 Design testbench for the fully-differential amplifier in the first integrator

The design verification consists of:

- AC analysis (closed-loop transfer function, stability of the main feedback loop, stability of the CMFB loop),
- DC analysis (differential output swing),
- and transient analysis in the system (output common-mode variation, *SNDR*).

8.2.3 Fully-differential amplifier in the second integrator

Figure 8.8 displays the second integrator with the detailed schematic of the switch-capacitor DAC and the quantizer consisting of 10 comparators. The second stage integrates the difference between the feedforward signal (charge of C_S) and the signal coming from the second feedback loop of the modulator (sum of the charges of the DAC-element capacitances C_D). The integration is

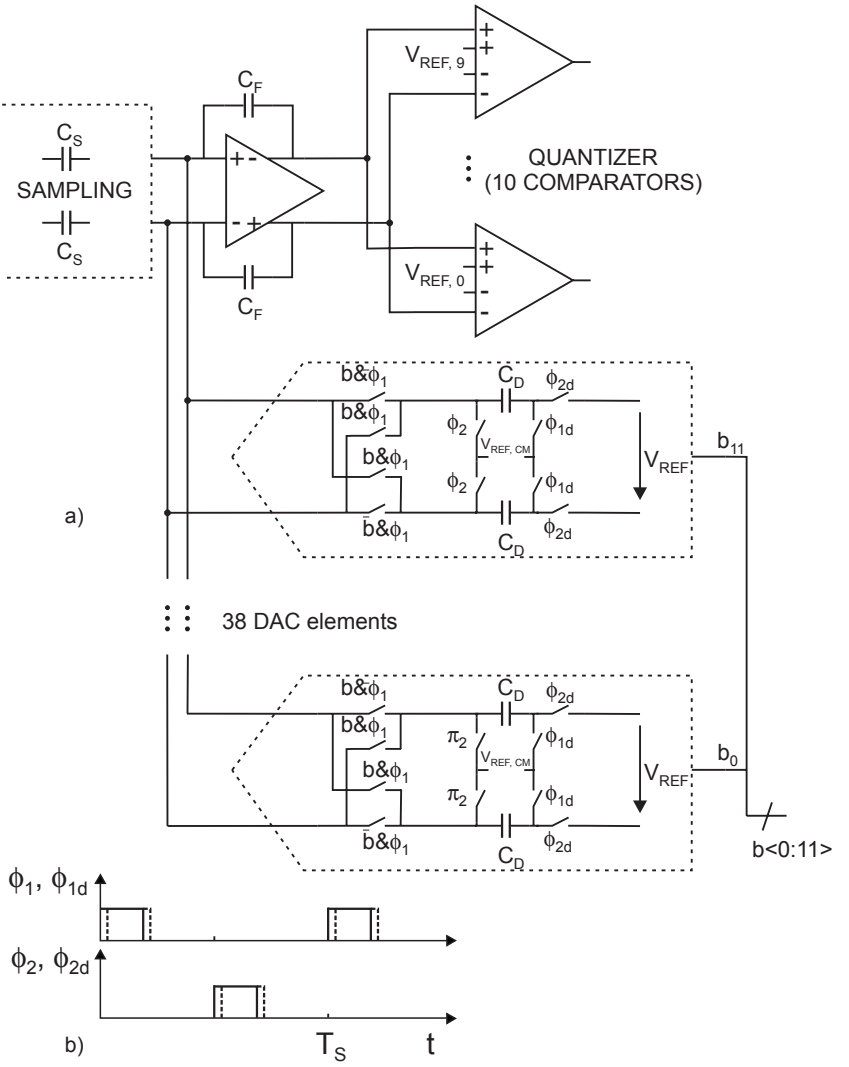


Figure 8.8 a) Second integrator, switched-capacitor DAC and quantizer and b) time diagram of the control voltages of DAC switches

performed only once per clock cycle, when both Φ_1 and Φ_{1d} control voltages are high. The design testbench approximates the negative feedback capacitive network during the integration as shown in Figure 8.9. The amplifier load capacitance is the sum of the input parasitic capacitances of the comparators constituting the quantizer, and its value is equal to 10 x 7 fF = 70 fF.

Table 8.5 Specification list for the fully-differential amplifier in the second integrator

current consumption	$\Sigma I < 1.2$ mA
sourcing/sinking capability at the output	$I_{OUT} > 190$ μA
response delay	$t_D < 5$ ns
equivalent input transconductance	$g_m \approx 2$ mS
R_{OUT} in the output range	> 3.2 MΩ
main feedback loop stability	$PM > 70°$
CMFB loop stability	$PM > 70°$
differential output swing	$\Delta V_{OUT} > \pm 0.5$ V

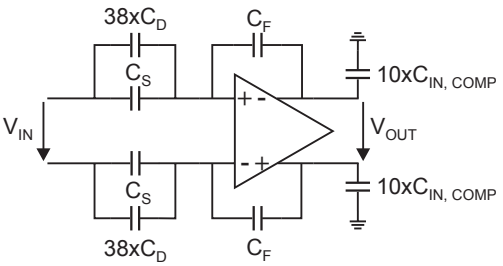


Figure 8.9 Design testbench for the fully-differential amplifier in the second integrator

Table 8.5 gives an overview of the specifications. The main design requirement is a fast time-domain response, since the output signal settling time must be shorter than the integration cycle. This, in fact, imposes a high output current sourcing/sinking capability.

The minimal required current at the output, as well as the equivalent transconductance and the output resistance are extracted from the behavioral simulations. Due to large circuit currents, the output resistance becomes critical because of its non-linear behavior. The more accurate modeling of this behavior permitted us to extract the minimum required value in the output range. The design analysis includes:

- AC analysis (open-loop stability, CMFB loop stability),
- DC analysis (differential output swing, output resistance variation),
- and transient analysis in the system (*SNDR*).

8.3 Topology selection

Careful examination of the specification sets for the three amplifiers shows that there are four identical design requirements:

- gain (A_0 > 70 dB),
- gain bandwidth frequency (f_{GBW} > 70 MHz, or even higher in the case of the second integrator),
- main feedback loop stability (PM > 70°),
- and CMFB loop stability (PM > 70°).

The most suitable candidate to achieve this gain-gain bandwidth-stability design requirement is the fully-differential folded-cascode OTA depicted in Figure 8.10. Its most important property, analyzed in chapter 5, is that it is a single-stage amplifier providing a gain equivalent to two stages and, since there is no compensation capacitance, a frequency behavior of a single stage.

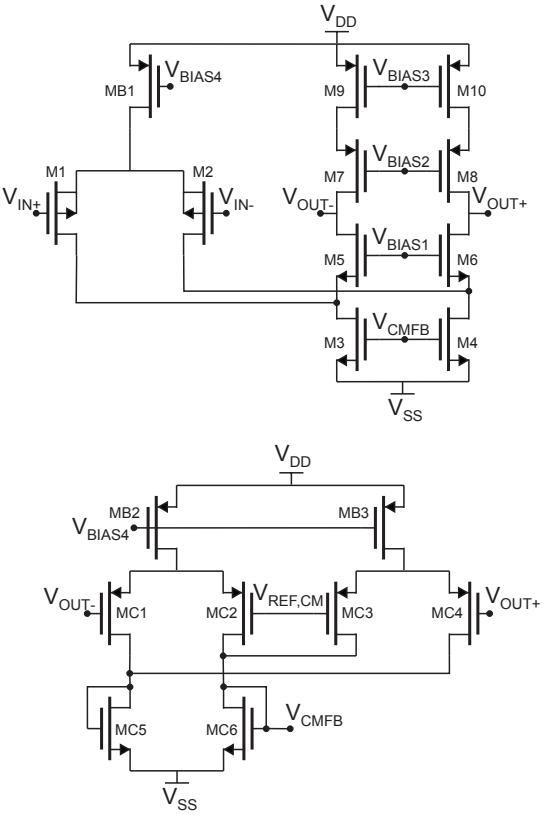


Figure 8.10 Fully-differential folded cascode OTA with its CMFB circuit

The design of each amplifier is, therefore, based on a folded cascode OTA design recipe with additional requirements, such as:

- differential output swing and rejection of the differential signal in the common-mode signal at the output for the first integrator,
- input common-mode range, differential output swing, input common-mode rejection ratio and linearity for the unity-gain buffer,
- differential output swing and high current sourcing/sinking capability for the second integrator.

Each set of specifications obviously needs different design trade-offs. Thus, in the subsections that follow we will first analyze the technology limits, and then determine the necessary topology variants.

8.3.1 Analysis of technology limits

For the given 0.18 μm CMOS technology, EKV model parameters are used for hand-calculations, while the final circuit design is confirmed by simulations with the BSIM model. The EKV model parameters needed for this analysis are listed in Table 8.6. The parameters are extracted, using the BSIM2EKV converter, described in chapter 3, from the BSIM model provided with the technology design kit. The minimum drawn transistor width is $W_{MIN} = 0.5\ \mu m$, and the minimum drawn transistor length is $L_{MIN} = 0.4\ \mu m$. These values are used for the hand-calculations, the simulations and the layout. Then, during the fabrication process all sizes are shrunk by a factor of two.

Table 8.6 EKV model parameters needed for hand-calculations for the given 0.18 μm CMOS technology

VTO_N	0.37 V
VTO_P	- 0.36 V
KP_N	490 μV/A²
KP_P	82 μV/A²
COX	8 mF/m²
n	1.22
U_{aN}	2 V/μm
U_{aP}	4 V/μm

Table 8.7 Folded cascode OTA gain, non-dominant pole frequency and phase margin approximation needed to explore the technology limits

gain	$A_0 \approx \frac{g_{m1,2}}{g_{ds1,2}} \cdot \frac{g_{m5,6}}{g_{ds5,6}}$
non-dominant pole	$f_{ndp} = \frac{1}{2\pi} \cdot \frac{g_{m5,6}}{\sum(C_P)_{FP}} \approx \frac{1}{2\pi} \cdot \frac{g_{m5,6}}{3C_{GS5,6}}$
phase margin	$PM = 180° - atan\left(\frac{f_{pd}}{f_{GBW}}\right) - atan\left(\frac{f_{npd}}{f_{GBW}}\right)$

To explore the technology limits, the circuit parameters of interest, such as gain, non-dominant pole frequency and phase margin are simplified as in Table 8.7. The goal is to determine the maximum gain when the stability condition is fulfilled. It is considered that the non-dominant pole of the amplifier is the pole closest to the origin of all the non-dominant poles in the feedback loop and that the negative feedback factor is equal to one (which is the worst-case situation).

The stability condition, $PM \geq 70° + 10°$, gives the minimal non-dominant pole frequency with regard to the specified gain bandwidth frequency as

$$f_{ndp} \geq 6 \cdot f_{GBW}. \tag{8.1}$$

The security margin of 10° is added because of important differences between the BSIM and EKV capacitance models [4].

To ensure maximal gain, the transistors M1,2 and M5,6 are placed at the limit of weak inversion, that is $IF_{1,2} = IF_{5,6} = IF = 0.1$. Given the fact that both sets of transistors contribute to the sum of the parasitic capacitances in the folded point and weak inversion is the most unfavorable region concerning the size of the intrinsic capacitances, we will assume that $L_{1,2} = L_{5,6} = L$.

The gate-source capacitance of the transistors M5,6, used for the non-dominant pole frequency estimation, is written as a function of the inversion factor as

$$C_{GS5,6} = COXW_{5,6}L \cdot c_{GS5,6} = COX\frac{I_{Dsat5,6}}{2 \cdot n \cdot KP_n \cdot IF \cdot U_T^2}L^2 \cdot c_{GS5,6} \tag{8.2}$$

with the intrinsic capacitance $c_{GS5,6}$ approximated by

$$c_{GS5,6} = \frac{2}{3} \cdot \frac{(x_f + 1) \cdot (x_f - 1/2)}{(x_f + 1/2)^2}. \tag{8.3}$$

This expression is derived from the definition in (2.50), when the transistor M5 operates in saturation, as shown in chapter 2.

Substituting the gate-source capacitance expressed in (8.2) and (8.3) in the stability condition (8.1), the maximal transistor length is easily calculated and is given by

$$L \le \sqrt{\left. \frac{IF \cdot KP_n \cdot U_T}{2 \cdot \pi \cdot COX \cdot 6 \cdot f_{GBW} \cdot \frac{(x_f + 1) \cdot (x_f - 1/2)}{(x_f + 1/2)^2}} \right|_{IF = 0.1}} \approx 0.7\,\mu\text{m}. \tag{8.4}$$

The gain of the folded cascode OTA is approximated here as a two-stage intrinsic gain. Since

$\left(\frac{g_m}{I_{Dsat}}\right) \approx \frac{1}{n \cdot U_T}$ in weak inversion, and $g_{DS} \approx \frac{I_D}{L \cdot U_a}$, it follows that

$$\begin{aligned} A &\approx \left(\frac{g_{m1,2}}{I_{Dsat1,2}}\right) \cdot \frac{I_{Dsat1,2}}{g_{DS1,2}} \cdot \left(\frac{g_{m5,6}}{I_{Dsat5,6}}\right) \cdot \frac{I_{Dsat5,6}}{g_{DS5,6}} \\ &= \left(\frac{1}{n \cdot U_T}\right)^2 \cdot U_{aN} \cdot U_{aP} \cdot L^2. \end{aligned} \tag{8.5}$$

Hence, the maximal gain is 72 dB. Nevertheless, when the circuit is simulated with the BSIM model, this value is reduced by at least 10 dB, as shown in subsection 3.5.2 (the given conversion example is for the BSIM model library file used in this design example). There are two main reasons for this degradation: a lower g_{ds} value and a lower g_m/I_{Dsat} value, as a result of different approaches to model the transistor behavior in weak inversion. This is discussed in section 3.5, and thus must be taken into account when the EKV model is used for hand-calculations. Consequently, the stability and the gain, required in this practical example, cannot be satisfied at the same time.

To achieve the specified gain, the circuit needs an additional gain stage. It is usually implemented as a common-source stage following the folded cascode OTA. The circuit becomes a two-stage amplifier, and thus the compensation capacitance is mandatory for stability. The other possibility is a gain-boosting technique, described in [5]. This makes it possible to keep a single-stage-amplifier frequency behavior and achieve the three-stage-

amplifier gain. Both solutions are discussed in section 7.2 and section 7.3 as topology variants.

8.3.2 Topology variants to fulfill the additional design requirements

CMFB loop stability - If the main amplifier has a high gain, then the CMFB loop also has a high gain as it uses one part of the amplifier gain path. This improves the rejection ratios (*CMRR*, *PSRR*) of the fully-differential amplifier, but on the other hand degrades the stability of the CMFB loop.

An elegant solution to this problem is a modification of the gain path in the main amplifier by splitting the current bias source driven by the CMFB control voltage into two different sources: the first one with a large current and a fixed control voltage and the second one with a small current and controlled by the CMFB amplifier. The gain of the main amplifier does not change, while the phase margin of the CMFB loop transfer function increases. The described topology is presented in section 7.7.

Large differential output swing - The folded cascode OTA can have a large swing at the output, if all bias voltages are correctly determined. The limitation comes from the differential pair at the input of the CMFB amplifier. However, the differential input range of the CMFB amplifier is improved when this differential pair is linearized. This topology variant is discussed in section 7.5.

Rejection of the differential signal in the common-mode signal at the output - The main contributor to the presence of the differential signal in the output common-mode signal is again the CMFB amplifier. In fact, each differential pair at the input of the CMFB amplifier senses a difference between the common-mode reference voltage $V_{REF,\,CM}$ and the output signal from one side of the main amplifier. Consequently, two differential pairs have different input common-mode levels and the *CMRR* of the CMFB amplifier must be improved. This can be achieved by cascoding its current bias sources, as shown in section 7.6.

Large input common-mode range - The basic analog structure that mainly determines the input common-mode range is the differential pair at the folded cascode OTA input. Section 7.4 shows how to overcome this problem. The standard differential pair is replaced by a complementary differential pair, including one NMOS differential pair and one PMOS differential pair connected in parallel.

Table 8.8 summarizes the topology variants for the fully-differential folded cascode OTA and its CMFB amplifier used in this practical design example.

Table 8.8 Overview of topologies required to fulfill the additional design requirements

design requirement	topology variant
high gain	second gain stage
high gain	gain boosting
input common-mode range	complementary differental pair
differential output swing (differential input range)	linearized differential pair at the CMFB input
rejection of the differential signal in the common-mode output signal	cascoded current bias sources in the CMFB amplifier
CMFB loop stability	split sources in the amplifier gain path

8.4 Design of the fully-differential high-gain amplifier

The design of the amplifier in the second integrator is discussed in this section. Although it is designed at the end of the system-level design sequence, its design procedure is presented here before the other two amplifiers. This is important because this design flow is the simplest one, primarily focused on gain enhancement and stability, and represents the basis for the other two procedural design flows.

According to the specification list from Table 8.5 (see subsection 8.2.3) the amplifier in the second integrator has to provide a high output current, i.e a fast time-domain response, and a high gain. All feedback loops must be stable as well. Therefore, it is realized as a single-stage amplifier as previously decided, using a folded cascode OTA (see Figure 8.10) and a gain-boosting stage added for gain enhancement (section 7.3).

On the other hand, to ensure the differential output swing, the input pairs of the CMFB amplifier are linearized and their bias mirrors are cascoded (section 7.5, section 7.6). Due to large circuit currents, the bias current ratio I_3/I_{CMFB} is high, and thus the main amplifier is implemented with split sources to ensure stability of the CMFB loop (section 7.7). The folded cascode OTA and its CMFB amplifier with all necessary topology modifications are shown in Figures 8.11 and 8.12.

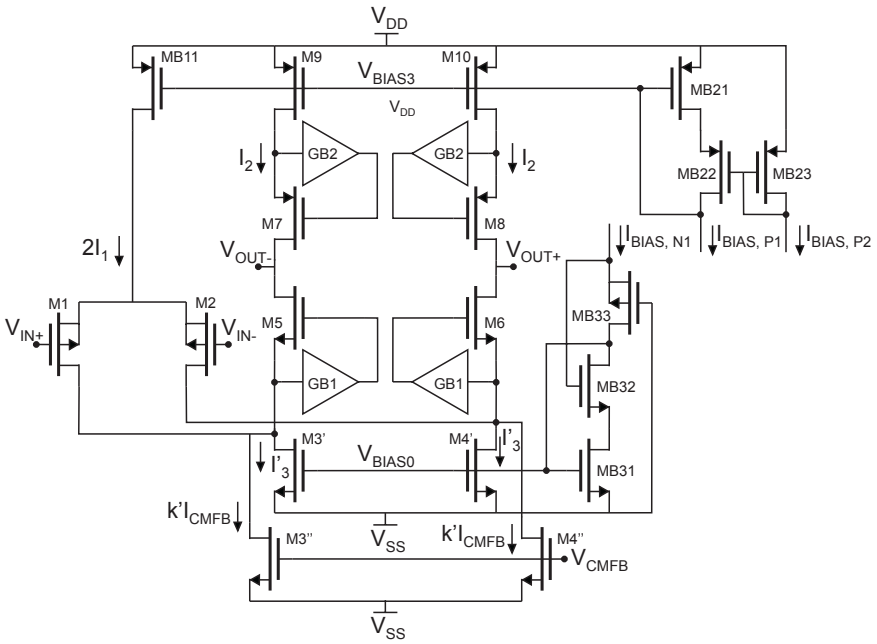


Figure 8.11 Fully-differential folded cascode OTA with gain-boosting and split current bias sources

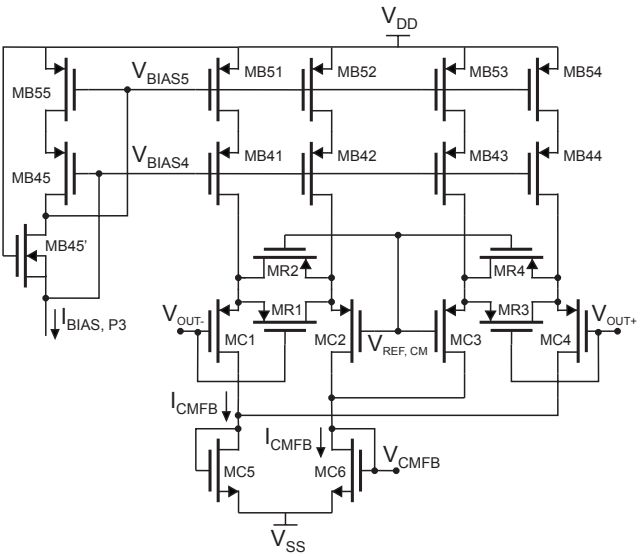


Figure 8.12 Linearized CMFB amplifier with cascoded current bias sources

8.4.1 Equivalent output load

The amplifier feedback during the integration is approximated in Figure 8.9. Although the capacitance of each DAC element can be charged positively or negatively and connected to a positive or negative input terminal of the amplifier (according to [1]), it can be considered that half of the capacitances are always connected on each side. The closed loop gain changes with each integration cycle. To simplify the design procedure, it is assumed that in the worst case situation the amplifier has to provide an output current sufficient to charge/discharge all load, feedback and parasitic capacitances. Thus, the equivalent load is determined for the open-loop situation, depicted in Figure 8.13. The feedback and load (from the quantizer stage) capacitances give

$$C_L = 10 \cdot C_{IN,COMP} + \frac{C_F \cdot (C_S + 38 \cdot C_D)}{C_F + C_S + 38 \cdot C_D} \approx 0.26 \text{ pF}. \tag{8.6}$$

The capacitances $C_{p,OUT}$ and $C_{p,IN}$ are the parasitic capacitances seen at the output and the input of the amplifier, and are unknown at this design point. However, it is sufficient to guess that the total load capacitance is two times larger, that is $C_{L,tot} = 2 \cdot C_L \approx 0.6$ pF.

Since the capacitive feedback changes all the time, it is difficult to specify what is the required closed-loop behavior. Nevertheless, using the values extracted from the behavioral simulations in the system environment for the equivalent transconductance and the output resistance, the following parameters can be derived for the open loop case:

- gain $A_0 = g_{m1} R_{OUT} > 76$ dB,
- gain bandwidth frequency $f_{GBW} = \frac{1}{2} \cdot \frac{g_{m1}}{C_{L,tot}} \approx 600$ MHz.

The closed loop stability has to be confirmed within the system time-domain simulation, whereas the amplifier has to be designed to ensure $PM > 70°$.

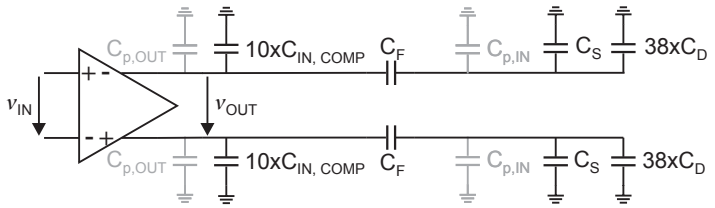


Figure 8.13 Open main feedback loop

8.4.2 Procedural design sequence

Circuit currents - The maximal large signal current provided at each output of the folded cascode OTA is proportional to the differential pair tail current, that is $I_{OUT+,-} = I_1$. From the specification for current sourcing/sinking capability at the output, when the security margins are added, the circuit currents are chosen as

$$\begin{aligned} I_1 &= 190\ \mu A, \\ I_2 &= 1.6 \cdot I_1 = 310\ \mu A, \\ I_3 &= I_1 + I_2 = I_3' + k' I_{CMFB} = 500\ \mu A. \end{aligned} \tag{8.7}$$

The bias transistors M3 = M3'+M3'' and M4 = M4'+M4'' are first designed as there is no source splitting. The current ratio k' is determined in the last design step with regard to the specification for the response-delay.

The current consumption of the main amplifier is 1 mA without biasing circuitry. Consequently, it is mandatory to minimize all other circuit currents: the current I_{CMFB} is set to 10 µA, the current budget for the four gain-boosting amplifiers is set to 10 µA, and the budget for all voltage and current bias circuits to 100 µ A.

Differential pair - The inversion factor of transistors M1,2 is determined from the specified equivalent transconductance and previously chosen circuit current

$$\left(\frac{g_m}{I_{Dsat}}\right)_{1,2} = \frac{2mS}{190\mu A} = 10.5 \rightarrow IF_{1,2} \approx 6.3\,. \tag{8.8}$$

Folded cascode pair - Transistors M5,6 are initially set in moderate inversion to achieve the maximal possible gain with

$$\left(\frac{g_m}{I_{Dsat}}\right)_{5,6} \approx 19.8. \tag{8.9}$$

To minimize the sum of the parasitic capacitances in the folded point, the transistor lengths of contributing transistors are set to $L_{1,2} = L_{5,6} = 1.5L_{MIN}$ and $L_{3,4} = 2L_{MIN}$.

The g_m/I_{Dsat} ratio of transistors M5,6 is then adjusted to respect the stability condition

$$f_{ndp} = \frac{g_{m5,6}}{2\pi(\Sigma C)_{FP}} > \frac{f_{GBW}}{tan(20°)} \approx 1.65\ \text{GHz}. \tag{8.10}$$

This finally gives $\left(\frac{g_m}{I_{Dsat}}\right)_{5,6} \approx 12 \rightarrow IF_{5,6} \approx 4.4$.

Current sources - Transistors M3,4 providing currents I_3 and transistors M7,8 and M9,10 providing currents I_2 are sized to ensure the required differential output range, plus ΔV = 50 mV of security margin on each side. Therefore, the available voltage headroom on each side is

$$V = \frac{(1.8 - (\Delta V + 0.5 + \Delta V))}{2} = 0.6 \text{ V}. \tag{8.11}$$

Since the minimal and maximal output voltages are

$$\begin{aligned} V_{MIN} &= V_{SS} + V = V_{SS} + V_{DSsat3,4} + \Delta V + V_{DSsat5,6} + \Delta V, \\ V_{MAX} &= V_{DD} - V = V_{DD} - |V_{DSsat9,10}| - \Delta V - |V_{DSsat7,8}| - \Delta V, \end{aligned} \tag{8.12}$$

it follows that

$$V_{DSsat3,4} + V_{DSsat5,6} = |V_{DSsat7,8}| + |V_{DSsat9,10}| = 0.5 \text{ V}. \tag{8.13}$$

The saturation voltage of transistors M5,6 is calculated from the inversion factor as

$$V_{DSsat5,6} = U_T \cdot (2\sqrt{IF_{5,6}} + 4) \approx 0.2 \text{ V}. \tag{8.14}$$

Hence, the saturation voltage of transistors M3,4 is $V_{DSsat3,4}$ = 0.3 V, which gives an inversion factor $IF_{3,4}$ = 14.

The transistors M7,8 and M9,10 are designed in the same way. The saturation voltages are chosen to be $V_{DSsat7,8}$ = 0.2 V and $V_{DSsat9,10}$ = 0.3 V, resulting in $IF_{7,8}$ = 4.4 and $IF_{9,10}$ = 14. The transistor lengths of M7,8 and M9,10 are set to ensure that $R_{UP} > R_{DOWN}$, and therefore $L_{7,8} = L_{9,10} = 2.5L_{MIN}$.

Gain-boosting stages - The output resistance of the folded cascode OTA with the transistors sized to ensure stability and output swing is $R_{OUT} \approx$ 64 KΩ, which is inferior to the required 3.2 MΩ. Thus, the required gain of the gain-boosting stages GB1 and GB2 is easily calculated and is equal to 34 dB. Since the gain-boosting amplifiers impose the DC voltages in the sources of transistors M3,4 and M7,8, the acceptable voltage headroom that preserves the specified differential output range is estimated to be one treshold voltage - approximately 0.4V. This imposes the inversion factor of the input transistors in each common-source stage to be IF = 0.5. Finally, as the current consumption is an important design parameter, the gain-boosting stages are designed with a minimal current budget of 10 μA for all four amplifiers

(without biasing circuits). For stability reasons, in order to minimize the contribution to the sum of the parasitic capacitances in the folded point, the maximal lengths of all transistors are set to $2L_{MIN}$.

When the GB1 and GB2 are implemented as common-source stages (as depicted in Figure 7.11), the resulting output resistances are $R_{DOWN} = 2.8$ MΩ, $R_{UP} = 4.4$ MΩ and $R_{OUT} \approx 1.7$ MΩ. Obviously, the auxiliary amplifier GB1, when designed with the conditions derived above, does not provide sufficient gain. As a result, it is re-designed and implemented

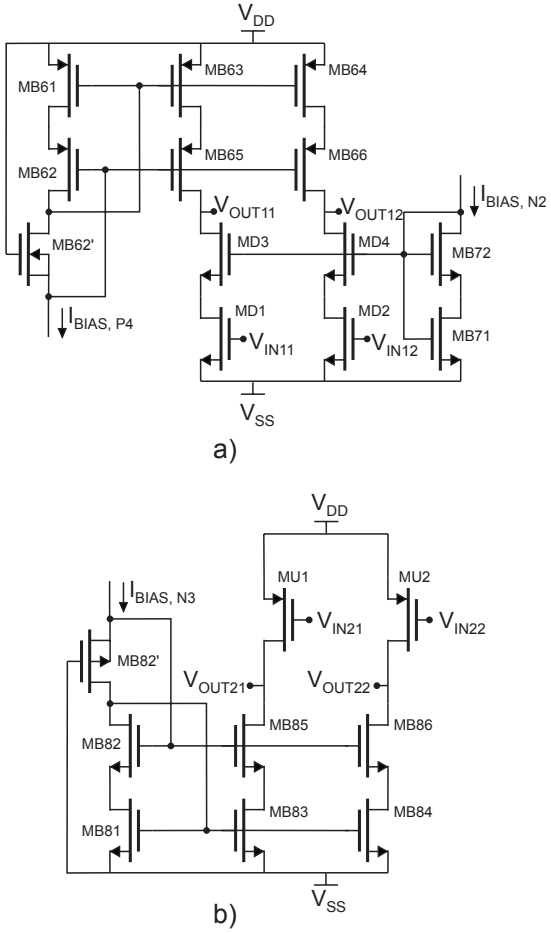


Figure 8.14 Gain-boosting amplifiers: a) cascode amplifier GB1, b) common-source amplifier GB2

as a cascode stage (Figure 8.14). The output resistances are now R_{DOWN} = 67 MΩ, R_{UP} = 4.4 MΩ and $R_{OUT} \approx$ 4 MΩ.

CMFB amplifier - The cascoded current bias sources and the input differential pairs in the CMFB amplifier are designed to be the same for the three amplifiers in the first and second stages of the Delta-Sigma modulator using the following strategy: to ensure the differential input range (i.e. the main amplifier output swing) and to provide the maximal gain. As a result, the transistors of the differential pair are placed at the limit of weak inversion with g_m/I_{Dsat} = 17.8, ensuring in this way the maximum equivalent transconductance. The sizes of the linearization transistors are set to be three times smaller than those of the corresponding differential pair transistors, because an optimal scaling factor that ensures a ±0.5 V swing is found to be m = 3. The diode-connected transistors MC5,6 are simply matched with previously sized current sources M3,4, since they represent a simple current mirror.

Split sources - The transistors M3, M4 are split in two in order to achieve stability of the CMFB loop. The minimal current controlled by the CMFB amplifier is determined from the system-level simulations since the response delay of the amplifier is the main design requirement, and can be confirmed only within the system simulations. The current ratio *k'* (see Figure 8.11) is changed in the range 1 to 10, where the maximal value gives the lowest acceptable phase margin of approximately 50°. The time-domain simulation within the system is repeated for each *k'*. Finally, k' = 8 is chosen as the best

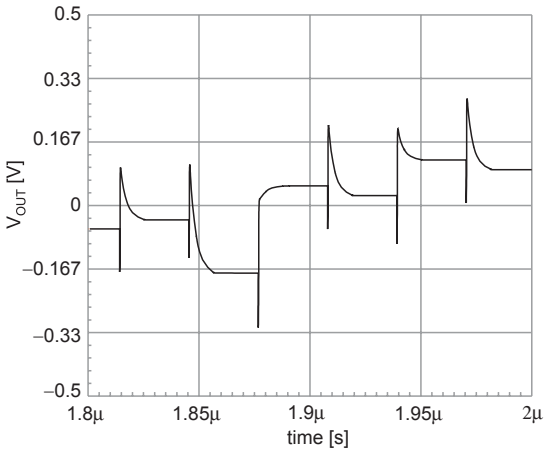


Figure 8.15 Differential output signal of the second integrator in the time-domain system simulation

compromise, resulting in *SNDR* = 93.1 dB, *PM* = 60° and an acceptable time-domain response (depicted in Figure 8.15), even though the phase margin is 10° below the specification. Thus, the currents are set in the following way: $I_3' + k' I_{CMFB}$ = 420 μA + 80 μA.

Transistor sizes - Complete schematics with final dimensions are presented in Figures 8.16, 8.17, 8.18 and 8.19. The transistor sizes are real sizes (shrunk by a factor of 2). The circuit currents are also indicated.

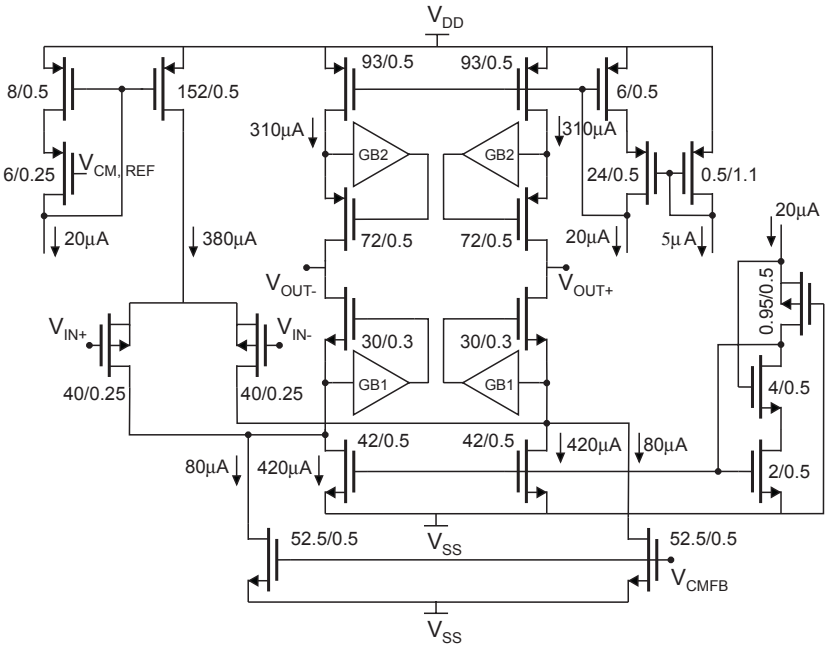


Figure 8.16 Complete schematic with final dimensions: main amplifier

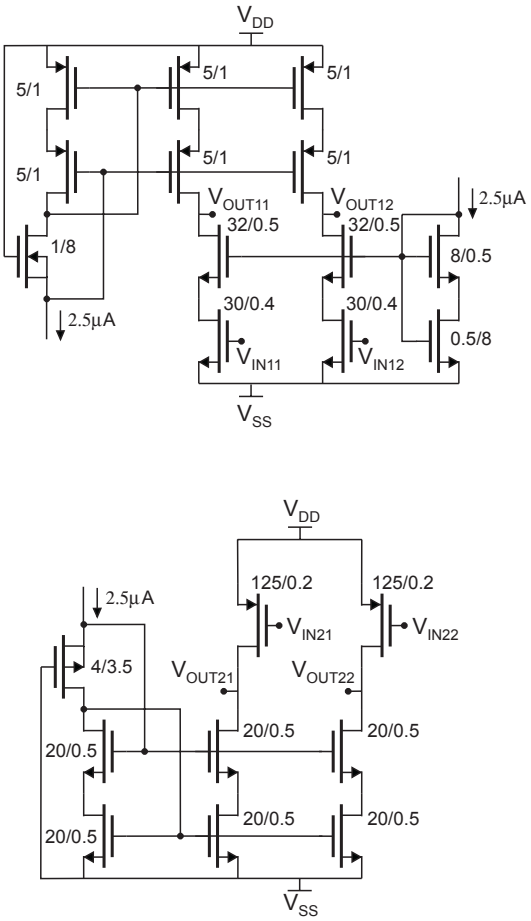


Figure 8.17 Complete schematic with final dimensions: gain-boosting amplifiers

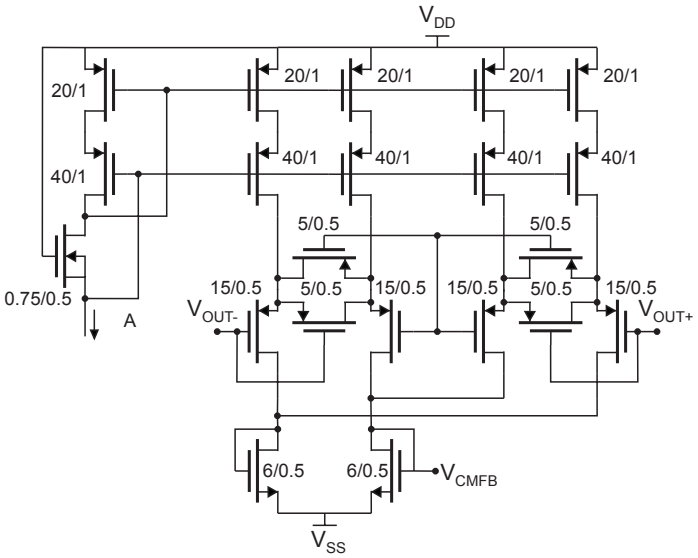


Figure 8.18 Complete schematic with final dimensions: CMFB amplifier

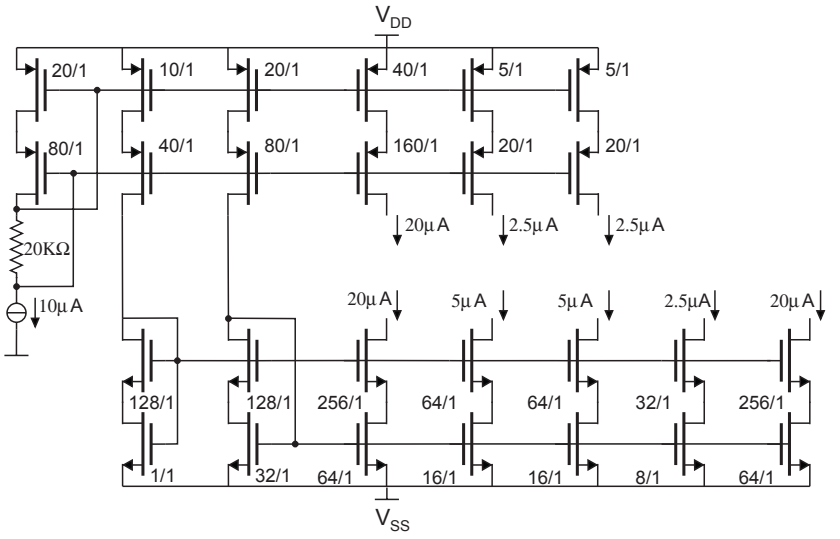


Figure 8.19 Complete schematic with final dimensions: biasing circuit

8.4.3 Simulation results

AC analysis - The transfer function of the CMFB loop is simulated using the testbench in Figure 8.20. The loop is opened at the CMFB amplifier input. The DC feedback is ensured by an RC network where $(2\pi RC)^{-1}$ = 1 μHz. The open CMFB loop transfer function is given by

$$H(j2\pi f) = \frac{v_{OUT+} + v_{OUT-}}{v_{IN,\,CM}}, \tag{8.15}$$

and its module and phase are shown in Figure 8.21 (the amplifier symbol represents the main amplifier connected with the CMFB amplifier).

The *PM* is lower than the specified 70°, but this is the maximal value that ensures an acceptable time-domain response, as discussed previously. On the other hand, the DC gain is only 23 dB and results in low rejection ratios. However, no degradation in *SNDR* values is noticed when the interactions with other circuit blocks are tested. The transfer function of the main amplifier is shown in Figure 8.22.

DC analysis - Since the capacitive feedback represents an open circuit in the DC domain, the simulations are performed with a resistive feedback. The feedback resistances are chosen in the following way: $R_1, R_2 \gg R_{OUT}$, as well as $R_2/R_1 = 10$ to provide a closed loop gain. It is estimated that the highest gain in the case of capacitive feedback, when all feedback capacitances are connected as in Figure 8.9, is equal to 10.

The differential output swing is determined for a gain error of 1%, and is shown in Figure 8.23 in two different cases: with and without linearization of the differential pair at the input of the CMFB amplifier. Obviously, the CMFB circuit without linearization severely degrades the swing of the differential output signal of the main amplifier.

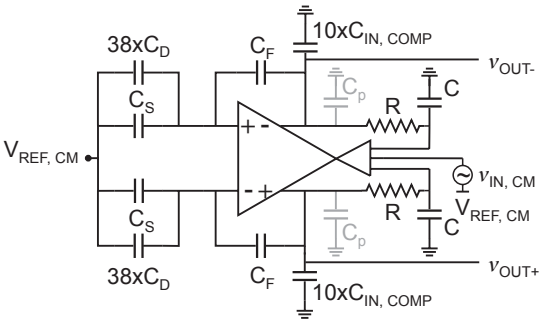


Figure 8.20 Open CMFB loop

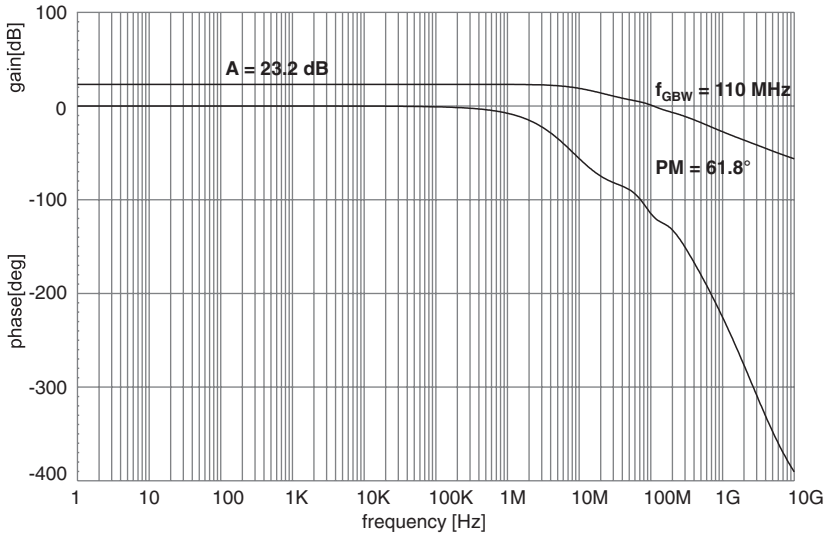


Figure 8.21 Open CMFB loop transfer function module and phase

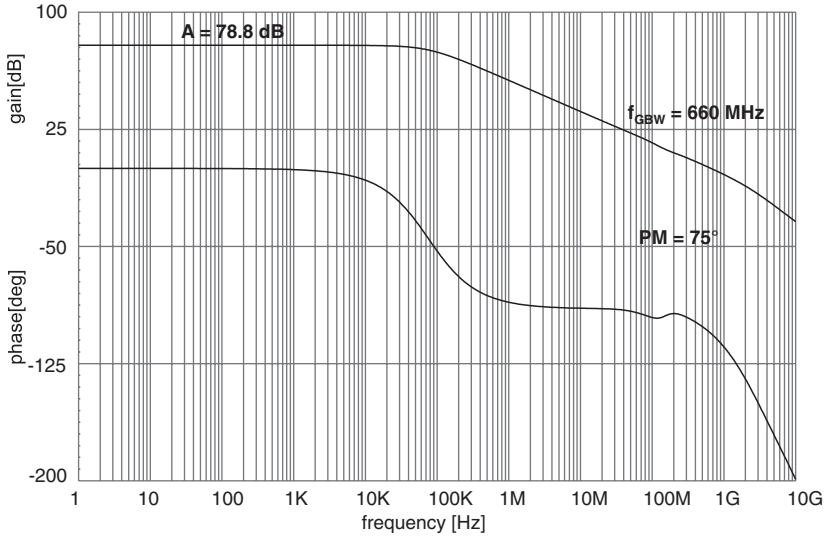


Figure 8.22 Main amplifier transfer function module and phase

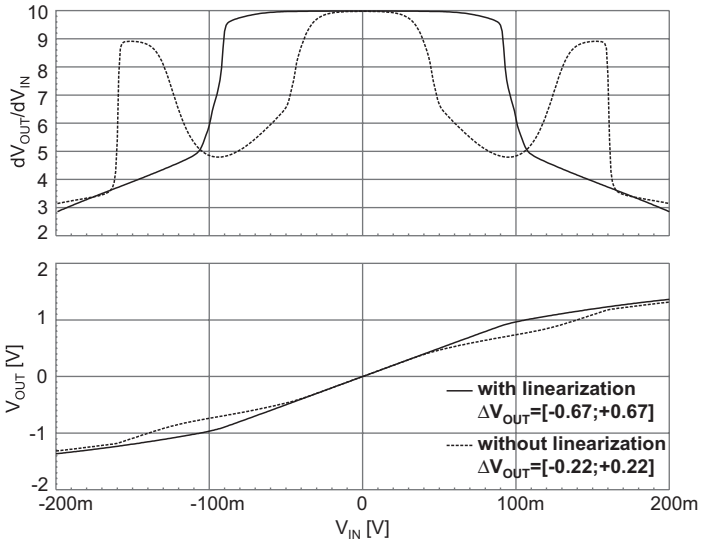


Figure 8.23 Differential output swing with and without linearization of the differential pair at the input of CMFB amplifier

In addition, the bias current mirror of the differential pair (as depicted in Figure 8.11) at the main amplifier input works at the limit of the linear region, because the available voltage headroom is only 0.2 V due to large circuit currents. The worst, typical, and best case simulations revealed the performance degradation. When the source voltage of the differential pair changes in the worst, typical, and best cases by approximately ±50 mV, the differential pair bias current changes by ±20 μA. Therefore, the bias mirror is re-designed as in Figure 8.24. Now, the two transistors MB11,12 do operate in the linear region, but the current is mirrored correctly, thanks to transistor MB13 that is sized as

$$\left(\frac{W}{L}\right)_{B13} = \frac{I_{BIAS,DP}}{2I_1}\left(\frac{2W}{L}\right)_{1,2} \tag{8.16}$$

that is with regard to the differential pair transistors. This made it possible to create approximately the same drain voltage for transistor MB12 as is the drain voltage of transistor MB11 imposed by the differential pair source voltage in the worst, typical or best case. The bias current mirror operating in the linear region does not degrade the amplifier performance, since the voltage headroom imposed by the source of the differential pair does not change significantly. In fact, the differential signal amplitude at its input is small, and thus a large differential input range is not required. At the same time, the negative feedback

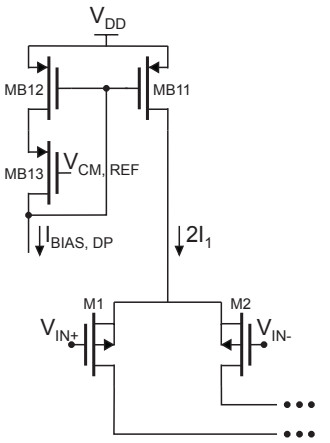


Figure 8.24 Modified bias circuit of the input differential pair (current mirror operating in the linear region)

Table 8.9 Time-domain system simulation results for the second integrator

interaction chain	*SNDR* [dB]
second integrator alone	93.9
sampling stage + second integrator	93.1
second integrator + quantizer	93.3
sampling stage + second integrator + switched-capacitor DAC	90.8

network keeps its input common-mode level constant and there is no need for a large common-mode input range.

Time-domain system simulations - The interaction with neighboring blocks at the transistor level is tested. The simulation results are summarized in Table 8.9.

8.5 Design of the fully-differential difference amplifier

The amplifier in the unity-gain buffer stage is a fully-differential voltage follower, as depicted in Figure 8.6. It is realized as a difference amplifier with two differential pairs at the input. One input of each differential pair is connected to one output terminal creating in this way a negative feedback that minimizes the difference between input and output signals.

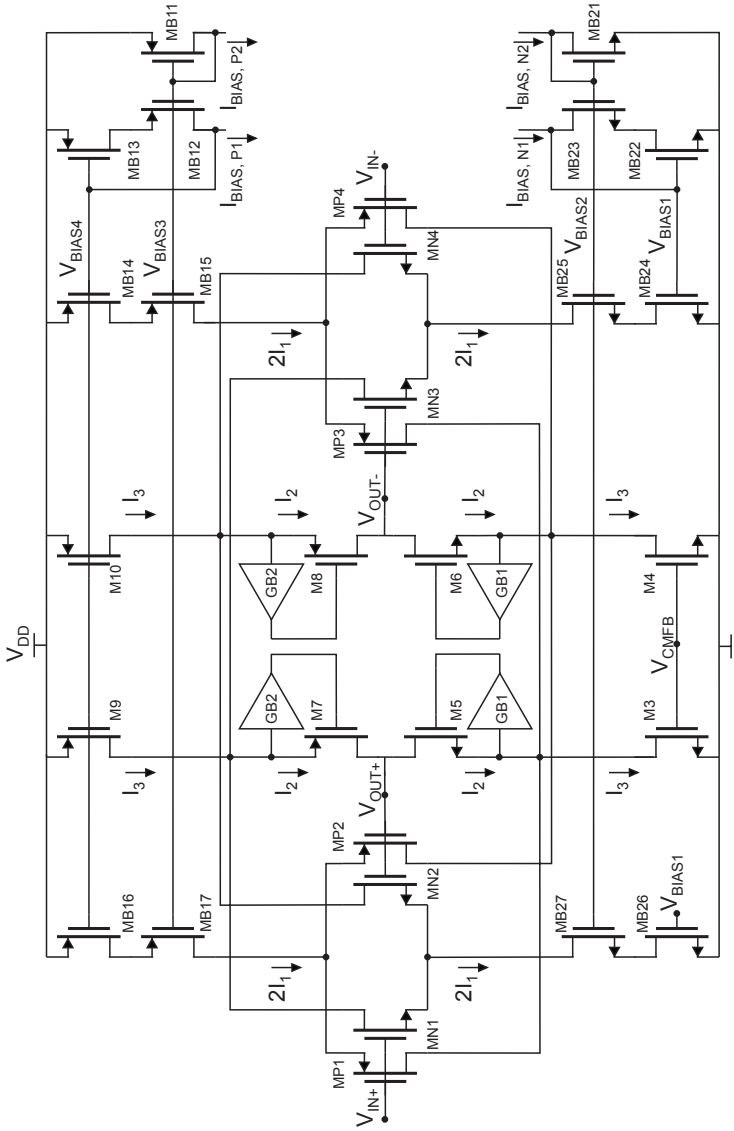


Figure 8.25 Fully-differential difference folded cascode OTA with gain-boosting

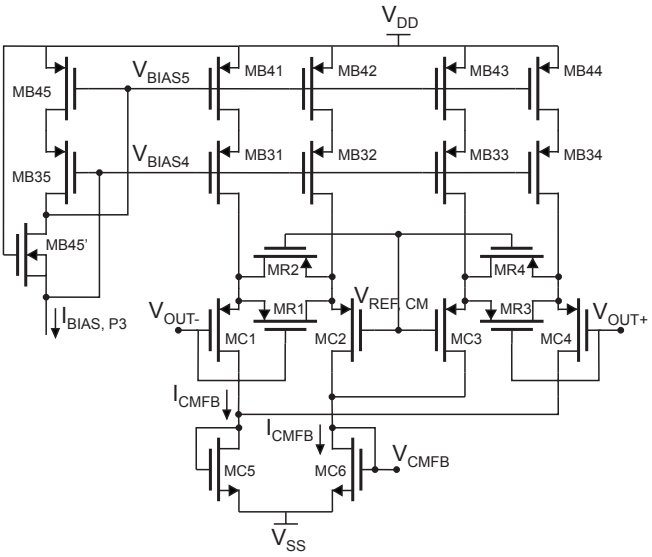


Figure 8.26 Linearized CMFB amplifier with cascoded current bias sources

The main design requirement is linearity. In other words, the output must follow the input signal in a specified ± 0.5 V differential signal range with an error of 0.1%. This implies a gain of at least 60 dB, when calculated from (7.29). However, according to the specifications extracted from time-domain simulations (Table 8.3), the required gain is 70 dB. Hence, the gain-boosting stages (section 7.3) are mandatory to achieve this value. Furthermore, the requirement to "follow" the input voltage in a specified signal range means ± 0.5 V of input common-mode range and ± 0.5 V of differential output swing. Therefore, each differential pair is implemented as a complementary differential pair (section 7.4), and the input pairs of the CMFB amplifier are linearized and their current bias sources are cascoded (section 7.5, section 7.6). Finally, to improve the rejection of common-mode variations in the differential input signal, the current bias sources of input pairs in the main amplifier are also cascoded. The circuit schematic with all topology modifications is presented in Figures 8.25 and 8.26.

8.5.1 Procedural design sequence

Circuit currents - The required output sourcing/sinking capability according to the simulation with the behavioral model is $I_{OUT+,-}$ = 100 µA. The maximal current that can be provided at each output is determined in the most unfavorable case: when the input signal has a large amplitude and is in

opposition to the output signal. In this case, V_{IN+} increases, and V_{IN-} decreases, while V_{OUT+} decreases, and V_{OUT-} increases. Consequently, NMOS transistors MN1,3 and PMOS transistors MP2,4 conduct all available bias current, denoted as $2I_1$ in Figure 8.25. Since the bias sources M3,4, and M9,10 impose the current I_3, then the current that flows through transistor M7 is I_3-$2I_1$-$2I_1$, and through M5 is I_3. Similarly, the current that flows through the transistor M8 is I_3, and through M5 is I_3-$2I_1$-$2I_1$. Hence, the current sourced/sunk at each output is $|I_{OUT+,-}| = 4I_1$. Accordingly, the circuit currents are chosen as follows

$$\begin{aligned} I_1 &= 25\ \mu A, \\ I_2 &= 2.2 \cdot I_1 = 55\ \mu A, \\ I_3 &= 2 \cdot I_1 + I_2 = kI_{CMFB} = 105\ \mu A. \end{aligned} \tag{8.17}$$

The current consumption of the main amplifier without biasing circuitry is I_3+$2I_1$+$2I_1$ = 310 μA. The other circuit currents are set similarly as in the case of the amplifier in the second integrator: I_{CMFB} is fixed at 10 μA, the budget for all gain-boosting stages at 10 μA, and the budget for all voltage and current bias circuitry at 80 μ A.

Estimation of the equivalent output load - The load capacitance is the sampling capacitance C_S, as discussed in subsection 8.2.1. However, the parasitic capacitances at the main amplifier output, as well as the input parasitic capacitances of the CMFB amplifier, which are also connected to the output, must be taken into account. As a result, the load capacitance is (over)estimated as

$$C_L = C_S + C_{p,\,OUT} + C_{p,\,IN,\,CMFB} \approx 1.5 \cdot C_S . \tag{8.18}$$

Circuit symmetry - The design of the fully-differential difference amplifier consists in the design of two complementary folded cascode amplifiers connected in parallel. If one amplifier is designed, then the other one can be derived by simply replacing the NMOS by PMOS transistors and vice versa, and setting the sizes as

$$\frac{(W/L)_N}{(W/L)_P} = \frac{KP_P}{KP_N} . \tag{8.19}$$

A key point here is to provide identical behavior when input terminal voltages decrease and increase. Therefore, the additional sizing conditions must be fulfilled to ensure circuit behavioral symmetry:

- The NMOS and PMOS differential pairs have the same equivalent transconductance, denoted in the text that follows as g_{m1}.

- The folded cascode pair M5,6 of NMOS differential pairs, and the folded cascode pair M7,8 of PMOS differential pairs have the same transconductance.
- The sum of parasitic capacitances in each folded point is the same.
- The equivalent resistances seen at the output: R_{DOWN} and R_{UP} are the same. This is fulfilled when

$$\frac{g_{m5,6}}{g_{DS5,6}} = \frac{g_{m7,8}}{g_{DS7,8}} \text{ and } g_{DS,N} + g_{DS9,10} = g_{DS,P} + g_{DS3,4}, \tag{8.20}$$

  where $g_{DS,N}$ and $g_{DS,P}$ are the output conductances of the NMOS and PMOS differential pairs.
- The transistors M3,4 and M9,10 representing the current sources have the same saturation voltage, i.e. the same inversion factor.

Input differential pairs - The equivalent transconductance of the amplifier from Figure 8.25 is equal to the sum of the equivalent transconductances of the NMOS and PMOS differential pairs, as seen in section 7.4, $g_{m,eq} = g_{m,eq,NMOS} + g_{m,eq,PMOS} = 2g_{m1}$. To ensure the specified gain bandwidth frequency, the minimal transconductance of each transistor is estimated as

$$(g_{m1})_{MIN} = \frac{g_{m,eq}}{2} = \frac{2\pi \cdot C_L \cdot f_{GBW}}{2} \approx 660\ \mu\text{S}, \tag{8.21}$$

and thus $(g_m/I_{Dsat})_1 > 13.2$ and $IF_1 > 3.5$. However, these are just the limit values. Since a gain of 70 dB is required in the whole signal range, the equivalent transconductance must be constant in this range. Therefore, the final value of the inversion factor of the differential pair transistors has to be chosen with respect to the required input common-mode voltage range of ± 0.5 V.

The common-mode voltage range is affected by the voltage headroom requirement of the current bias transistors $V_{DS,BIAS,N}$, $V_{DS,BIAS,P}$ and the gate-source voltage headroom of the differential pairs $V_{GS,N}$, $V_{GS,P}$. It can be expressed as

$$\begin{aligned} V_{SS} + V_{DS,BIAS,N} + V_{GS,N} &\le V_{REF,CM} - |\Delta V_{IN,CM}|, \\ V_{REF,CM} + |\Delta V_{IN,CM}| &\le V_{DD} - |V_{DS,BIAS,P}| - |V_{GS,P}|. \end{aligned} \tag{8.22}$$

According to the design specification, it follows that

$$V_{DS,BIAS,N} + V_{GS,N} = |V_{DS,BIAS,P}| + |V_{GS,P}| = 0.65\ \text{V}. \tag{8.23}$$

The important point here is that the current bias sources are implemented as cascode current mirrors. This ensures the working condition of each

differential pair (the cascode transistor protects the mirroring transistor from the variations of the source voltage of the differential pair), and improves the rejection of common-mode variations of the signal driven to the amplifier input from the previous stage. Taking this into account we finally obtain (without adding the security margins)

$$V_{GS} = V_{GS,N} = |V_{GS,P}| \leq 0.3 \text{ V, and} \tag{8.24}$$

$$V_{DS,BIAS} = V_{DS,BIAS,N} = |V_{DS,BIAS,P}| \leq 0.35 \text{ V}. \tag{8.25}$$

The inversion factor of differential pair transistors is thus

$$IF_1 = ln^2\left[1 + exp\left(\frac{V_{CM,REF} - |\Delta V_{IN,CM}| - V_{T0} - nV_{DS,BIAS}}{2nU_T}\right)\right] \approx 0.4 \tag{8.26}$$

which gives $(g_m/I_{Dsat})_1 \approx 25$. Nevertheless, in the simulations with the BSIM model, the maximum achieved transconductance with the given bias current and the calculated inversion factor is $g_{m1} \approx 500$ μS, and thus $(g_m/I_{Dsat})_1 \approx 20$. Even though the transconductance value is smaller than the minimal estimated value in (8.21), it results in gain bandwidth and phase margin simulated values within the required specifications, since the output capacitance is overestimated in (8.18). On the other hand, the simulated voltage headroom for the bias current mirror is approximately 0.3 V, which imposes the limit of weak inversion as the operating region.

Folded cascode pairs - The transistors M5,6 and M7,8 are sized with respect to the stability condition. The unity-gain loop stability is guaranteed if the amplifier non-dominant pole is at

$$f_{ndp} = \frac{1}{2\pi} \cdot \frac{g_{m5,6,7,8}}{(\Sigma C)_{FP}} \geq \frac{f_{GBW(noC_S)}}{tg(90° - PM)} \approx 440 \text{ MHz}, \tag{8.27}$$

where the gain bandwidth frequency is given here for a more critical situation, that is when the sampling capacitances switch roles and the output load is only the sum of the circuit parasitic capacitances. This gives $(g_m/I_{Dsat})_{5,6,7,8} \approx 17$, and $IF_{5,6,7,8} \approx 1.7$. The lengths of differential pair and folded cascode pair transistors are set to L_{MIN} in order to minimize the sum of parasitic capacitances in the folded point.

Current sources - As in the case of the amplifier in the second integrator, the voltage headroom on each amplifier side is determined by the specified differential output swing, and according to (8.11) it equals to 0.6 V. The minimal and maximal output voltages are then

$$V_{MIN} = V_{SS} + V = V_{SS} + V_{DSsat3,4} + \Delta V + V_{DSsat5,6} + \Delta V, \qquad (8.28)$$
$$V_{MAX} = V_{DD} - V = V_{DD} - |V_{DSsat9,10}| - \Delta V - |V_{DSsat7,8}| - \Delta V.$$

Using the inversion factor value determined in the previous step, we obtain $V_{DSsat5,6,7,8} \approx 0.17$ V, and thus $V_{DSsat3,4} = |V_{DSsat9,10}| \leq 0.3$ V. However, since the transistors M9,10 must be matched with the transistors of the current bias of the PMOS differential pair which have a limited voltage headroom, their saturation voltage is chosen to be $V_{DSsat3,4} = |V_{DSsat9,10}| = 0.2$ V. This finally gives $(g_m/I_{Dsat})_{3,4} = (g_m/I_{Dsat})_{9,10} \approx 14.8$. The lengths of transistors M3,4 and M9,10 are chosen as a compromise between stability and the required output resistance and set to $2.5L_{MIN}$.

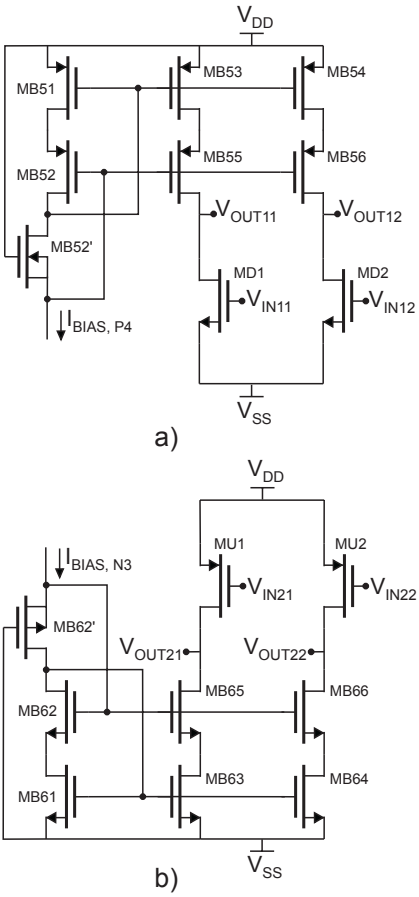


Figure 8.27 Gain-boosting amplifiers: a) common-source amplifier GB1, b) common-source amplifier GB2

Gain-boosting stages - The output resistance of the amplifier without gain enhancement is $R_{OUT} = R_{DOWN} \| R_{UP} \approx 155$ KΩ. The resulting amplifier gain is 43.7 dB, and thus the gain-boosting stages have to provide a gain of at least 30 dB. GB1 and GB2 are implemented as common-source amplifiers and depicted in Figure 8.27. The inversion factor of input transistors in each common-source stage is set to $IF = 0.5$, to keep the DC voltages in the drains of transistors M3,4 and M7,8 in the order of one threshold voltage. As in the previous case, the current budget for all four amplifiers is 10 μA and the maximal transistor lengths are set to $2L_{MIN}$, so as not to degrade the stability. After all transistors have been sized, the resulting amplifier output resistance is $R_{OUT} = 6.3$ MΩ, and the amplifier gain of 76 dB is achieved.

CMFB amplifier - The cascoded current bias sources and the input differential pairs in the CMFB amplifier are designed as in section 8.4. The diode-connected transistors MC5,6 are matched with the previously sized current sources M3,4, since they represent a simple current mirror. No stability problems in the CMFB loop occurred with the obtained CMFB loop gain of 45.8 dB.

Current bias mirrors of input differential pairs - In this step, the current bias mirrors of input differential pairs are designed with respect to the conditions derived in the previous steps, and re-designed with the additional requirement to improve the *CMRR*. As previously decided, the bias mirrors are implemented as cascode current mirrors. The PMOS part consists of transistors MB14,16 and cascode transistors MB15,17 biasing PMOS differential pairs. The NMOS part consists of transistors MB24,26 and cascode transistors MB25,27 biasing NMOS differential pairs. The initial transistor sizes are determined in the following way:

- For circuit symmetry and good matching of the circuit currents, the transistors MB14,16 are matched with the current sources M9,10, while MB24,26 are biased in the same way as M3,4. The transistor lengths are therefore set to $2.5L_{MIN}$.
- Taking into account that the available voltage headroom for bias sources, determined from the input common-mode range specification, is 0.3 V and that the calculated saturation voltage of the mirror transistors is 0.2 V, the cascode transistors are placed at the limit of weak inversion in order to have $V_{DSsat} = 0.1$ V.
- The lengths of cascode transistors are set to L_{MIN}, in order to minimize the sum of parasitic capacitances seen in the sources of the differential pairs, since this degrades the time-domain response.

The input common-mode rejection ratio of a fully-differential amplifier is defined in chapter 5 as

$$CMRR = \frac{A_{CMFB}}{A_C}. \tag{8.29}$$

A_{CMFB} is the gain of the CMFB loop, while A_C is the gain of the input common-mode signal to each amplifier output, and is given by

$$A_C = \frac{v_{OUT+,-}}{v_{IN,\,CM}} = \frac{1}{R_{BIAS}} R_{OUT}, \tag{8.30}$$

where R_{OUT} is the amplifier output resistance, and R_{BIAS} is the output resistance of the bias mirror.

The amplified input common-mode signal contributions that appear in the common-mode output signal ($v_{OUT+} + v_{OUT-}$)/2 are minimized by the negative CMFB loop. The higher the gain of the common-mode loop, the better is the rejection of this unwanted component in the common-mode output signal. Nevertheless, the gain of the CMFB loop in this example is not an independent variable. It is determined as the maximal possible gain that can be achieved without degrading the differential output swing and respecting the stability of the feedback loop. Hence, the only way to improve the *CMRR* is to decrease the gain of the input common-mode signal. The independent variable is R_{BIAS}, since R_{OUT} is determined by the main amplifier gain and gain bandwidth specifications. Therefore, it follows that its required value is

$$R_{BIAS} \geq R_{OUT} \cdot \frac{CMRR}{A_{CMFB}} \approx 3.2\ \text{M}\Omega. \tag{8.31}$$

However, the resulting output resistance of the current bias mirrors, obtained respecting the design conditions above, is approximately 0.5 MΩ. Obviously, this is far from the required value. Nevertheless, it represents the maximum possible value that can be achieved with the given technology and with all cascode current mirror transistors operating in saturation.

One possible solution is to augment R_{BIAS} by pushing the transistors MB14,16 and MB24,26 into the linear region. This is done by placing MB15,17 and MB25,27 in moderate inversion and thus lowering the drain voltage of transistors MB14,16 and MB24,26. It is estimated that the operation in the linear region is not critical, since the drain voltage of mirror transistors is set by the cascode transistors and does not change as long as the cascode transistors operate in the saturation region. Since the current mirroring is degraded in this way, the sizes of transistors MB14,16 and MB24,26 have to be adjusted to provide the required bias current values.

The resistance seen in the source of cascode transistors MB15,17 and MB25,27 now becomes larger than $1/g_{DS}$, and depends on the source voltage of the cascode transistors (i.e. the drain voltage of transistors MB14,16 and MB24,26). It is thus determined by the inversion factor of these transistors, which has to be tuned in several iterations to find the optimal value. Finally, the value $IF = 1.4$ is chosen.

Transistor sizes - Complete schematics with final dimensions are presented in Figures 8.28, 8.29, 8.30 and 8.31. The transistor sizes are real sizes. The circuit currents are also indicated.

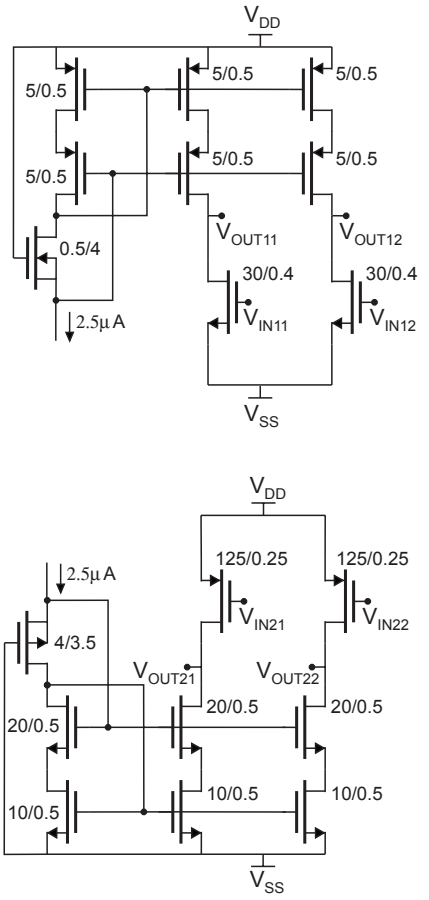


Figure 8.28 Complete schematic with final dimensions: gain-boosting amplifiers

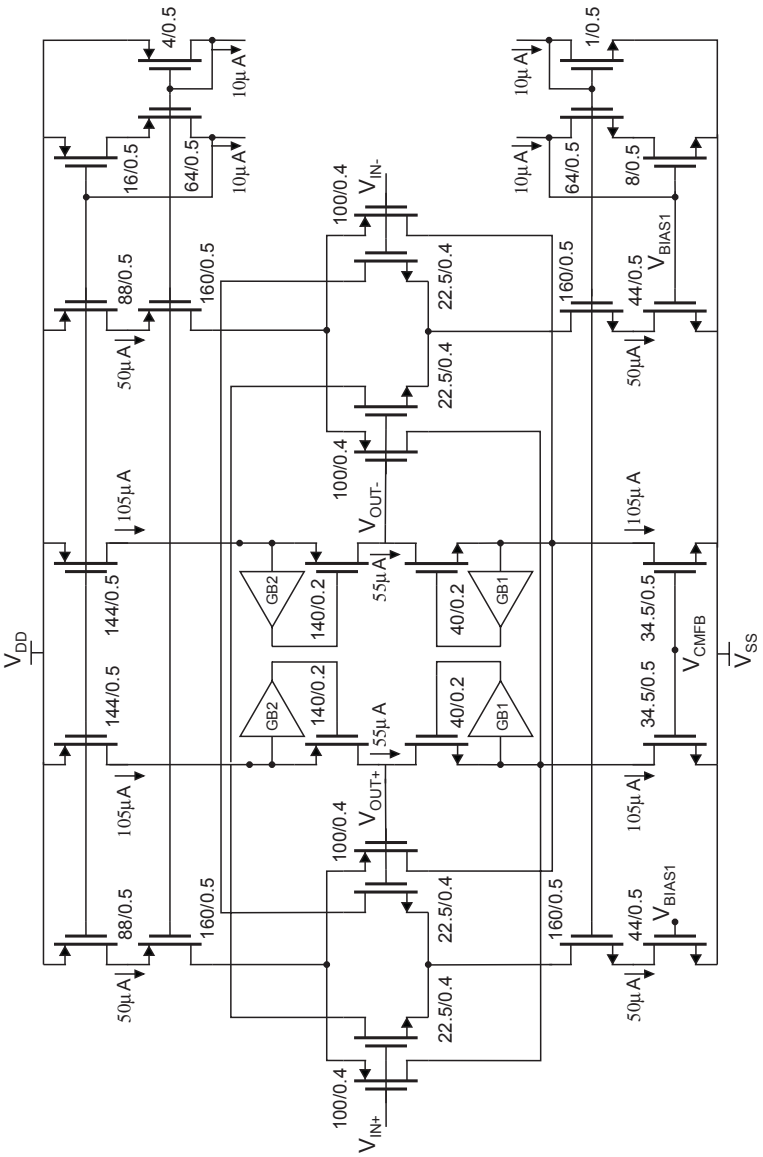


Figure 8.29 Complete schematic with final dimensions: main amplifier

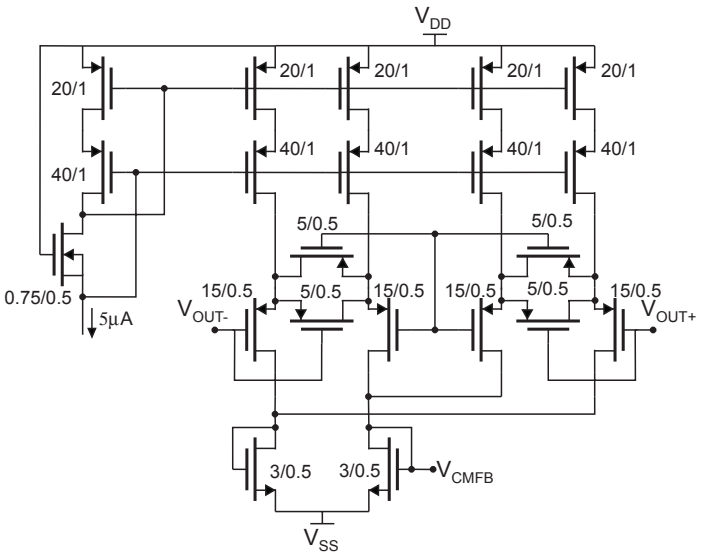


Figure 8.30 Complete schematic with final dimensions: CMFB amplifier

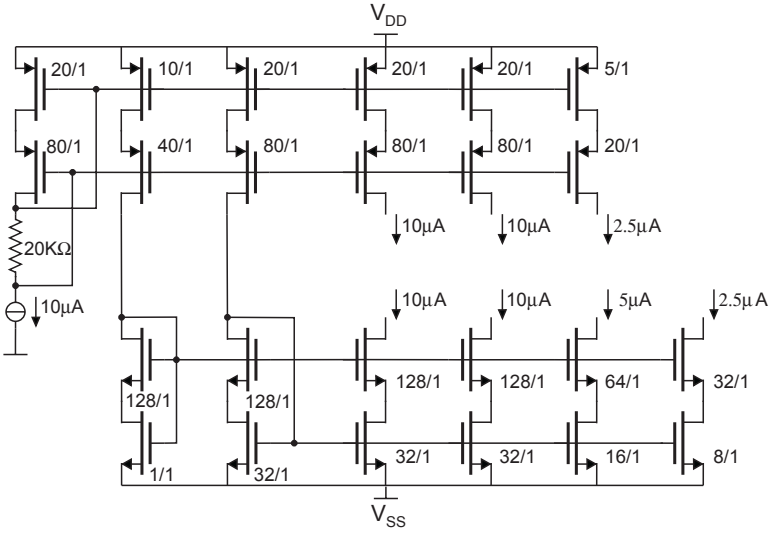


Figure 8.31 Complete schematic with final dimensions: biasing circuit

8.5.2 Simulation results

AC simulation - The open loop transfer function is determined using the testbench from Figure 8.32. The DC feedback is ensured by an RC network where $(2\pi RC)^{-1} = 1\ \mu$Hz. The transfer function module and phase are shown in Figure 8.33 in two cases: with and without sampling capacitance at the output. The phase margin $PM = 52°$ when the sampling capacitance is not connected is acceptable, because this situation occurs within a very short lapse of time. Furthermore, no degradation is observed in the time-domain system simulations.

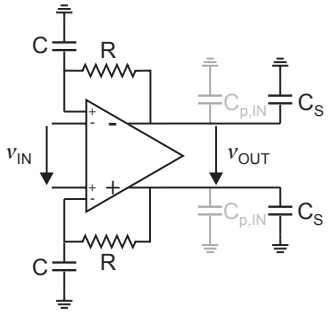


Figure 8.32 Open main feedback loop

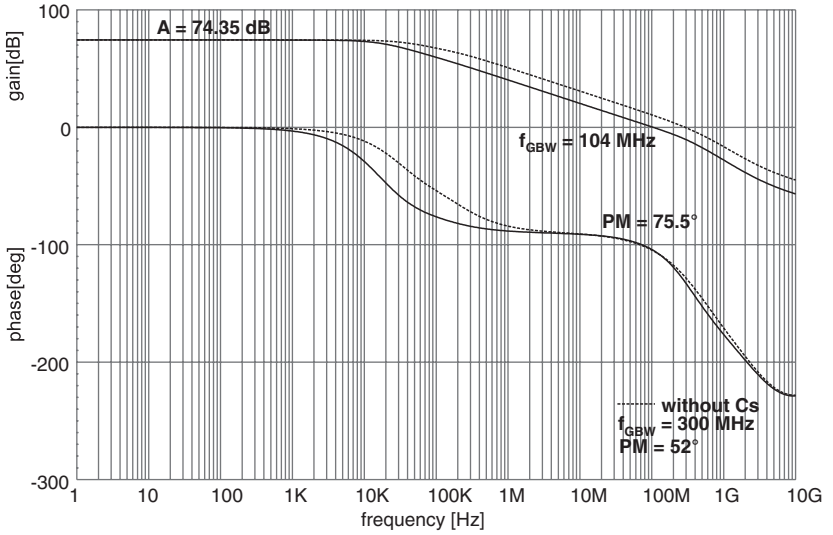


Figure 8.33 Open main feedback loop transfer function module and phase (with and without load capacitance)

The transfer function of the CMFB loop is simulated using the testbench in Figure 8.34 (the amplifier symbol represents the main amplifier connected with the CMFB amplifier). The loop is opened at the CMFB amplifier input. The DC feedback is ensured by the RC network where $(2\pi RC)^{-1} = 1\ \mu Hz$. The open CMFB loop transfer function is given by

$$H(j2\pi f) = \frac{v_{OUT+} + v_{OUT-}}{v_{IN,\,CM}}, \tag{8.32}$$

and its module and phase are shown in Figure 8.35 in two cases: with and without sampling capacitance at the output.

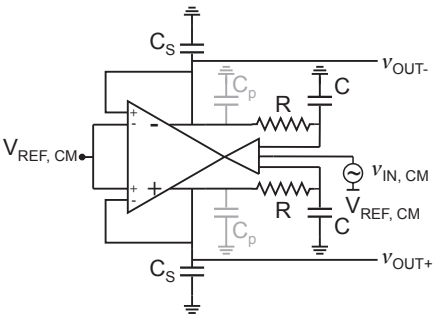


Figure 8.34 Open CMFB loop

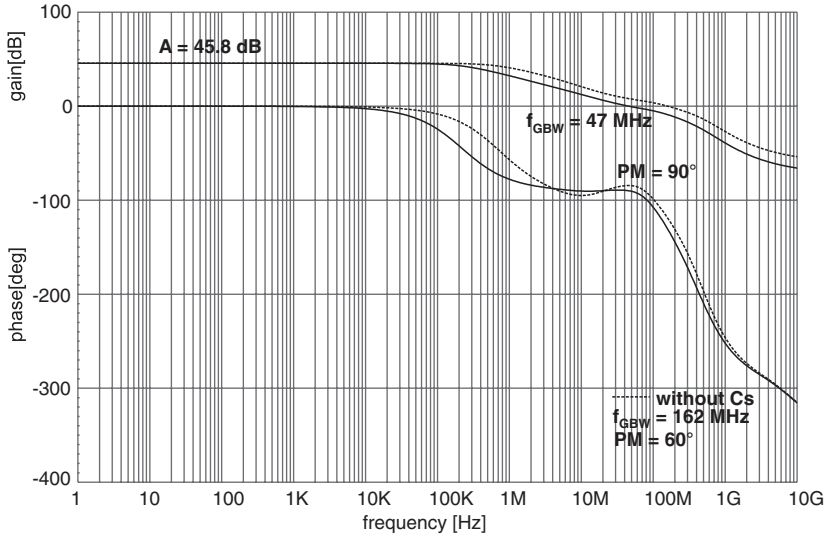


Figure 8.35 Open CMFB loop transfer function module and phase (with and without load capacitance)

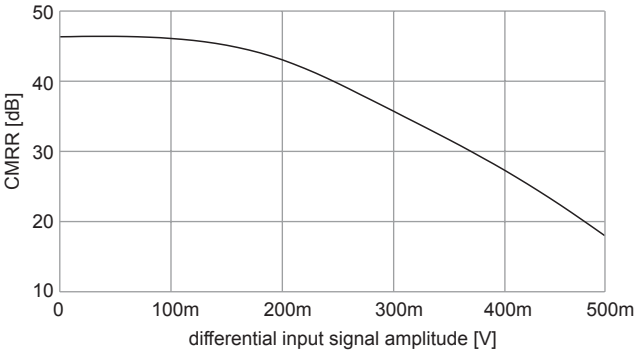


Figure 8.36 CMRR in the differential input range

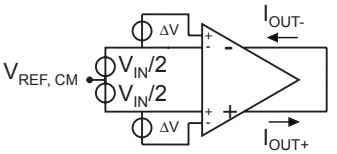


Figure 8.37 Circuit used to determine the equivalent transconductance

Figure 8.36 displays the *CMRR* in the differential input signal range. For large signal amplitudes the rejection ratio is of the order of 20 dB, and obviously the *CMRR* is not improved in this case. The problem is solved by imposing an acceptable common-mode variation of ± 5 mV at the output of the first integrator. When the amplifier of the first integrator is re-designed respecting this condition, and the interaction of the two stages is tested, the obtained *SNDR* is within the specified range and is equal to 94.1dB.

DC analysis - The equivalent transconductance is simulated for the input signal range using the circuit depicted in Figure 8.37. The small-signal difference, representing the difference between the input and output signals in unity-gain configuration, is set to $2\Delta V = 1\ \mu V$. The equivalent transconductance of the difference amplifier is determined as $I_{OUT+}/\Delta V$. It is shown in Figure 8.38 for typical, worst and best cases.

Due to constant equivalent transconductance in the input signal range, the amplifier gain is also constant in the input signal range and the resulting unity-gain error is within specified values as displayed in Figure 8.39.

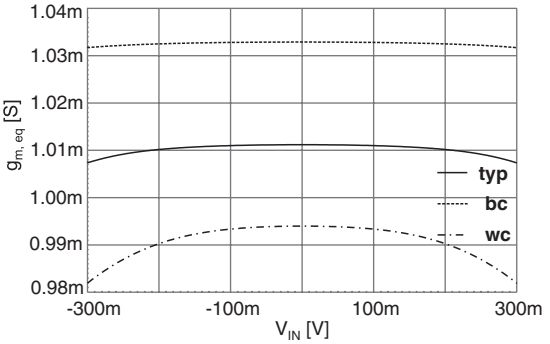


Figure 8.38 Equivalent transconductance in typical, worst and best cases

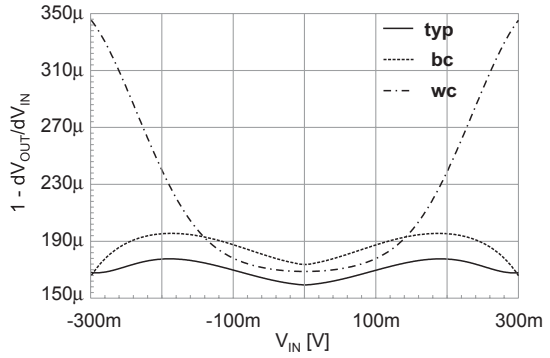


Figure 8.39 Unity-gain error in input signal range in typical, worst and best cases

Table 8.10 Time-domain system simulation results for the unity-gain buffer

interaction chain	*SNDR* [dB]
unity-gain buffer alone	92.3
unity-gain buffer + sampling stage	92.9

Time-domain system simulations - The simulation results, when the interaction with neighboring blocks is checked, are summarized in Table 8.10.

8.6 Design of the fully-differential two-stage amplifier

The amplifier in the first integrator is implemented as a two-stage amplifier to provide the required DC gain. A folded cascode OTA in the first stage is followed by a common-source output stage that ensures a high differential output swing. Since the capacitive load at the output is small, i.e. of the order of 0.1 pF, the stability of the main feedback loop is achieved using Miller cascode compensation (section 7.2). The CMFB amplifier has, as in the case of the two other amplifiers, linearized input differential pairs and cascoded current bias sources (section 7.5, section 7.6). To ensure the stability of the CMFB loop, the folded cascode OTA is realized with split sources (section 7.7). The described circuit topology is derived from the specifications given in Table 8.4 and is shown in Figures 8.40 and 8.41.

8.6.1 Equivalent output load

The capacitive load at the output of the first integrator is the input parasitic capacitance of the unity-gain buffer stage as shown in subsection 8.2.2 (Figure 8.7). The required integrator frequency $f_I = 2.6$ MHz is realized with a negative feedback network consisting of resistance $R_I = 6\ \mathrm{K\Omega}$ and capacitance $C_I = 10.8$ pF.

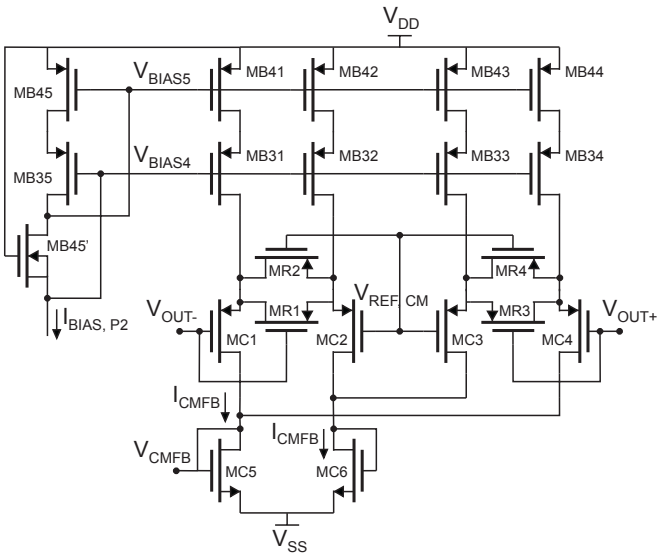


Figure 8.40 Linearized CMFB amplifier with cascoded current sources

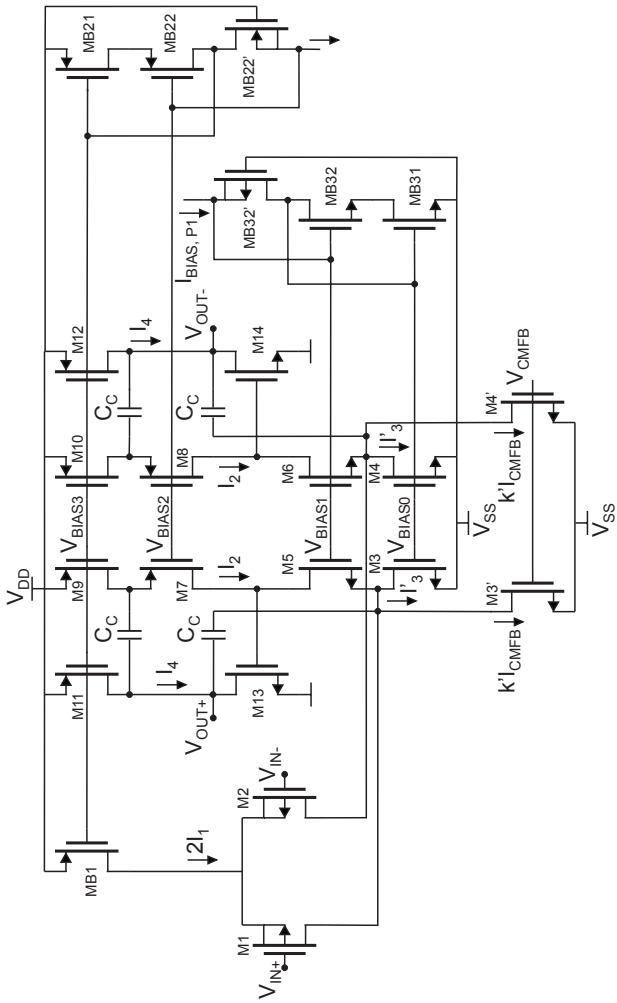


Figure 8.41 Fully-differential folded cascode OTA with common-source output stage

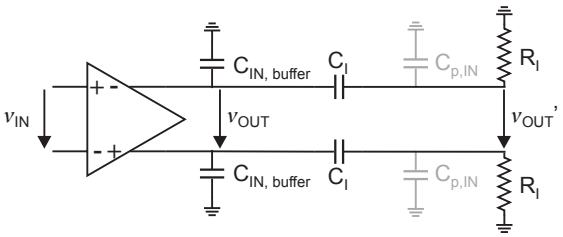


Figure 8.42 Open main feedback loop

The equivalent output load is determined, as in the case of the second integrator, from the analysis of the open main feedback loop depicted in Figure 8.42. The open loop transfer function is a product of the main amplifier transfer function $H_1(j2\pi f)$, approximated here as

$$H_1(j2\pi f) = \frac{v_{OUT}}{v_{IN}} = \frac{A_0}{\left(1 + j\frac{\omega}{\omega_{dp}}\right)\left(1 + j\frac{\omega}{\omega_{ndp}}\right)}, \tag{8.33}$$

and the feedback transfer function $H_2(j2\pi f)$, given by

$$H_2(j2\pi f) = \frac{v'_{OUT}}{v_{OUT}} = \frac{j\omega R_I C_I}{1 + j\omega R_I(C_I + C_{p,IN})} \approx \frac{j\omega R_I C_I}{1 + j\omega R_I C_I}, \tag{8.34}$$

since the integrator capacitance is larger than the input parasitic capacitance, $C_I >> C_{p,IN}$.

Figure 8.43 shows the open loop transfer function module. Obviously, the gain bandwidth frequency of the open loop is the same as the gain bandwidth frequency of the amplifier. Each pole in the main amplifier transfer function introduces a - 90° phase shift, and the zero in the feedback transfer function introduces a phase shift of + 90° as far as the integrator frequency, where it is canceled by the pole. Therefore, the stability condition is written as

$$f_{ndp,AMP} \geq k \cdot f_{GBW}. \tag{8.35}$$

In other words, the main feedback loop stability depends on the frequency response of the main amplifier, and is not affected by the negative feedback network. As a result, it follows that for the stability analysis the input parasitic capacitance of the unity-gain buffer stage represents the output load.

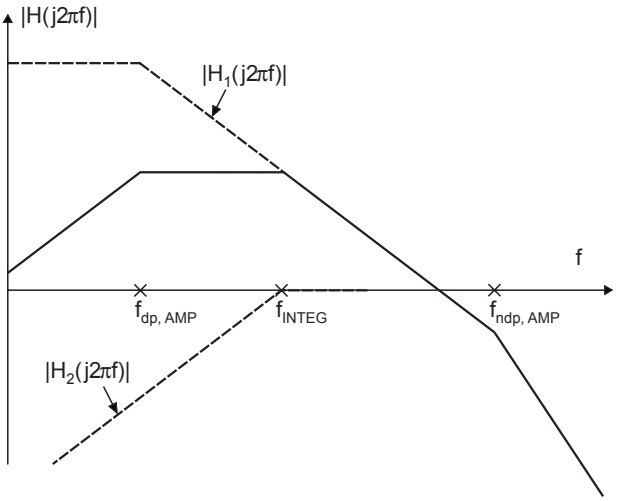


Figure 8.43 Open main feedback loop transfer function module

8.6.2 Additional design requirements for the continuous-time integrator

The transfer function module of an ideal integrator decreases with a slope of - 20 dB/dec, whereas the phase shift is - 90°. Figure 8.44 illustrates the closed loop transfer function module in the case of the real integrator, which is realized using an amplifier with finite gain and finite gain bandwidth, given by

$$H(j\omega) = A_0 \frac{(1 - j\omega / \omega_{z,\,INTEG})}{(1 + j\omega / \omega_{dp,\,INTEG})(1 - j\omega / \omega_{ndp,\,INTEG})}. \tag{8.36}$$

The transfer function is calculated using the behavioral model of the amplifier consisting of an input g_m-cell, an equivalent output resistance R_{OUT} and an equivalent output load capacitance C_L, and an negative feedback network with the integrator frequency $f_I = 1/(2\pi R_I C_I)$. The DC gain of the amplifier in this case is $A_0 = g_m R_{OUT}$. The zero and pole frequencies are summarized in Table 8.11.

The integrator function is realized in the frequency range where the frequency response corresponds to the ideal integrator, i.e. the module decreases with - 20 dB/dec and the phase shift is - 90°, as stated previously. Therefore, the real integrator starts to integrate at

$$f_{INTEG,\,START} = 10 \cdot f_{dp,\,INTEG} \tag{8.37}$$

and if the non-dominant poles of the amplifier do not appear at frequencies lower than the zero frequency it finishes at

$$f_{INTEG,\,END} = \frac{1}{10} \cdot f_{z,\,INTEG}. \tag{8.38}$$

As a result, when the amplifier in the continuous-time integrator is designed, its gain and gain bandwidth are imposed by the frequency range of the integrator function.
The minimal DC gain of the amplifier is easily calculated as

$$A_0 \geq \frac{10 \cdot f_I}{f_{INTEG,\,START}}. \tag{8.39}$$

In the case of a single-stage amplifier, the gain bandwidth frequency is determined as $f_{GBW} = g_{m1}/(2\pi C_L)$, where $g_{m1} = g_m$ of the input g_m-cell in the behavioral model. Therefore,

$$g_{m1} \geq 10 \cdot 2\pi C_I \cdot f_{INTEG,\,END} \rightarrow f_{GBW} \geq \frac{10 C_I \cdot f_{INTEG,\,END}}{C_L}. \tag{8.40}$$

If $C_I > C_L$, then $f_{GBW} > f_{z,INTEG}$, and the amplifier stability condition guarantees that the non-dominant poles of the amplifier do not appear at frequencies lower than the zero frequency.

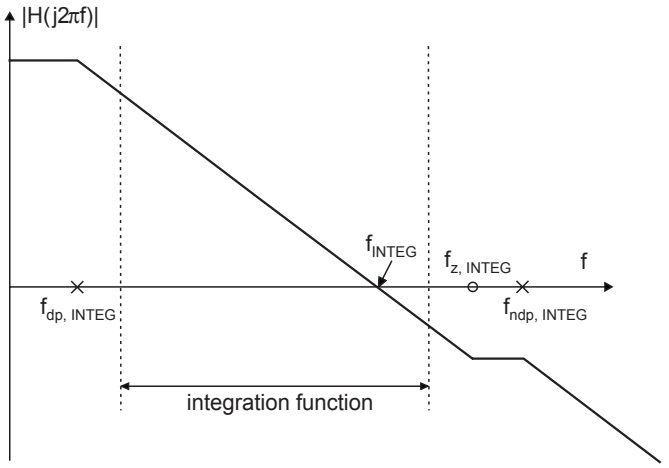


Figure 8.44 Closed loop transfer function module in the case of the real integrator

Table 8.11 Zero and pole frequencies in the closed transfer function of the real integrator

dominant pole $f_{dp,INTEG}$	$\frac{1}{2\pi R_I C_I g_m R_{OUT}} = \frac{f_I}{A_0}$
zero $f_{z,INTEG}$	$\frac{g_m}{2\pi C_I}$
non-dominant pole $f_{ndp,INTEG}$	$\frac{g_m}{2\pi C_L}$

In the case of a two-stage amplifier, the g_m of the input g_m-cell in the behavioral model can be interpreted as a product of the gain of the first stage A_1 and the equivalent transconductance of the second stage g_{m2}, that is

$$g_m = A_1 \cdot g_{m2}. \tag{8.41}$$

Since the stability of the two-stage amplifier is achieved using the compensation capacitance C_C, the gain bandwidth frequency becomes $f_{GBW} = g_{m1}/(2\pi C_C)$ and is lower than the frequencies of the zero and the non-dominant pole of the integrator. Hence, the non-dominant pole of the amplifier introduces an additional phase shift and determines the end of the integration

$$f_{INTEG,\,END} = \frac{f_{ndp,\,AMP}}{10}. \tag{8.42}$$

If the amplifier is designed with the stability condition $f_{ndp,\,AMP} \geq 10 \cdot f_{GBW}$, then the minimal gain bandwidth frequency is given by

$$f_{GBW} \geq f_{INTEG,\,END}. \tag{8.43}$$

8.6.3 Procedural design sequence

Circuit currents - Since there is no output current sourcing/sinking capability requirement for this continuous-time integrator, the circuit currents are chosen to fit the total current consumption budget. Similarly to the previously designed amplifiers the I_{CMFB} = 10 μA, whereas the current budget for the voltage and current biasing circuitry is 60 μA. The main amplifier currents are chosen as

$$I_1 = 20\ \mu A, \quad I_2 \geq 2 \cdot I_1 = 25\ \mu A, \quad I_3 = I_1 + I_2 = 35\ \mu A, \quad I_4 = 22\ \mu A. \tag{8.44}$$

Compensation capacitance - The maximal acceptable value of the compensation capacitance is 1 pF, determined according to the practical implementation requirements and the acceptable surface in the layout that can be occupied by the four compensation capacitors. However, the initial value is chosen with regard to the output load capacitance as

$$C_C = 5 \cdot C_L = 0.5\ \text{pF}. \tag{8.45}$$

This value is adjusted to 0.42 pF in the optimization design step where the transconductances of transistors M5,6 and M13,14 are changed in order to fulfill the stability conditions and to ensure under-damped complex-conjugate poles.

Differential pair - To ensure the integrator stop frequency, the gain bandwidth frequency has to be greater than 70 MHz. Therefore, the equivalent transconductance of the differential pair is $g_{m1,2} \geq 440\ \mu\text{S}$. From here it is easily calculated that $(g_m/I_{Dsat})_{1,2} \geq 22$. The transistor lengths are in this case set to L_{MIN}.

Folded pair and current sources - The transistor lengths of the folded pair transistors M5,6, as well as of the current source transistors M3,4 and M7,8, and M9,10 are set to $5L_{MIN}$ to achieve the maximal possible output resistance. In the initial design step, all transistors are placed at the limit of moderate inversion. The transconductance of transistors M5,6 is adjusted in the stability optimization step. The resulting output resistance of the first stage is $R_{OUT1} = 1\ \text{M}\Omega$, and thus a gain of 48 dB is achieved with this stage.

Output stage - The common-source stage is designed to provide low output impedance, and therefore the lengths of transistors M13, M14 are set to L_{MIN}. The resulting output impedance is $R_{OUT2} = 330\ \text{K}\Omega$. The required gain of the second stage is approximately 25 dB, which implies $g_{m13,14} \geq 50\ \mu\text{S}$ and $(g_m/I_{Dsat})_{13,14} \geq 2.5$. Nevertheless, the inversion factor is chosen taking into account the specification for the differential output swing of the amplifier, and thus the transistors M13,14 are placed in moderate inversion. The resulting $(g_m/I_{Dsat})_{13,14} \approx 14.5$ also ensures that the DC voltage imposed on the output terminals of the first stage is in the range $V_{REF,CM} \pm 0.2$ V.

Optimization step: stability of the open main feedback loop - The stability of the two-stage amplifier with Miller compensation is discussed in section 7.2. The calculations from the previous steps give $g_{m13,14} < g_{m1,2}$, and to improve the stability the equivalent transconductance of the differential pair has to be lowered. When $(g_m/I_{Dsat})_{1,2}$ is set to 18, and the value of the compensation capacitor to 0.42 pF, $PM = 80°$ is obtained. On the other hand, to decrease the damping factor, $g_{m5,6}$ is increased. After several iterations $(g_m/I_{Dsat})_{5,6}$ is set to 14. The resulting open loop transfer function is presented in subsection 8.6.4.

CMFB amplifier - The cascoded current bias sources and the input differential pairs are the same as in the case of the two other amplifiers. The transistors MC5,6 are matched with M3,4.

Split sources - The only way to improve the stability of the CMFB loop is to reduce its gain, $k \cdot g_{mC1} \cdot R_{OUT1} \cdot A_2$, by decreasing the current ratio k. For this continuous-time integrator, it is confirmed by time-domain simulations that the minimal value of $k = 1$ already gives a satisfactory result.

Transistor sizes - Complete schematics with final dimensions are presented in Figures 8.45, 8.46 and 8.47. The transistor sizes are real sizes. The circuit currents are also indicated.

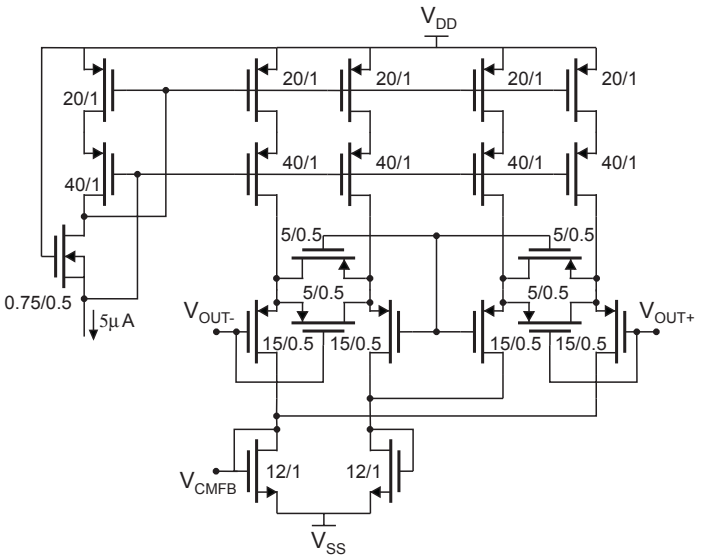


Figure 8.45 Complete schematic with final dimensions: CMFB amplifier

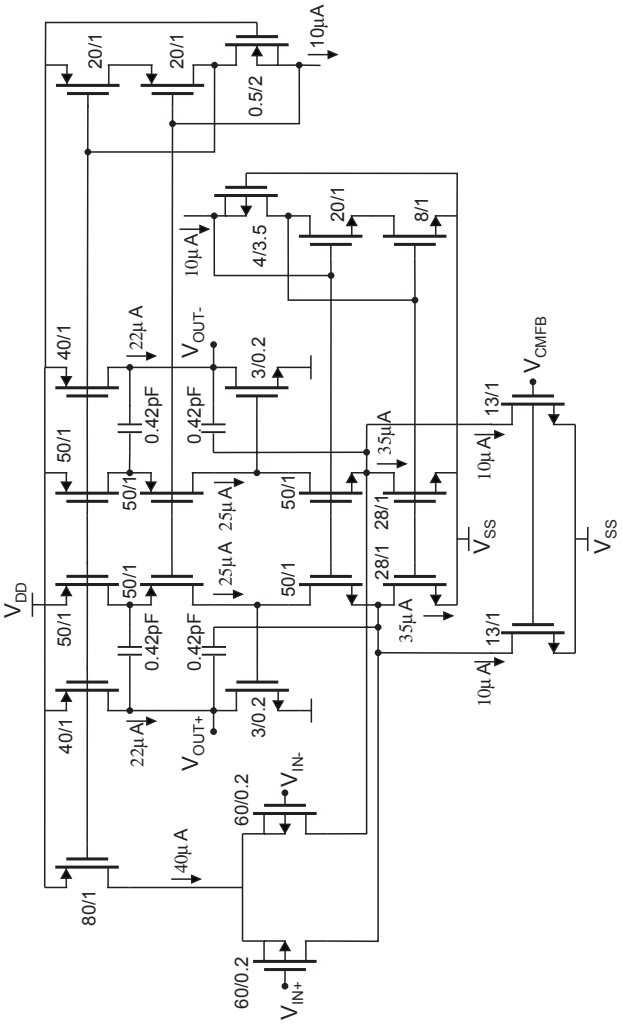


Figure 8.46 Compete schematic with final dimensions: main amplifier

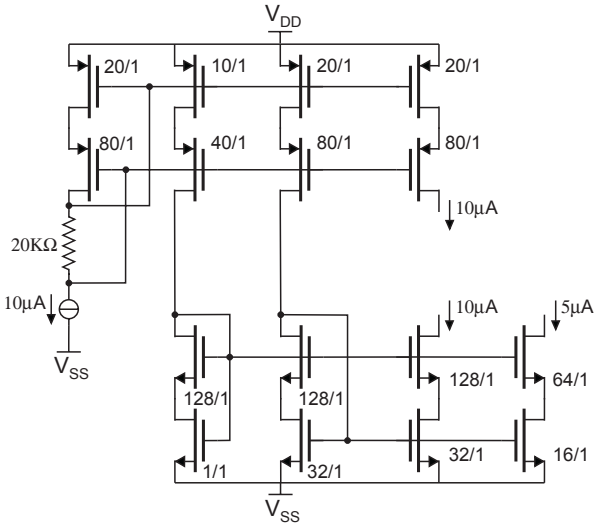


Figure 8.47 Complete schematic with final dimensions: biasing circuit

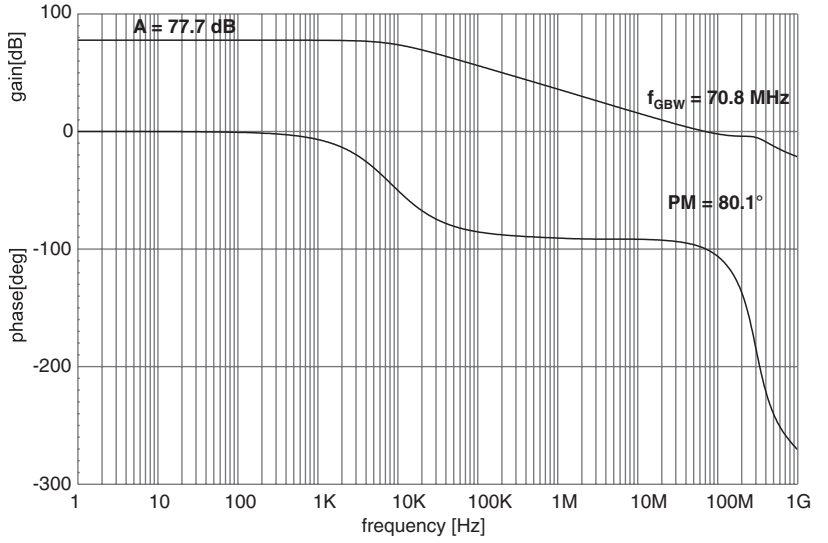


Figure 8.48 Open main feedback loop transfer function module and phase

8.6.4 Simulation results

AC analysis - The transfer function module and phase of the open main feedback loop are shown in Figure 8.48.

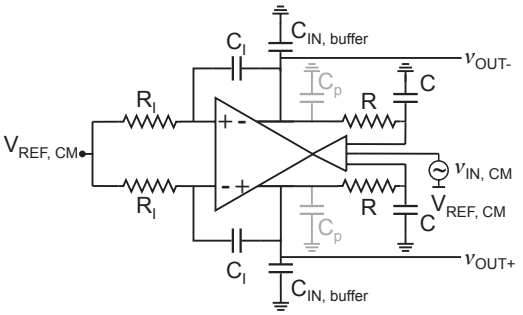


Figure 8.49 Open CMFB loop

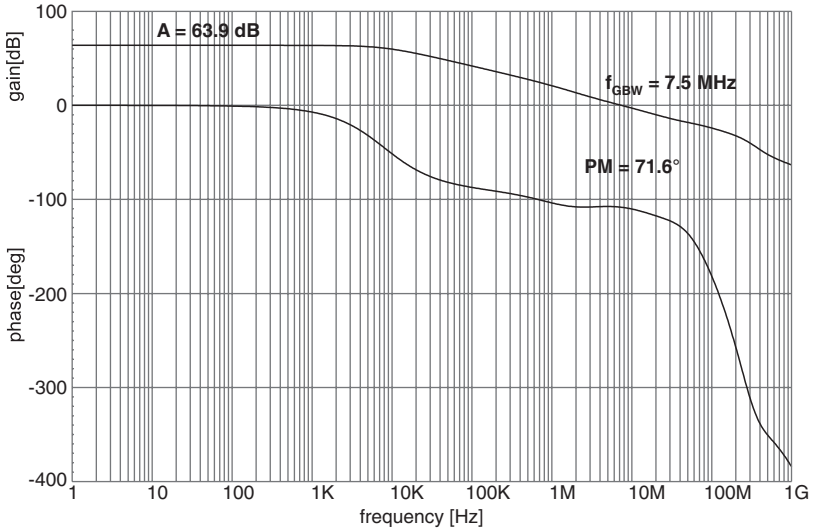


Figure 8.50 Open CMFB loop transfer function module and phase

The open CMFB loop is simulated using the circuit in Figure 8.49 (the amplifier symbol represents the main amplifier connected with the CMFB amplifier). The resulting module and phase are shown in Figure 8.50.

DC analysis - The differential output swing is checked by adding a resistance in parallel with the integrator capacitance and setting the closed loop gain in this way to 100. The result is displayed in Figure 8.51. Since the differential pairs at the input of the CMFB amplifier are linearized, the differential output swing is superior than the specified signal range. In addition, the variations of the common-mode output level are reduced to ±3 mV.

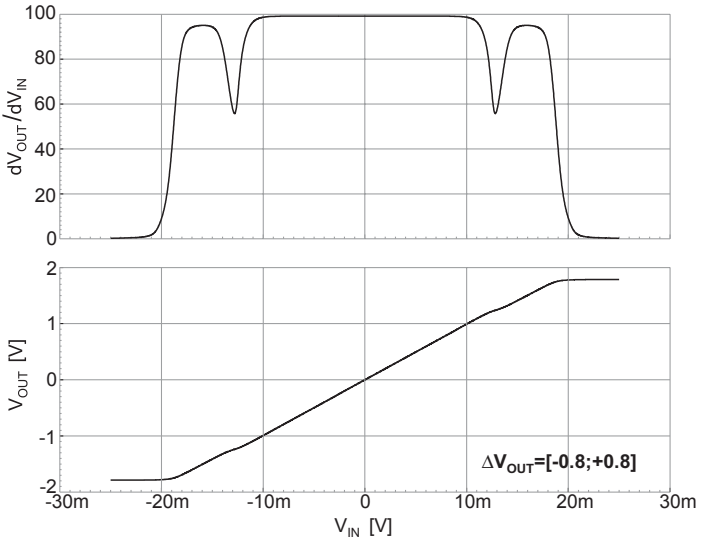


Figure 8.51 Differential output swing of the two-stage amplifier with a linearized CMFB amplifier

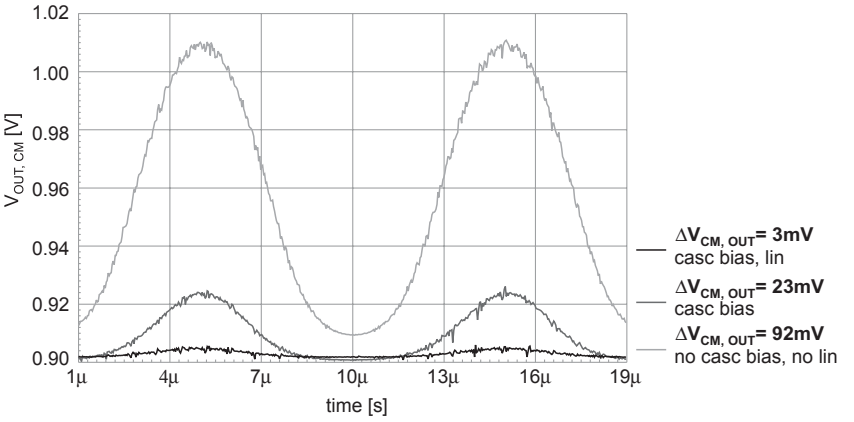


Figure 8.52 Output common-mode variation in three different cases

Time-domain system simulations - Figure 8.52 illustrates the improvement in the rejection of the differential signal in the common-mode output signal by replacing the differential pairs at the input of the CMFB amplifier by linearized differential pairs and cascoding their current bias sources. This enables to overcome the problems of interaction with the unity-

Table 8.12 Time-domain system simulation results for the first integrator

interaction chain	*SNDR* [dB]
first integrator alone	93.8
first integrator + unity-gain buffer	94.1
first integrator + unity-gain buffer + current-steering DAC	94.3

gain buffer stage. The *SNDR* values extracted from time-domain simulations are given in Table 8.12.

8.7 Conclusion

Three different amplifiers that are parts of a Delta-Sigma modulator system are designed in this chapter using the structured design approach and a fully-differential folded cascode procedural design scenario proposed earlier in this thesis. The specifications for each topology are derived from the system-level requirements using the behavioral models and time-domain simulations. The amplifiers are designed using the extracted EKV model parameters, whereas the final design sizes are confirmed using simulations with the BSIM model parameters. The performances, confirmed from the time-domain simulations with system blocks implemented at the transistor level, are within the specified range.

Bibliography

[1] S. Pesenti, “Hybrid Continuous-discrete-time Delta-sigma A/D Converters with Autoranging Algorithm”, PhD thesis N°3941, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2007

[2] Y. Geerts, M. Steyaert, W. Sansen, “Design of Multi-Bit Delta-Sigma A/D Converters”, Springer, May 2002, ISBN: 1402070780

[3] M. Kayal, Z. Randjelovic, “Auto-zero differential difference amplifier”, Electronics Letters, Volume 36, pp. 695-696, April 2000

[4] P. Bendix, “Detailed comparison of the SP2001, EKV, and BSIM3 models”, Technical Proceedings of the 2002 International Conference on Modelling and Simulation of Microsystems, Nanotech 2002, Volume 1, pp. 649-652

[5] K. Bult, G. Geelen, “A fast settling CMOS op amp for SC circuits with 90-dB DC gain”, Journal of Solid-State Circuits, Volume 25, pp. 1379-1384, December 1990

Index

E

F

G

M

P

S

T

V