1993

A digital tuning scheme for digitally programmable integrated continuous-time filters and techniques for high-precision monolithic linear circuit design and implementation

Chong-Gun Yu
Iowa State University

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

U·M·I

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Order Number 9414041

A digital tuning scheme for digitally programmable integrated continuous-time filters and techniques for high-precision monolithic linear circuit design and implementation

Yu, Chong-Gun, Ph.D.

Iowa State University, 1993

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

A digital tuning scheme for digitally programmable integrated continuous-time filters and techniques for high-precision monolithic linear circuit design and implementation

by

Chong-Gun Yu

A Dissertation Submitted to the

Graduate Faculty in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Department: Electrical Engineering and Computer Engineering
Major: Electrical Engineering (Microelectronics)

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa
1993

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like to express my heartful thanks to my advisor, Dr. Randall L. Geiger. I can never thank him enough for his cordial guidance, encouragement, and support during my Ph.D. studies at Texas A&M University and Iowa State University. His technical excellence and foresight have been of great help to me and have had an important effect on my life.

I would like to thank Dr. Stanley G. Burns, Dr. Hsien-Sen Hung, Dr. Marwan Hassoun, and Dr. Peter Sherman for serving as my graduate committee and for sharing their busy schedules to help me with my research. I also would like to acknowledge the academic assistance from many faculties and friends, especially Dr. W. G. Bliss, Dr. T. M. Scott, and Dr. K. H. Loh.

I also would like to thank Dr. R. K. Hester of Iowa State University and Mr. W. Dietrich of Texas Instruments Inc. for their valuable discussions and comments on one of my research topics. Dr. B. Brandt of Texas Instruments Inc. is acknowledged for his assistance with the circuit fabrication, and Texas Instruments Inc. is acknowledged for support used on this project.

Finally, I would like to thank my parents and my brothers for their self-sacrificing love and encouragement.

CHAPTER 1. INTRODUCTION

As the title indicates, this dissertation contains multiple topics and can be classified into two parts in a large way or five topics in a detailed way. The first part is concerned with a digital tuning scheme for digitally programmable integrated continuous-time filters [4],[8]-[10]. The second part includes four topics:

- Nonideality consideration for high precision amplifiers – Analysis of random common-mode rejection ratio [5],[7].
- An automatic offset compensation scheme with ping-pong control for CMOS operational amplifiers [2],[6].
- Very low voltage circuits and operational amplifiers using floating gate MOS transistors [3].
- An accurate and matching-free V_T extractor using a ratio-independent SC subtracting amplifier and a dynamic current mirror [1].

which can be grouped under the name of *techniques for high-precision monolithic linear circuit design and implementation.*

Since the five topics are not directly related to each other, each topic is presented in a separate chapter, and each chapter has a full organization including introduction, main body, and conclusion. Actually, these topics are based on the journal or conference papers which have been already published or are to be published. Although the topics are not directly connected with each other, they can be correlated in that they are all concerned with theories and techniques for high-precision linear circuit design and implementation.

To obtain high-precision linear integrated circuits, the causes degrading their accuracy must be well understood, and appropriate measures should be taken to compensate for them. These are what the five topics are concerned with. In the first topic, continuous-time filters are digitally tuned for high-precision continuous-time filtering. This is one of the attempts to digitally solve the problems such as process variations and parasitic effects which are inherent

to monolithic linear circuits and degrade their performance by deviating the fabricated circuits from desired circuits. To obtain high-precision linear integrated circuits, it is required that op-amps which are the most fundamental circuits be implemented with high accuracy. Thus, in the second topic, the nonidealities associated with op-amps are analyzed. The analyses include random CMRR and offset and their statistical characteristics which will be of much help to obtaining high-precision integrated op-amps. In the third topic, op-amp offset voltages are automatically compensated by digital means. The offset is one of the important obstacles pulling op-amps out of high-precision, and thus must be compensated for high-precision applications. With increased low-voltage circuit applications, low-voltage as well as high-precision linear integrated circuits become essential. In the fourth topic a promising scheme to obtain high-precision linear circuits that can be operated with a very low power supply is proposed. Finally, a high-precision threshold voltage extraction scheme applicable to many real-time on-chip applications is discussed. The threshold voltage extractor is also required for the low-voltage circuit implementation.

Chapter 2 is concerned with a digital tuning scheme. For high-precision monolithic filtering, a continuous-time filter must be tuned after fabrication because of large component variations and undesired parasitic effects. This problem has motivated the introduction of many tuning techniques. In this chapter, as a new promising high-precision tuning technique, a digital tuning scheme is presented for digitally programmable/tunable continuous-time filters which have many broadband applications where accuracy as well as reconfigurability to various filter functions in a wide range of frequency are crucial for the system. The tuning scheme consists of two steps, system identification (ID) and adjustment. Various methods for continuous-time filter identification are investigated on the basis of accuracy and efficiency from both time requirements and silicon implementation viewpoints. An adjustment tuning algorithm is presented which uses the system ID results to estimate process dependent parameters and then to calculate filter control parameters for adjustment. Extensive Monte-Carlo based simulation results and some experimental results are also presented to evaluate the performance of the digital tuning scheme. Finally, the first statistical characterization scheme for rigorously assessing the performance of any tuning algorithm is introduced.

In Chapter 3, nonideal factors which play a key role in performance and yield in high-precision applications of operational amplifiers are rigorously investigated. Of necessity, the combined effects of both deterministic and statistical parameters must be incorporated. The statistical characteristics of the common-mode rejection ratio and the offset of two-stage CMOS op-amps are investigated. The op-amp errors associated with finite open-loop gains, finite

CMRRs, and nonzero offset voltages are compositely analyzed.

In many op-amp applications, offset cancellation or reduction is critical because an amplifier input offset voltage limits the capability of the system. An automatic offset compensation scheme for CMOS operational amplifiers is presented in Chapter 4. Offset is reduced by digitally adjusting the bias voltage of a programmable current mirror which is used as the load of the differential input stage. A 100% operating duty cycle is obtained by using a ping-pong structure. The offset compensation scheme is inherently time and temperature stable since the offset compensation is periodically performed with the ping-pong control. The proposed circuit has been fabricated using a 1.0-μm n-well CMOS process. The measured offset voltages of the test circuits are less than 400μV in magnitude.

With emergence of an increasing number of battery-operated applications, low voltage circuit techniques have moved into the limelight. Following the trend, a threshold voltage tunable op-amp structure that can be operated with a very low power supply (e.g. 0.5V) is presented in Chapter 5. Since the threshold voltage of a floating gate transistor can be precisely controlled, use of floating gate MOS transistors as the basic circuit element makes it possible to obtain much lower voltage circuits than achievable by other approaches using device size scaling techniques where the threshold voltage variation is significantly increased as the device size is decreased. Good matching can also be achieved by tuning the threshold voltages. A two-step threshold voltage tuning scheme is presented. Due to the long term charge retention property of the floating gate transistors, the threshold voltage tuning does not have to be done frequently, and thus, near continuous-time operation of the op-amp can be achieved.

In Chapter 6, an accurate threshold voltage extraction scheme for MOS transistors is presented. In contrast to alternative methods recently reported in the literature, the proposed scheme does not need a matched replica of the transistor under test. Moreover, the scheme can be accurately implemented in a matching-free way. Thus, the scheme has potential of extracting threshold voltages much accurately than other techniques of which the performances depend upon the test device matching as well as other component matching in their extraction circuits. The proposed scheme is implemented using a ratio-independent switched-capacitor subtracting amplifier and a dynamic current mirror. Nonideal effects associated with these circuits are thoroughly investigated.

CHAPTER 2. A DIGITAL TUNING SCHEME FOR DIGITALLY PROGRAMMABLE INTEGRATED CONTINUOUS-TIME FILTERS

2.1 Introduction

While digital filters have many advantages due to their inherent characteristics such as signal processing in digital form and easiness in design, analysis, and testing automation, continuous-time (analog) filters may offer advantages over digital filters in many respects. Integrated continuous-time filters require orders of magnitude less die area than comparable digital filters and do not need any domain transformation (continuous-time to discrete-time or vice versa) which is essential for digital filtering when the signal to be processed is in continuous-time form. Many problems which are involved in the peripheral parts required for the transformation such as A/D and D/A converters, anti-aliasing filters and reconstruction circuits are main drawbacks in digital filtering [19]. The most attractive feature of the pure continuous-time filtering is the capability of operating at higher frequencies than its sampled-data counterparts such as switched-capacitor, switched-current, and digital filters [11]-[15]. These reasons make the continuous-time filtering preferable in certain applications.

In spite of the advantages of continuous-time filters, their use has been limited because of the significant discrepancy between the designed and fabricated filter characteristics due to large component tolerances and parasitics. Therefore, tuning is essential, and it is the most serious and challenging problem which must be overcome to obtain high-performance continuous-time filtering. In the past many kinds of techniques have been developed for tuning of various continuous-time filters [13]-[37]. These can be categorized as functional methods, deterministic methods with automatic tuning algorithms, and automatic on-chip methods with analog tuning loops associated with master-slave techniques.

Functional Tuning: In functional tuning components are adjusted on a one-by-one basis while an excitation is applied to the circuit and measurements are made. Although functional

tuning may be efficient in some cases, it is generally precluded because it is inherently slow and depends on unclear heuristics. As early versions of tuning techniques, the functional methods and the deterministic methods have been applied to mainly hybrid integrated active RC filters for low-frequency filtering applications.

Deterministic Tuning: In deterministic tuning [26]-[33] the necessary adjustments are analytically calculated from a set of component measurements. Many deterministic tuning algorithms have been proposed, and some of them have been successfully implemented and used for analog filter products. Three representative deterministic tuning algorithms which have been compared in [26], [27], are a least squares method [31], a sequential tuning algorithm [28],[29], and a large-change-sensitivity method [30]. In these methods only a subset of the components, usually resistors, are adjusted in an irreversible and increasing manner based upon the measurements of the remaining components, usually capacitors and resistors. The methods require solving complex nonlinear equations derived from circuit analysis at a set of discrete frequency points. Therefore, a resistor and frequency selection procedure is essential in these methods and should be handled carefully because the performance of these algorithms is directly related to these choices.

The adjustment method of deterministic tuning is trimming the circuit resistors by physical or chemical methods, so the trimming errors are unavoidable. Sometimes the required resistor trim tolerances are so tight that satisfactory results can not be obtained by the state of the art of the available trimming technology. The trimming error, of course, can be minimized if the closed-loop method of the sequential tuning algorithm is used [28]. Drawbacks of these methods are that much task is required for deriving the required expressions and a sophisticated on-line computer with a large software support system is also required for its implementation. In addition these algorithms have been applied only at low frequencies, and the tunability has not been considered at higher frequencies where parasitics have much more dominant effects on the filter performance than component tolerances. These algorithms have only limited success at compensating for component variations, and thus, have been restrained to tuning circuits already close to the desired circuits.

Analog Loop Tuning: The most popular approaches to tuning of high-frequency continuous-time filters have been to use on-chip analog tuning loops such as phase locked loops (PLL) and vector locked loops (VLL) based upon the master-slave concept [13]-[23]. These methods have been applied to high-frequency OTA-C continuous-time filter implementations, and good results

have been achieved. Most early versions of these techniques used one single reference frequency only for pole resonance frequency control, and Q-control was not considered necessary under the assumption that the quality factor has sufficient accuracy by component ratios. For highly selective filters, this assumption is not correct any more because parasitic effects produce significant errors in the effective quality factor and thus the gain at the resonance frequency. To overcome this problem more complex tuning schemes have been reported [13]-[17] which control two objectives, quality factor control as well as frequency control. It has been shown that controlling more objectives results in better performance at the expense of more complexity of tuning circuits. The primary limitation of these kinds of methods using the master-slave scheme is the mismatch errors between reference circuits and main filters and other undesired effects associated with the large extra analog tuning circuits.

Digital Tuning: Although attention has been concentrated on high-frequency filtering as a promising application area of the continuous-time filters, many broadband continuous-time filtering applications such as telephony and radio also exist where accuracy as well as reconfigurability to different types of filter functions in a wide range of frequency are required for the system [39]. For these kinds of applications, digitally programmable and digitally tunable continuous-time filter architectures have been developed [38]-[43]. This kind of reconfigurable filter should be served by more general tuning schemes that can perform tuning of various filter transfer functions. A feasible way to do that is opening a tuning host or employing a tuning microprocessor to perform a well developed software tuning algorithm. This tuning scheme is referred to as *digital tuning.*

In the digital tuning, the actual filter performance is measured, and the measured data in digital form are transferred to a computer/microprocessor that performs a tuning algorithm to digitally control the continuous-time filter. This technique does not require component matching as in the master-slave scheme. Although the digital tuning also has some disadvantages that it requires an external tuning host computer/microprocessor and precision filter performance measurement circuitry and that the performance is limited by the quantization effects of the digital control mechanism and the accuracy of the performance measurement circuit, it has potential of high accuracy as well as applicability to high-Q and high-frequency applications if a good tuning algorithm is provided.

The comparison of the tuning schemes mentioned above is summarized in Table 2.1. A few digital tuning schemes have been reported in [40],[41]. Their work has demonstrated the

Table 2.1: Comparative summary of the tuning schemes

	Functional Tuning	Deterministic Tuning	Analog Loop Tuning	Digital Tuning
Main Idea	Heuristics	Circuit Analysis	Master-Slave Scheme	System ID Adjustment
Adjustment Method	Laser Trimming	Laser Trimming	Analog Control	Digital Control
Required Measurement	Frequency Responses	A Set of Components	No	Frequency Responses
Excitation	Yes	No	Yes	Yes
Frequency Range	Low	Low	High	High
Charac-terristics	- Efficient when filters are simple - Inherently slow	- Formidable task to derive the required expressions - Required resistor trim tolerance are tight - Large software support system - Require accurate system model - Applicable at only low frequency range - Suffer from aging	- On-chip automatic tuning - Require large extra analog tuning circuits - Mismatch errors between master and slave filters - Signal interference - Limited accuracy	- Simplified by decomposing it into two phases - Require software support system - Require digital contol circuits and measurement circuits - Potential of high accuracy

applicability of the digital tuning methods to high-precision and high-frequency filter applications. However, their tuning schemes were developed for only 2nd-order bandpass filters, and extensions to versatile filter types or to higher-order filter functions have not been considered. The objective of this research is thus to develop a more general digital tuning scheme for digitally controllable continuous-times filters such that it can be well applicable to any type or any order filter functions. The whole tuning procedure is simplified by dividing it into two phases: system identification (ID) and adjustment. The transfer function of the filter to be tuned is first identified from input-output time-domain samples. Based upon the identified results, new filter control parameters are estimated to adjust the filter. This procedure is repeated until a tuned filter is obtained.

Section 2.2 describes a general and simplified formulation of the digital tuning problem. In Section 2.3 the digitally programmable continuous-time filter structure [38, 39] is briefly reviewed which is the basic test structure for the proposed tuning algorithm. Nonidealities of

the filter structure which have strong effects on the tuning performance are also investigated.

In Section 2.4, various continuous-time system ID methods are described. Two approaches are investigated. One is via the z-domain system ID followed by z-to-s transformation. The other is based on the s-domain system ID following frequency response measurements. An iterative complex least squares (ICLS) algorithm [9] which can be referred to as the s-domain counterpart of the z-domain generalized least squares (GLS) algorithm is proposed as a robust s-domain system ID method. Very accurate domain transformation (s-to-z and z-to-s) methods based on complex LS algorithms are also proposed and compared with the well-known bilinear transformation method. Frequency response measurement algorithms required for the second approach are comparatively reviewed on the basis of accuracy and implementation cost [4]. Extensive simulations based on the Monte Carlo method are performed to investigate the performance of the continuous-time system ID methods.

In Section 2.5, a tuning (adjustment) algorithm [8] related to the physical filter structure is introduced. The proposed tuning algorithm combined with the system ID method is extensively simulated to evaluate the whole digital tuning scheme. Yield calculation of the tuned filters is also performed to more clearly investigate the performance of the tuning scheme. Some experimental tuning results are presented in this section. Finally, conclusive remarks including the contribution of this research, limitations, and possible future works are given in Section 2.6.

2.2 Tuning Problem Formulation

The digital tuning procedure is partitioned into two phases: system identification, and adjustment as shown in Fig 2.1. The input-output samples of the actual continuous-time filter which has an unknown transfer function $T_a(s)$ are collected by a performance measurement circuit. The input-output data are then fed to a system ID algorithm to estimate the parameters of a model $T_i(s)$. The estimated model parameters C_i which is the coefficient vector of the identification model $T_i(s)$ is compared with the desired model parameters C_d which is the coefficient vector of the desired model $T_d(s)$. Based on the comparison, new filter control parameters are calculated by an adjustment algorithm and then the physical filter components are adjusted by ΔG where G is the filter component vector.

The initially implemented continuous-time filter response $T_a(s)$

$$T_a(s) = \frac{\sum_{i=0}^{m_a} b_{ai} s^i}{\sum_{i=0}^{n_a} a_{ai} s^i}$$

usually differs from the desired response $T_d(s)$ due to the component variations and parasitic

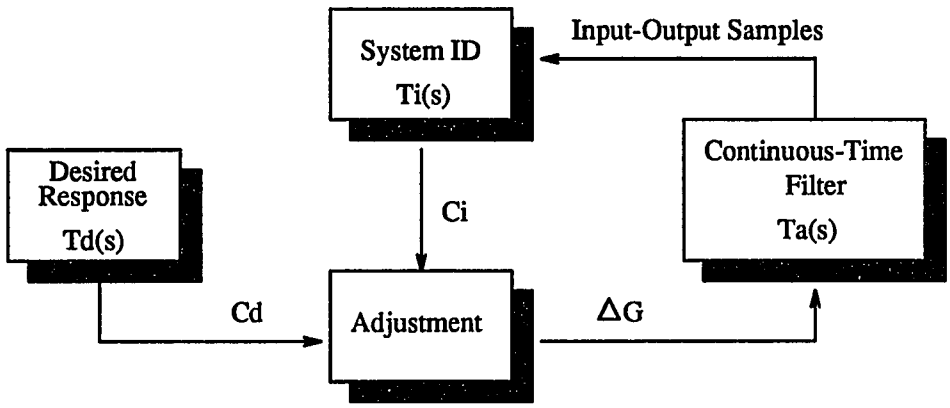


Figure 2.1: Block diagram of the digital tuning scheme

effects. The actual filter response $T_a(s)$ consists of two components and can be expressed, for convenience, as

$$T_a(s) = T_c(s)T_p(s).$$

where $T_c(s)$ is modeled, controllable/programmable and capable of realizing the desired response.

$$T_c(s) = \frac{\sum_{i=0}^{m_c} b_{ci}s^i}{\sum_{i=0}^{n_c} a_{ci}s^i}.$$

The order of $T_c(s)$ is the same as that of $T_d(s)$. The $T_p(s)$ is due to parasitics. The order of $T_p(s)$ is unknown, so the order of the actual filter response $T_a(s)$ is also unknown but will be greater than that of the desired response $T_d(s)$. In reality, $T_a(s)$ can not be expressed as a simple multiplication of two independent terms $T_c(s)$ and $T_p(s)$ since they are usually correlated. If the coefficient vector C_c and the component vector G are defined by

$$\begin{aligned} C_c &= [a_{c0} \ldots a_{cn_c}, b_{c0} \ldots b_{cm_c}]^T \\ G &= [R_1 \ldots R_k, C_1 \ldots C_l]^T, \end{aligned}$$

then the coefficient C_c is a function of G,

$$C_c = f(G) \tag{2.1}$$

It is now obvious that the tuning problem is to control the controllable components of the component vector G such that $T_a(s)$ approaches to $T_d(s)$. The whole tuning procedure is then as follows:

1. Measure the time-domain input and output samples of the continuous-time filter to be tuned.
2. Identify the actual filter with a system ID model $T_i(s)$ using the measured data.
3. Compare between the desired response $T_d(s)$ and the identified result $T_i(s)$ and determine the required ΔC_c such that $T_a(s)(\approx T_i(s))$ approaches to $T_d(s)$.
4. Map ΔC_c to the required ΔG and make component adjustment.
5. Repeat the above steps until tuning is completed.

It can be seen that the tuning performance will highly depend on the accuracy of the system ID. The system ID model $T_i(s)$ should have the same order as the desired response $T_d(s)$ to easily obtain the ΔC_c and ΔG and thus to make the adjustment algorithm simple. Thus, the system ID must be robust in the presence of parasitics because the order of the system ID model is usually less than that of the actual filter to be identified.

2.3 Digitally Programmable Continuous-Time Filter Architecture

The digitally programmable continuous-time filter architecture [39] is shown in Fig. 2.2. This structure has been selected specifically as a basic test vehicle for investigating the performance of the digital tuning algorithm. The system consists of an analog bus, a digital bus, a local digital controller, a performance monitor and a number of digitally programmable biquadratic sections. The structure of each biquadratic block is shown in Fig. 2.3, which consists of 5 programmable operational transconductance amplifiers (OTA), two programmable capacitor arrays, an analog buffer stage, and six analog configuration switches. More details of the architecture and characteristics of each elements can be found in [38]-[43].

The ideal transfer function of the programmable biquadratic block is given by

$$\frac{V_{out}}{V_{in}} = \frac{(B_{hp})s^2 + (\frac{g_{m4} - g_{m3}B_{bp}}{C_7})s + (\frac{g_{m1}g_{m3} + g_{m2}g_{m3}B_{lp}}{C_6C_7})}{s^2 + (\frac{g_{m5}}{C_7})s + \frac{g_{m2}g_{m3}}{C_6C_7}} \tag{2.2}$$

where the B variables can be 0 or 1 depending upon the switch settings. The pole frequency ω_o and the quality factor Q of the biquad are given by

$$\omega_o = \sqrt{\frac{g_{m2}g_{m3}}{C_6C_7}}$$

$$Q = \sqrt{\frac{g_{m2}g_{m3}}{C_6C_7}} \Big/ \left(\frac{g_{m5}}{C_7}\right)$$

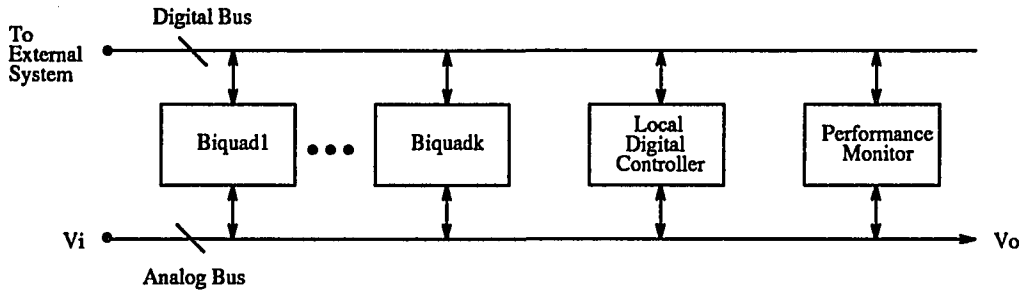


Figure 2.2: Digitally programmable continuous-time filter architecture

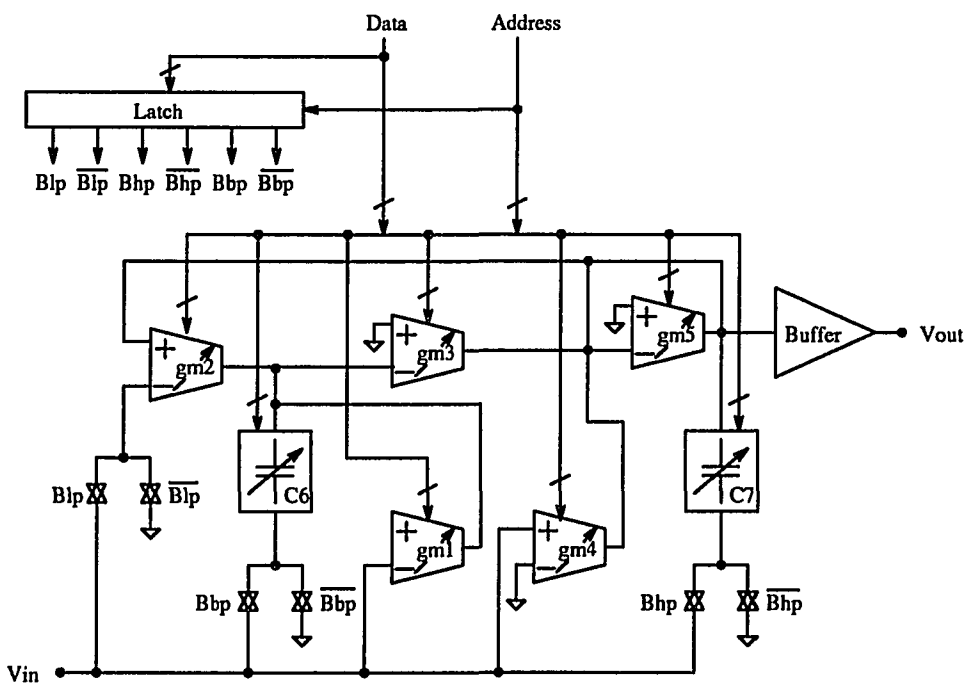


Figure 2.3: Biquadratic building block

The model for $T_c(s)$ and $T_i(s)$ should be

$$T_c(s) = T_i(s) = \frac{s^2 + b_1 s + b_0}{s^2 + a_1 s + a_0} \tag{2.3}$$

which has only 4 degrees of freedom, instead of 5 in ordinary second order rational transfer functions. The coefficient vector and the component vector are then

$$\begin{aligned} C_c &= [a_0, a_1, b_0, b_1]^T \\ G &= [g_{m1} \ldots g_{m5}, C_6, C_7]^T \end{aligned}$$

The coefficients are given respectively by

$$a_1 = \frac{g_{m5}}{C_7} \tag{2.4}$$

$$a_0 = \frac{g_{m2} g_{m3}}{C_6 C_7} \tag{2.5}$$

$$b_1 = \frac{g_{m4} - g_{m3} B_{bp}}{C_7} \tag{2.6}$$

$$b_0 = \frac{g_{m1} g_{m3} + g_{m2} g_{m3} B_{lp}}{C_6 C_7} \tag{2.7}$$

These equations show that C_c is a function of G as in (2.1). From (2.4)-(2.7), it follows that we can get independent or sequential adjustment of the transfer function coefficients.

Compared to conventional analog filters, the digitally programmable continuous-time filter structure has many advantages which are favorable to the filter adjustment. The digitally controllable structure provides more accuracy, flexibility, and simplicity for filter adjustments as follows.

1. Since the operational transconductance amplifier (OTA) gains are programmed or adjusted by digitally controlled voltages, more accurate adjustments can be possible than in the conventional analog filters where resistors are usually adjusted by laser trimming. The speed of digital adjustments is much faster than that by trimming.

2. Digital tuning can be more flexible than the conventional deterministic tuning because in the digitally programmable continuous-time filter the OTAs can be adjusted to have any gains in a given range at any time, but in the conventional filters resistors can be adjusted only in an irreversible and increasing manner.

3. Since the digitally programmable continuous-time filter consists of cascaded biquadratic blocks, the individual second-order sections of the filter can be separately tuned to adjust the whole filter, and the independent or sequential adjustability of the filter parameters

can make the computing procedures for the tuning corrections very simple not including any linearization procedures or matrix inversions as in the conventional deterministic tuning [26]-[33].

4. The digital tuning does not suffer from the mismatch errors which are the major drawbacks in the conventional analog loop tuning [13]-[23], because the digitally programmable continuous-time filter does not have any reference circuits which should be well matched with main filters.

On the other hand, this structure also has some nonidealities which affects the tuning performance. The nonidealities associated with OTA-C integrators and their effects are discussed in the following subsection.

2.3.1 Effects of Nonideal OTA-C Integrators

A basic component consisting of the biquadratic blocks is the OTA-C integrator shown in Fig. 2.4(a). The ideal transfer function in (2.2) holds under the assumption that the integrators are ideal, and it has been derived using the ideal model shown in Fig. 2.4(b). The ideal integrator transfer function is

$$T_{int}(s) = \frac{V_{out}}{V_{in}} = \frac{g_{mo}/C_L}{s} = \frac{\omega_u}{s}$$

where g_{mo} is the dc transconductance gain of the OTA and ω_u is the unity-gain frequency given by

$$\omega_u = \frac{g_{mo}}{C_L}. \tag{2.8}$$

The frequency responses of the ideal integrator is shown in Fig. 2.5(a). In reality each OTA has nonidealities such as parasitic poles and zeros associated with internal nodes and finite output impedance resulting in a finite dc gain of the OTA or the OTA-C integrator.

The parasitic poles and zeros of an OTA can be modeled by a single pole (ω_p) as follows. If an OTA with infinite output impedance has n parasitic poles and m parasitic zeros, then the transfer function of the OTA-C integrator can be written as

$$T_{int}(s) = \frac{\omega_u}{s} \frac{(1+s/\omega_{z1})\cdots(1+s/\omega_{zm})}{(1+s/\omega_{p1})\cdots(1+s/\omega_{pn})} \tag{2.9}$$

Since the parasitic poles and zeros are usually located at higher frequencies than the unity-gain frequency ω_u, by assuming the frequency range of interest $\omega << \omega_{pi}, \omega_{zi}$, equation (2.9) can be

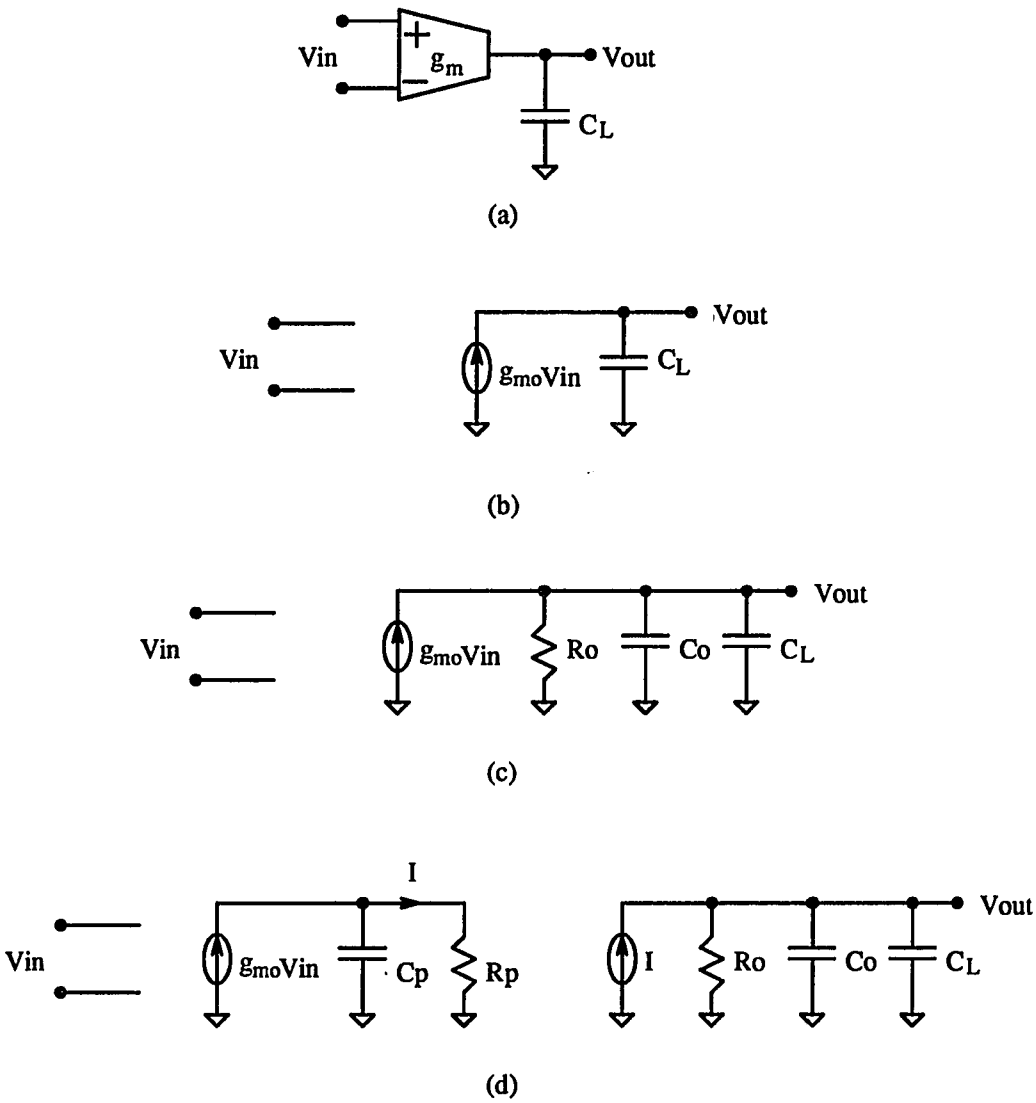


Figure 2.4: OTA-C integrator models (a) Circuit diagram (b) Ideal model (c) Model including the finite output impedance (d) Model including both the effective parasitic pole and the finite output impedance

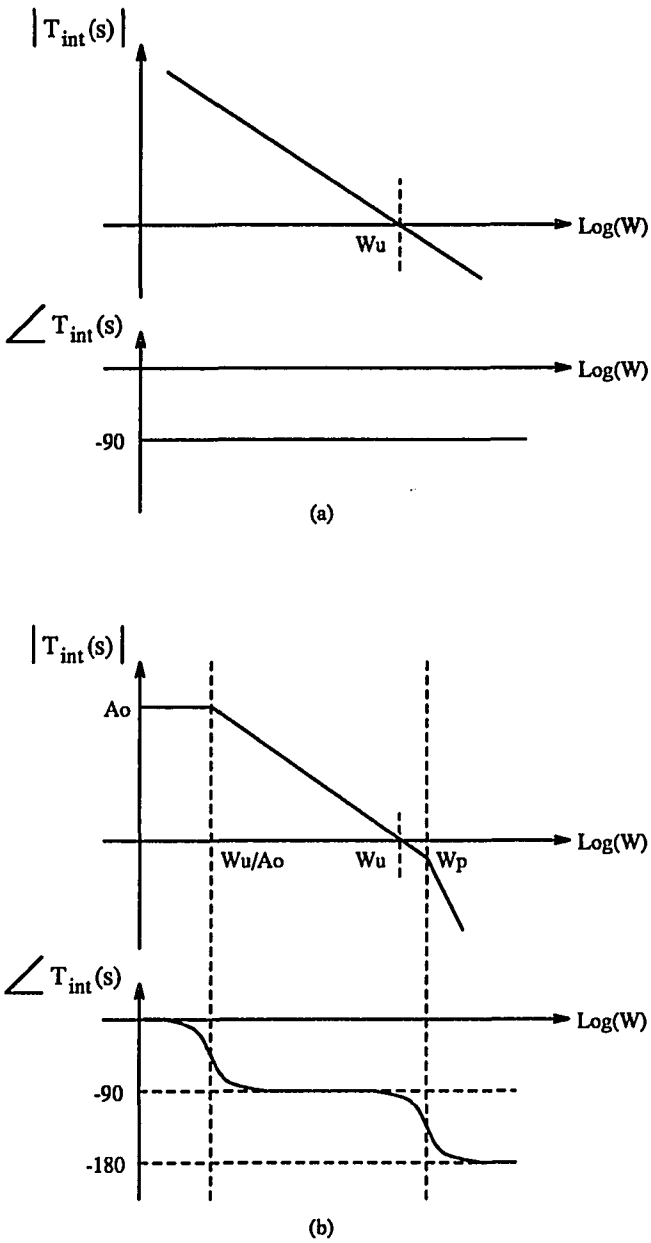


Figure 2.5: Frequency responses of (a) Ideal integrators (b) Nonideal integrators

approximated by

$$\begin{aligned} T_{int}(s) &\simeq \frac{\omega_u}{s}\frac{1}{(1+s/\omega_{p1})\cdots(1+s/\omega_{pn})(1-s/\omega_{z1})\cdots(1-s/\omega_{zm})} \\ &\simeq \frac{\omega_u}{s}\frac{1}{1+s\left(\sum_{i=1}^{n}\frac{1}{\omega_{pi}}-\sum_{i=1}^{m}\frac{1}{\omega_{zi}}\right)} \end{aligned}$$

Defining the effective parasitic pole ω_p as

$$\omega_p = 1/\left(\sum_{i=1}^{n}\frac{1}{\omega_{pi}}-\sum_{i=1}^{m}\frac{1}{\omega_{zi}}\right),$$

the integrator transfer function and the OTA transconductance gain can be approximated by

$$\begin{aligned} T_{int}(s) &\simeq \frac{\omega_u}{s}\left(\frac{1}{1+s/\omega_p}\right) \\ g_m(s) &\simeq \frac{g_{mo}}{1+s/\omega_p}, \end{aligned} \tag{2.10}$$

where ω_u is as defined in (2.8).

The effects of the effective parasitic pole on the integrator is the high frequency roll off of the gain response and the excess phase lag at the unity-gain frequency as can be seen from Fig. 2.5(b). This excess phase error causes the Q-enhancement effect as can be seen later. If we assume that all OTAs of the biquad have the same effective parasitic pole, then a more realistic transfer function for the biquad of Fig. 2.3 can be obtained by substituting (2.10) into the ideal model (2.2):

$$T_a(s) \simeq \frac{\tau^2 s^4 + 2\tau s^3 + (1+\tau\frac{g_{m4}-g_{m3}}{C_7})s^2 + \frac{g_{m4}-g_{m3}}{C_7}s + \frac{g_{m1}g_{m3}}{C_6C_7}}{\tau^2 s^4 + 2\tau s^3 + (1+\tau\frac{g_{m5}}{C_7})s^2 + \frac{g_{m5}}{C_7}s + \frac{g_{m2}g_{m3}}{C_6C_7}}, \tag{2.11}$$

where $\tau = 1/\omega_p$, $B_{lp} = 0$, and $B_{hp} = B_{bp} = 1$. A 4th or higher order transfer function is thus more appropriate for describing the biquad instead of the ideal 2nd-order function. This phenomenon is called the over-ordering effect.

The over-ordering effect due to the parasitic poles and zeros can be explained in another way. By assuming $\omega_o\tau << 1$, for medium to high Q biquads, the actual pole frequency and quality factor can be characterized by [68]

$$\begin{aligned} \omega_{o_a} &\simeq \omega_o \\ Q_a &\simeq \frac{Q}{1-2\omega_o\tau Q}. \end{aligned} \tag{2.12}$$

Now, it can be easily seen that the effect of the parasitic poles and zeros on the biquad is the significant Q-enhancement. We define $\omega_o\tau$ as the over-ordering factor. Generally, the over-ordering factor will be very small for low frequency applications because ω_o is much less than

$\omega_p(=1/\tau)$ at low frequencies. In this case the over-ordering effects can be negligible, but at high frequencies the over-ordering effects will be substantial. If the effective parasitic pole is located at 10 times higher frequency than the pole frequency, i.e. the over-ordering factor $\omega_o\tau = 0.1$, and the filter is designed with Q equal to or greater than 5, then it can be seen from (2.12) that the result is an oscillatory circuit. Therefore, predistortion techniques should be adapted to implement high-Q biquads. If we can estimate or guess the effective parasitic pole frequency, then we can use the following predistorted Q value at the initial implementation to prevent the filter from oscillation:

$$Q_{dist} < \frac{Q}{(1+2\omega_o\tau_e Q)},$$

where τ_e is an estimated value. Since it is hard to estimate τ accurately, the rule of thumb is to use an over-predistorted Q value.

The model including the finite output impedance but assuming no parasitics in $g_m(s)$ is shown in Fig. 2.4(c) where R_o and C_o are the output impedance and output capacitance of the OTA and C_L is the integrating capacitance. Assuming $C_L >> C_o$, the integrator transfer function is

$$\begin{aligned} T_{int}(s) &= \frac{g_{mo}R_o}{1+sC_LR_o} \\ &= \frac{\omega_u}{\omega_u/A_o+s} \end{aligned} \tag{2.13}$$

where $A_o = g_{mo}R_o$ is the finite dc gain of the OTA and $\omega_u = g_{mo}/C_L$ is the unity-gain frequency. The effects of the OTA finite dc gain on the integrator is the finite integrator dc gain (ideally infinity) and the phase lead at low frequencies as shown in Fig. 2.5(b).

Using the nonideal integrator equation (2.13), the actual pole frequency and quality factor of the biquad are given by

$$\begin{aligned} \omega_{o_a} &\simeq \omega_o \\ Q_a &\simeq \frac{Q}{1+2Q/A_o} \end{aligned}$$

It can be seen that the phase lead due to the finite dc gain causes Q-degradation and thus partially compensates for the Q-enhancement effect due to the parasitic poles and zeros.

Combining the parasitics and finite dc gain effects, the real integrator can be modeled as Fig. 2.4(d) where the effective parasitic pole ω_p is represented by the internal parasitic capacitance C_p and resistance R_p. The transfer function of the integrator becomes

$$T_{int}(s) \simeq g_{mo}R_o\left(\frac{1}{1+sR_oC_L}\right)\left(\frac{1}{1+sR_pC_p}\right) \tag{2.14}$$

$$= \frac{\omega_u}{(\omega_u/A_o+s)(1+s/\omega_p)} \tag{2.15}$$

Using this, the actual quality factor can be readily approximated by

$$Q_a \simeq \frac{Q}{1+2(1/A_o-\omega_o\tau)Q}$$

and the actual pole frequency by $\omega_{o_a} \simeq \omega$.

The Q-degradation effect due to the finite dc gain is usually negligible unless A_o is too small. Thus, the 4th-order rational function of (2.11) can be used to model the actual over-ordered biquad. However, if we use the 4th-order model for system ID, we will lose the independent or sequential adjustments of the transfer function coefficients because in our target architecture, there is no attempt to tune or cancel the parasitics themselves. A straightforward method is to then identify the over-ordered actual system with a low order model. The system ID methods presented in the next section make it possible to use a low order model for identifying the over-ordered system with good accuracy. We can thus maintain near independence of adjustment of the transfer function coefficients even in the presence of significant parasitics.

2.4 System Identification

The problem of system ID can be referred to as the estimation of the system model parameters by observing the system input and output samples as shown in Fig. 2.6. Linear time-invariant continuous-time system can be modeled as

$$T(s) = \frac{\sum_{i=0}^{m} b_i s^i}{1+\sum_{i=1}^{n} a_i s^i}.$$

where $n \geq m$. The problem is thus to estimate the coefficients a_i and b_i using the sampled input and output data $\{x(k)\}$ and $\{y(k)\}$.

This problem must be solved efficiently and accurately since as indicated in Section 2.2, accurate identification of the continuous-time filter to be tuned should be preceded for filter adjustments. The system model $T(s)$ is the mathematical equation representing the relationship between the input and output at all times. A feasible way to obtain such a model is to apply appropriate inputs to the filter and observe its outputs. The observed (sampled) input and output data are then processed to estimate the model. It is desirable that the order of the system model be the same as that of the actual physical filter. However, as mentioned in Section 2.3, the order of the actual filter can not be detected easily and is usually greater than that of the desired response because of the over-ordering effects due to parasitics. In that case, a lower-order model should be used to estimate the over-ordered system because the model for system ID should have the same order as the desired response for easy adjustment. The

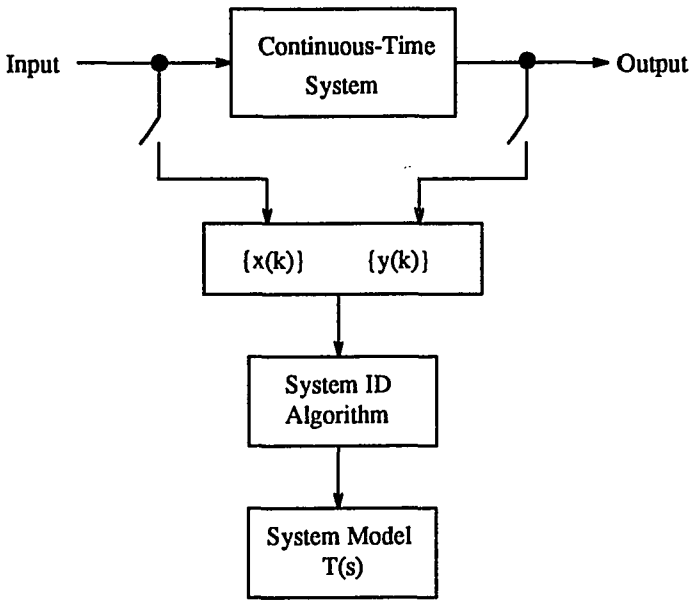


Figure 2.6: Block diagram representing the system ID problem

problem of identifying over-ordered systems with lower-order models is deferred for the present and will be discussed in Section 2.4.5. Until then, it is assumed that the order of the system to be identified is known, and a model which has the same order as the actual system is used for system ID.

There are several ways to get an estimation of $T(s)$ from the input-output samples as shown in Fig. 2.7. The z-domain model $T(z)$ is a discrete-time equivalent to the s-domain model $T(s)$ and is described as

$$T(z) = \frac{\sum_{i=0}^{m_z} b_{zi} z^{-i}}{1 + \sum_{i=1}^{n_z} a_{zi} z^{-i}}.$$

The z-domain system ID (path $< 1 >$) is defined as a problem to estimate the discrete-time model $T(z)$ from the input-output samples $\{x(k)\}$ and $\{y(k)\}$. On the other hand, the s-domain system ID (path $< 4 >$) is defined here as a problem to estimate the continuous-time model $T(s)$ from frequency response data $\{T(j\omega_k)\}$. The z-domain system ID problem has attracted major attention since the estimation of the parameters of a discrete-time model is more straightforward although most systems are of the continuous-time type. From Fig. 2.7 we can find four different methods to get $T(s)$ from $\{x(k)\}$ and $\{y(k)\}$, i.e.,

Method 1 : $< 2 >$

Method 2 : $< 1 > - < 6 >$

Method 3 : $< 3 > - < 4 >$

Method 4 : $< 3 > - < 5 > - < 6 >$

Method 1 can be called the "direct method" and the others the "indirect method." Most direct methods [62, 63, 64] are based upon the differential equations which are the basic models for continuous-time systems. Because of the difficulty to accurately estimate the derivatives from sampled input-output data, the differential equations are transformed into integral equation forms. To calculate multiple integrations from sampled data, prefiltering is performed such as the numerical integration, the bilinear transformation, and orthogonal functions. After prefiltering a discrete-time ID model which contains the continuous-time model parameters can be obtained. Now, various parametric system ID algorithms can be applied to the discrete-time model to estimate the continuous-time model parameters.

In the indirect methods the original problem can be decomposed into a few simpler problems. Method 2 utilizes the input-output samples to first estimate a discrete-time model using z-domain system ID algorithms and then determine an equivalent continuous-time model

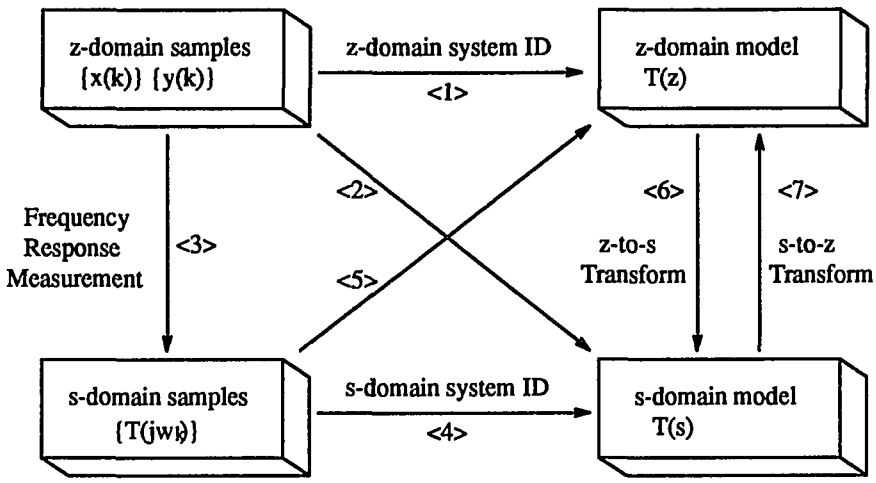


Figure 2.7: Relations among sample data and system models

through z-to-s transformation [58]. In method 3, frequency responses (magnitude and phase) of the continuous-time system are estimated at a set of frequencies from the input-output samples and then the s-domain model is estimated based upon the frequency response data using s-domain system ID algorithms [40, 9]. Another indirect method (method 4) is possible through path $< 3 > - < 5 > - < 6 >$. In path $< 5 >$, the discrete-time model is estimated from frequency response data [57, 56].

The indirect methods are considered the candidates for the continuous-time filter identification to avoid the complicated prefiltering problem required in the direct method. In this section method 2 and 3 will be investigated and method 3 is precluded since it looks somewhat inefficient. Among many well known parameter estimation techniques such as least squares, maximum likelihood, correlation, instrumental variable, and so on, we will primarily utilize the method of least squares because it is conceptually simple and applicable to most practical situations. In fact, the LS method can be applicable to all the paths shown in Fig. 2.7.

2.4.1 z-domain System Identification

In the z-domain system ID, a set of time-domain input and output samples are used to estimate the coefficients of the transfer function $T(z)$ which models the system to be identified.

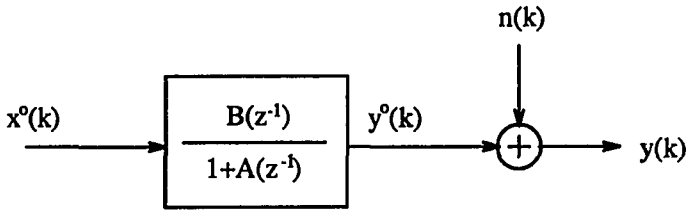


Figure 2.8: Basic system configuration for z-domain system ID

The most popular and basic system configuration for z-domain system ID is shown in Fig. 2.8 [45, 46, 55] where the followings are assumed:

1. All the noise effects corrupting the input and output signals of the system can be lumped into a single additive noise source $n(k)$ at the output.
2. The input $x^o(k)$ can be observed without any noise since it is a specifically designed test signal or a control signal.
3. The noise $n(k)$ is a stationary random process with zero mean and is uncorrelated with both the true system input and output, $x^o(k)$ and $y^o(k)$.

The z-domain model $T(z)$ of the system to be identified can be expressed as

$$T(z) = \frac{Y^o(z)}{X^o(z)} = \frac{B(z^{-1})}{1 + A(z^{-1})} = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + \cdots + b_m z^{-m}}{1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_n z^{-n}}.$$

where $X^o(z)$ and $Y^o(z)$ are the Z-transformations of $x^o(k)$ and $y^o(k)$ respectively. If we define q as the shifting operator such that

$$q^{-i}[x(k)] = x(k-i),$$

the system can be expressed by the difference equation

$$[1 + A(q^{-1})]y^o(k) = B(q^{-1})x^o(k) \tag{2.16}$$

where

$$\begin{aligned} B(q^{-1}) &= b_0 + b_1 q^{-1} + b_2 q^{-2} + \cdots + b_m q^{-m} \\ A(q^{-1}) &= a_1 q^{-1} + a_2 q^{-2} + \cdots + a_n q^{-n}. \end{aligned}$$

The observed output data contaminated by the noise source $n(k)$ is

$$y(k) = y^o(k) + n(k). \tag{2.17}$$

Substituting (2.17) into (2.16), the system equation becomes

$$[1 + A(q^{-1})]y(k) = B(q^{-1})x^o(k) + v(k) \tag{2.18}$$

where

$$v(k) = [1 + A(q^{-1})]n(k). \tag{2.19}$$

The problem of z-domain system ID is now to estimate the parameters a_i and b_i in (2.18) from the observed input-output samples $\{x^o(k)\}$ and $\{y(k)\}$. Many methods are possible to solve the problem. Among them least squares (LS) algorithms [44]-[50] that are the most popular methods will be discussed. The ordinary LS algorithm which usually leads to a biased estimate and the generalized LS (GLS) algorithms where the bias is compensated using noise whitening filters are investigated, and their performances are evaluated through extensive simulations.

2.4.1.1 z-domain Least Squares Algorithm In the z-domain LS method the well-known LS algorithm is directly applied to (2.18). The equation can be rewritten as

$$\begin{aligned} y(k) &= -A(q^{-1})y(k) + B(q^{-1})x^o(k) + v(k) \\ &= \phi_k^T \boldsymbol{\theta} + v(k) \end{aligned}$$

where

$$\begin{aligned} \phi_k &= [-y(k-1), \cdots, -y(k-n), x^o(k), \cdots, x^o(k-m)]^T \\ \boldsymbol{\theta} &= [a_1, \cdots, a_n, b_0, \cdots, b_m]^T. \end{aligned}$$

With a set of N measurements at $k = 1, 2, 3, \ldots, N$, we can have the following matrix equation:

$$\mathbf{y} = \Phi\boldsymbol{\theta} + \mathbf{v} \tag{2.20}$$

where

$$\begin{aligned} \mathbf{y} &= [y(n+1), y(n+2), \cdots, y(N)]^T \\ \mathbf{v} &= [v(n+1), v(n+2), \cdots, v(N)]^T \\ \Phi &= [\phi_{n+1}, \phi_{n+2}, \cdots, \phi_N]^T \end{aligned}$$

or

$$\Phi = \begin{bmatrix} -y(n) & \cdots & -y(1) & x^o(n+1) & \cdots & x^o(n-m+1) \\ -y(n+1) & \cdots & -y(2) & x^o(n+2) & \cdots & x^o(n-m+2) \\ \vdots & & \vdots & \vdots & & \vdots \\ -y(N-1) & \cdots & -y(N-n) & x^o(N) & \cdots & x^o(N-m) \end{bmatrix}.$$

The dimension of matrix Φ is $(N-n)\times(n+m+1)$. The LS estimate of $\boldsymbol{\theta}$ can be obtained by minimizing the sum of squared errors, $\mathbf{v}^T\mathbf{v}$, with respect to $\boldsymbol{\theta}$. By solving

$$\frac{\partial(\mathbf{v}^T\mathbf{v})}{\partial\boldsymbol{\theta}} = 0,$$

the LS estimate is given by

$$\hat{\boldsymbol{\theta}}_{LS} = (\Phi^T\Phi)^{-1}\Phi^T\mathbf{y}. \tag{2.21}$$

From (2.21) and (2.20), the LS estimate can be rewritten by

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{LS} &= \boldsymbol{\theta} + (\Phi^T\Phi)^{-1}\Phi^T\mathbf{v} \\ &= \boldsymbol{\theta} + \Phi^\dagger\mathbf{v} \end{aligned}$$

where $\Phi^\dagger$ is the pseudoinverse of Φ. Thus, the bias of the z-domain LS estimate becomes

$$\begin{aligned} \mathbf{b} &= E[\hat{\boldsymbol{\theta}}_{LS} - \boldsymbol{\theta}] \\ &= E[\hat{\boldsymbol{\theta}}_{LS}] - \boldsymbol{\theta} \\ &= E[\Phi^\dagger\mathbf{v}], \end{aligned}$$

where E is the expectation operator. If Φ and $\mathbf{v}$ are statistically independent and $E[\mathbf{v}] = \mathbf{0}$, then the LS estimate $\hat{\boldsymbol{\theta}}_{LS}$ is consistent and unbiased [45]. However, the residual vector v is usually correlated with the matrix Φ even when the output noise $\{n(k)\}$ is an uncorrelated white noise sequence. The biasedness of the ordinary LS estimation has been explicitly shown in [55] and [46].

2.4.1.2 z-domain Generalized Least Squares Algorithms To compensate for the bias of the LS estimate, it is assumed in z-domain GLS algorithms that the residual $v(k)$ is the output of a linear filter with a white Gaussian noise input $e(k)$. Generally, $v(k)$ can thus be modeled as an autoregressive moving-average (ARMA) process, i.e.,

$$v(k) = \frac{D(q^{-1})}{1 + C(q^{-1})}e(k). \tag{2.22}$$

Since it has been shown in [65, Section 8.6] that the AR model (where $D(q^{-1}) = 1$) is superior to the ARMA model or the MA model (where $C(q^{-1}) = 0$) from the viewpoint of stability and computational efficiency, the AR model based on [51] and a simple modified MA model based on [61] are investigated.

First, if $v(k)$ is modeled as a pth-order AR process

$$[1 + C(q^{-1})]v(k) = e(k) \tag{2.23}$$

where

$$C(q^{-1}) = c_1 q^{-1} + c_2 q^{-2} + \cdots + c_n q^{-p},$$

then by combining (2.23) and (2.19),

$$[1 + A(q^{-1})][1 + C(q^{-1})]n(k) = e(k).$$

The coefficients of $C(q^{-1})$ should be determined such that $[1+A(q^{-1})][1+C(q^{-1})]n(k)$ becomes a white noise sequence with zero mean. Defining the following:

$$\begin{aligned}
\mathbf{c} &= [c_1, c_2, \cdots, c_p]^T \\
\mathbf{e} &= [e(p+1), e(p+2), \cdots, e(N)]^T \\
\mathbf{z}_k &= [-v(k-1), -v(k-2), \cdots, -v(k-p)]^T \\
\mathbf{Z} &= [\mathbf{z}_{p+1}, \mathbf{z}_{p+2}, \cdots, \mathbf{z}_N]^T,
\end{aligned}$$

we have

$$\mathbf{v} = \mathbf{Z}\mathbf{c} + \mathbf{e}.$$

The vector $\mathbf{c}$ can be obtained by a LS estimate

$$\hat{\mathbf{c}} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{v}. \tag{2.24}$$

Combining (2.23) and (2.18), we have the new system equation

$$[1 + A(q^{-1})][1 + C(q^{-1})]y(k) = B(q^{-1})[1 + C(q^{-1})]x^o(k) + e(k). \tag{2.25}$$

By defining

$$\tilde{y}(k) = [1 + C(q^{-1})]y(k) \tag{2.26}$$

$$\tilde{x}^o(k) = [1 + C(q^{-1})]x^o(k), \tag{2.27}$$

(2.25) becomes

$$[1 + A(q^{-1})]\tilde{y}(k) = B(q^{-1})\tilde{x}^o(k) + e(k).$$

We can have a modified matrix equation

$$\tilde{\mathbf{y}} = \tilde{\mathbf{\Phi}}\boldsymbol{\theta} + \mathbf{e}$$

and the new LS estimate

$$\hat{\boldsymbol{\theta}} = (\tilde{\mathbf{\Phi}}^T\tilde{\mathbf{\Phi}})^{-1}\tilde{\mathbf{\Phi}}^T\tilde{\mathbf{y}}. \tag{2.28}$$

Now, it is clear that the new LS estimate of $\boldsymbol{\theta}$ will be consistent because the new system model has an uncorrelated residual error vector $\mathbf{e}$. The standard proofs of consistency for LS estimators rely upon the uncorrelatedness of the residual error vector [53]. In most GLS algorithms the parameters of the system, a_i and b_i, and the parameters of noise model, c_i are alternatively estimated in an iterative process. The iterative procedure is as follows:

Step 1 Set $C(q^{-1}) = 1$.

Step 2 Form $\tilde{y}$ and $\tilde{x}^o$ using equation (2.26) and (2.27).

Step 3 Obtain the LS estimate using equation (2.28).

Step 4 Stop if converged.

Step 5 With $A(q^{-1})$ and $B(q^{-1})$ estimated, compute the residual $\{v(k)\}$ using equation (2.18) and then estimate $C(q^{-1})$ using equation (2.24). Go to Step 2.

This GLS algorithm based on the pth-order AR model will be denoted as GLS(ARp).

In the second method $v(k)$ is modeled as a modified MA process

$$v(k) = [1 + A'(q^{-1})]e(k) \tag{2.29}$$

where $A'(q^{-1})$ is the previous estimate of $A(q^{-1})$ and thus the coefficients of $A'(q^{-1})$ are simply obtained from the previous estimate $\hat{\boldsymbol{\theta}}$. From equation (2.19) and (2.29)

$$n(k) = \frac{1 + A'(q^{-1})}{1 + A(q^{-1})}e(k).$$

If this algorithm converges, then $A'(q^{-1})$ converges to $A(q^{-1})$. Thus, the algorithm will work very well when $\{n(k)\}$ is a white noise sequence. This GLS algorithm based on the modified MA model will be denoted as GLS(MAp), where 'p' does not represent the order of the noise model as in GLS(ARp) but implies that the model parameters are obtained from the *previous* estimates. The iterative procedure for GLS(MAp) can be readily formulated using the similar method as that for GLS(ARp).

2.4.1.3 Simulation Results The z-domain LS algorithm and GLS algorithms are compared through computer simulations. First, the characteristics of the noise associated with the system and the measurement process should be assumed for simulations. We will consider three cases as shown in Fig. 2.9.

In Model 1, output samples are corrupted with the additive correlated noise which is the output of a noise shaping filter when the input is a white Gaussian noise $e(k)$. In Model 2, output samples are corrupted directly with a white Gaussian noise $e(k)$, which is one commonly encountered case in system ID. In Model 1 and 2, only output samples are noisy and the input samples are assumed to be observed without any noise. However, there are many practical situations where the input as well as the output can not be observed without noise. These cases are simulated with Model 3 where the input and output samples are contaminated with white Gaussian noise $m(k)$ and $n(k)$, respectively.

A second-order system of which the transfer function is

$$T(z) = \frac{1.0}{1.0 - 0.5z^{-1} + 0.5z^{-2}}$$

is chosen for simulations. The ideal z-domain parameter polynomials and the ideal parameter vector are thus

$$\begin{aligned} A(z^{-1}) &= -0.5z^{-1} + 0.5z^{-2} \\ B(z^{-1}) &= 1.0 \\ \boldsymbol{\theta} &= [-0.5, 0.5, 1.0]^T \end{aligned}$$

The input $\{x^o(k)\}$ is a sequence of independent Gaussian random variables with unity variance ($\sigma_s^2 = 1.0$) and zero mean. A noise z-domain parameter polynomial $C(z^{-1}) = 0.7z^{-1}$ is chosen for Model 1. The zero mean independent Gaussian random noise $e(k)$, $m(k)$, and $n(k)$ have the same variance σ_n^2.

From 300 input-output samples (data length $N = 300$), the parameters of the system were estimated using three different algorithms for three different noise models at various signal to noise ratios (SNR). The SNR in dB is defined as $10log(\sigma_s^2/\sigma_n^2)$. The simulated sample mean of squared estimation errors ($E_s[\boldsymbol{\theta}_e^T\boldsymbol{\theta}_e]$) are shown in Fig. 2.10 (a), (b), and (c) for noise model 1, 2, and 3, respectively. The sample mean operator E_s is defined by

$$E_s[X] = \frac{1}{N_s}(X_1 + X_2 + \ldots + X_{N_s}),$$

where N_s is the sample size. The estimation error vector $\boldsymbol{\theta}_e$ is

$$\boldsymbol{\theta}_e = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}.$$

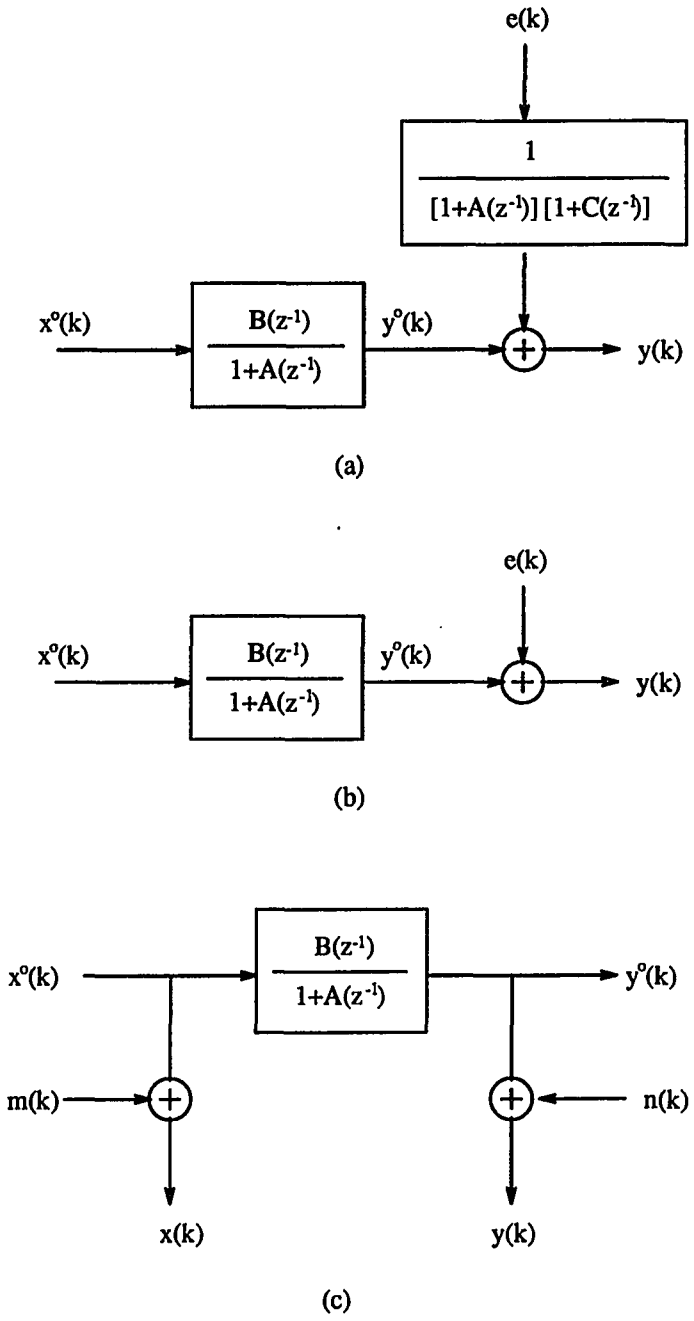


Figure 2.9: System configuration for z-domain system ID with noise model (a) Model 1 (b) Model 2 (c) Model 3

The iteration limit for GLS methods was set to 10. Every result was computed from 100 independent simulations, which means that the sample size is 100.

In most cases GLS algorithms show better results than the ordinary LS algorithm, which implies that the bias is reduced by using GLS algorithms. For noise model 1, the GLS algorithm using first-order AR model, GLS(AR1), gives the best results since the real output noise was assumed in noise model 1 such that the residual $v(k)$ of the system equation became an AR process. As expected, the GLS algorithm based on the modified MA model, GLS(MAp), exhibits the best results for noise model 2. On the other hand, the bias reduction by GLS algorithms is not good for noise model 3 compared to that for noise model 1 or 2. The GLS(AR1) still shows better performance for all simulated noise levels than the LS algorithm. The convergence rate of z-domain GLS algorithms are shown in Fig. 2.11. The iteration number 0 corresponds to the results of the LS algorithm. For most cases, the GLS algorithms converge within 5 iterations.

So far, we have investigated only the GLS algorithms using AR and modified MA models. However, the residual can be modeled as a pure MA model or an ARMA model as shown in (2.22). To compare them, the GLS algorithms using a second-order MA model GLS(MA2) and using a first-order ARMA model GLS(ARMA1) were simulated and their results are depicted in Fig. 2.12. The methods to estimate the parameters of the MA and ARMA models can be found in [65, 66, 45]. The simulation results indicate that the GLS algorithms show similar performance and that they give better results than the LS algorithm.

2.4.2 s-domain System Identification

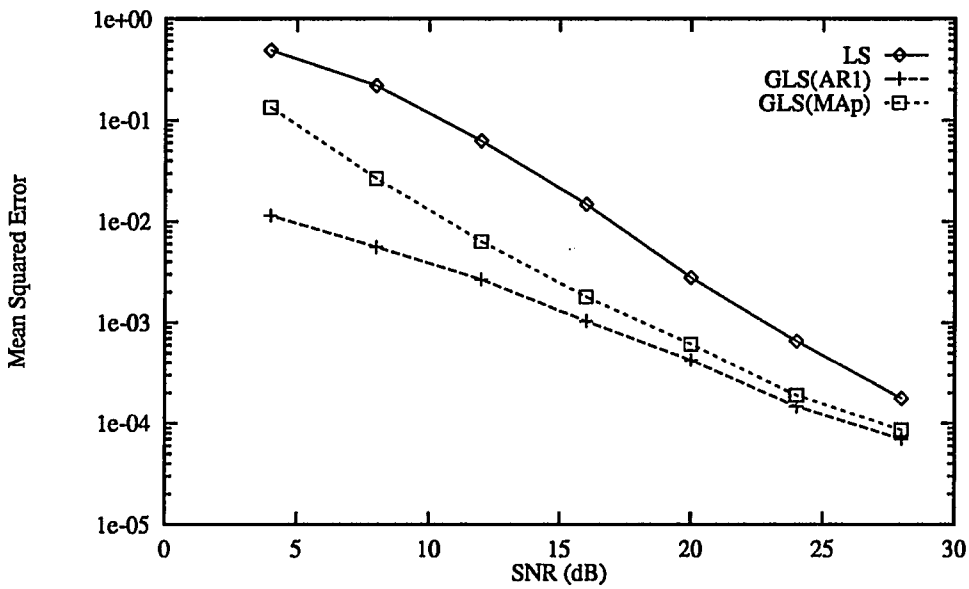
As shown in Fig. 2.7, the s-domain system ID uses a set of frequency response data (gain and phase responses) to estimate the coefficients of a model $T(s)$. In this section s-domain least squares algorithms are investigated as frequency-domain parametric continuous-time system ID methods.

Linear time-invariant continuous-time systems can be expressed as ($n \geq m$)

$$T(s) = \frac{B(s)}{1 + A(s)} = \frac{b_0 + b_1 s + b_2 s^2 + \cdots + b_m s^m}{1 + a_1 s + a_2 s^2 + \cdots + a_n s^n}.$$

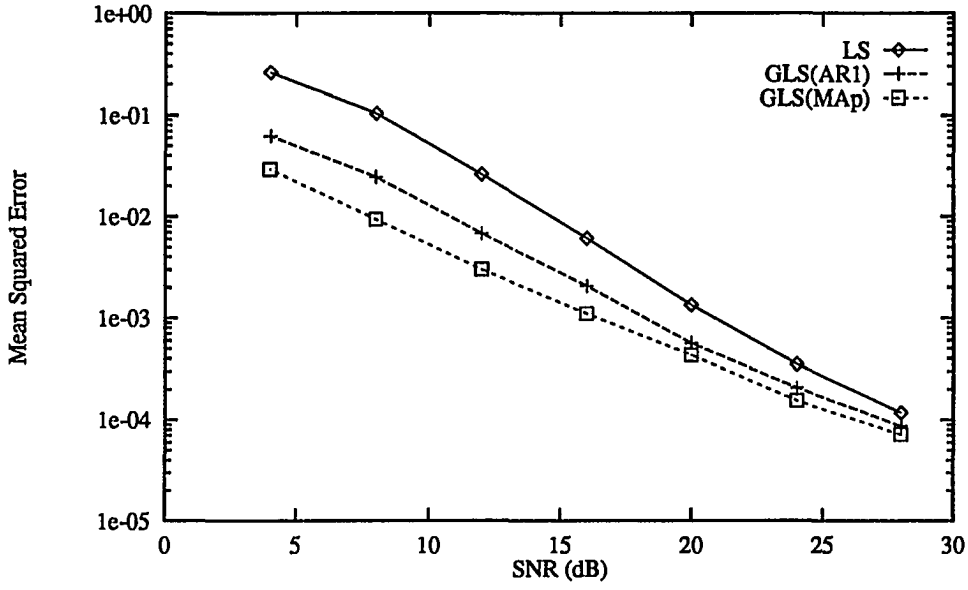
In many cases, frequency response data can readily be obtained or measured, but some measurement errors are unavoidably involved which is shown in Fig. 2.13. By using $s = j\omega$, the ideal system equation can be expressed in terms of $j\omega$.

$$[1 + A(j\omega)]T(j\omega) = B(j\omega) \tag{2.30}$$

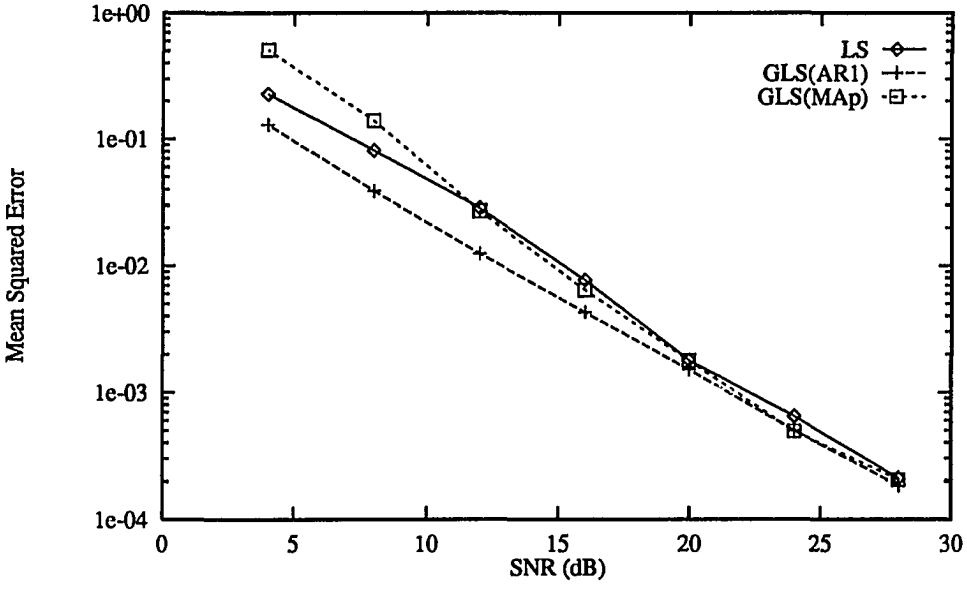


(a)

Figure 2.10: z-domain system ID results. Sample mean of the squared parameter estimation error ($E_s[\theta_e^T\theta_e]$) versus input signal to noise ratios in dB computed from 100 independent simulations (Data length N=300) with actual system noise models (a) Model 1



(b)



(c)

Figure 2.10: (continued) (b) Model 2 (c) Model 3

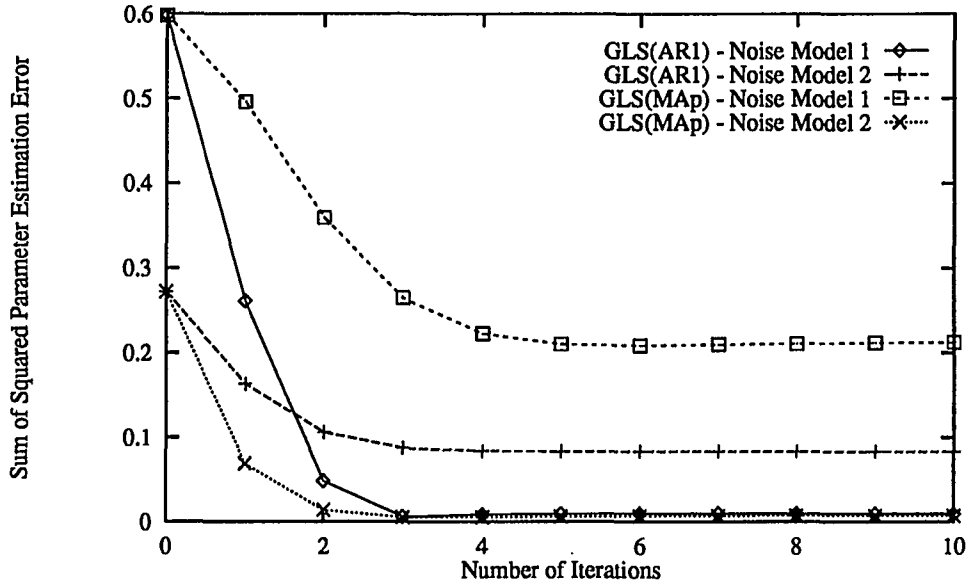


Figure 2.11: Rate of convergence of z-domain system ID algorithms (y-axis: $\theta_e^T\theta_e$, input signal to noise ratio SNR=3.0 (dB), data length $N = 300$)

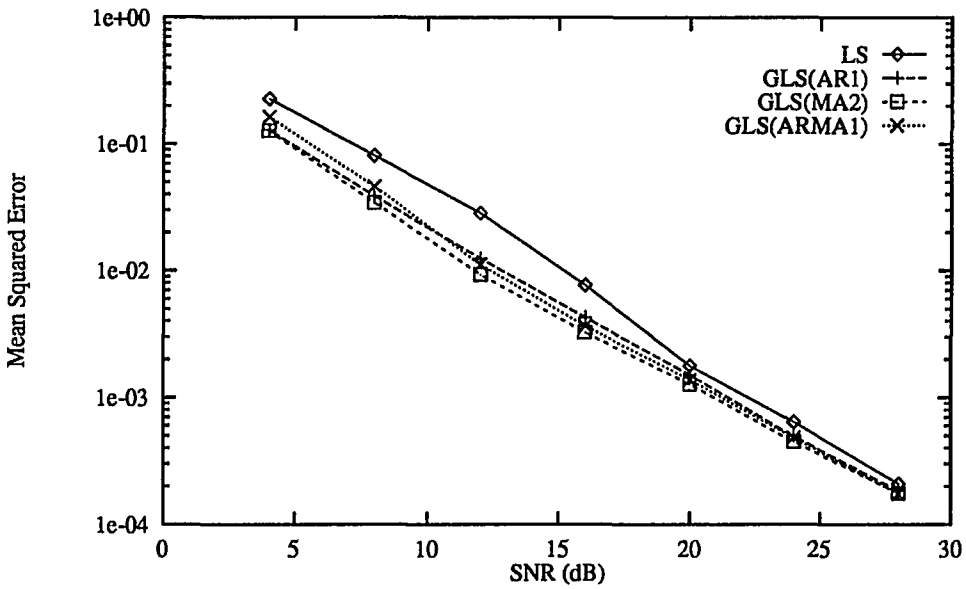


Figure 2.12: Performance comparison of z-domain LS algorithms with an actual system noise model of Model 3. The sample mean of the squared estimation error ($E_s[\theta_e^T\theta_e]$) were computed from 100 independent simulations with data length $N = 300$

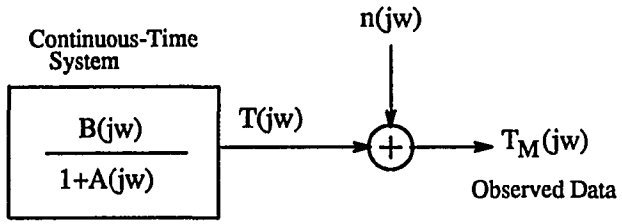


Figure 2.13: Basic system configuration for s-domain system ID

where

$$\begin{aligned} A(j\omega) &= a_1(j\omega) + a_2(j\omega)^2 + \cdots + a_n(j\omega)^n \\ B(j\omega) &= b_0 + b_1(j\omega) + b_2(j\omega)^2 + \cdots + b_m(j\omega)^m. \end{aligned}$$

The observed data $T_M(j\omega)$ which can be obtained from measured gain and phase responses are contaminated by $n(j\omega)$.

$$T_M(j\omega) = T(j\omega) + n(j\omega). \tag{2.31}$$

It is assumed that $n(j\omega)$ is an independent zero-mean additive complex random noise. Substituting (2.31) into (2.30), the system equation becomes

$$[1 + A(j\omega)]T_M(j\omega) = B(j\omega) + v(j\omega) \tag{2.32}$$

where

$$v(j\omega) = [1 + A(j\omega)]n(j\omega). \tag{2.33}$$

The problem of s-domain system ID is now to estimate the parameters a_i and b_i in (2.32) from the observed data $T_M(j\omega)$. Ordinary least squares (LS) algorithms and generalized least squares (GLS) algorithms have been used for the z-domain system ID in the previous section. As s-domain counterparts, the algorithms are applied for the above s-domain system ID problem. These s-domain LS and GLS algorithms are comparatively investigated based on the robustness and efficiency.

2.4.2.1 s-domain Least Squares Algorithm At $\omega = \omega_k$, equation (2.32) can be rewritten as

$$T_M(j\omega_k) = -A(j\omega_k)T_M(j\omega_k) + B(j\omega_k) + v(j\omega_k)$$

$$= \phi_k^T \boldsymbol{\theta} + v(j\omega_k)$$

where

$$\begin{aligned} \phi_k &= [-(j\omega_k)T_M(j\omega_k), \cdots, -(j\omega_k)^n T_M(j\omega_k), 1, (j\omega_k), \cdots, (j\omega_k)^m]^T \\ \boldsymbol{\theta} &= [a_1, \cdots, a_n, b_0, \cdots, b_m]^T. \end{aligned}$$

With a set of N measurements at ω_k, $k = 1, 2, 3, \ldots, N$, we can have the following matrix equation:

$$\mathbf{y} = \boldsymbol{\Phi}\boldsymbol{\theta} + \mathbf{v} \tag{2.34}$$

where

$$\begin{aligned} \mathbf{y} &= [T_M(j\omega_1), T_M(j\omega_2), \cdots, T_M(j\omega_N)]^T \\ \mathbf{v} &= [v(j\omega_1), v(j\omega_2), \cdots, v(j\omega_N)]^T \\ \boldsymbol{\Phi} &= [\phi_1, \phi_2, \cdots, \phi_N]^T \end{aligned}$$

or

$$\boldsymbol{\Phi} = \begin{bmatrix} -(j\omega_1)T_M(j\omega_1) & \cdots & -(j\omega_1)^n T_M(j\omega_1) & 1 & (j\omega_1) & \cdots & (j\omega_1)^m \\ -(j\omega_2)T_M(j\omega_2) & \cdots & -(j\omega_2)^n T_M(j\omega_2) & 1 & (j\omega_2) & \cdots & (j\omega_2)^m \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ -(j\omega_N)T_M(j\omega_N) & \cdots & -(j\omega_N)^n T_M(j\omega_N) & 1 & (j\omega_N) & \cdots & (j\omega_N)^m \end{bmatrix}.$$

The LS estimate of $\boldsymbol{\theta}$ can be obtained by minimizing the sum of squared complex errors, $\mathbf{v}^*\mathbf{v}$ (* denotes the complex conjugate transpose), with respect to $\boldsymbol{\theta}$. Thus, one substantial difference between the s-domain and z-domain LS equations is that the cost function of the s-domain LS estimation is in the frequency domain while the cost function of the z-domain LS estimation is in the time domain. By solving

$$\frac{\partial(\mathbf{v}^*\mathbf{v})}{\partial\boldsymbol{\theta}} = 0$$

the LS estimate is given by

$$\hat{\boldsymbol{\theta}}_{LS} = [Re(\boldsymbol{\Phi}^*\boldsymbol{\Phi})]^{-1} Re[\boldsymbol{\Phi}^*\mathbf{y}]. \tag{2.35}$$

From (2.34) and (2.35), the LS estimate can be rewritten by

$$\hat{\boldsymbol{\theta}}_{LS} = \boldsymbol{\theta} + [Re(\boldsymbol{\Phi}^*\boldsymbol{\Phi})]^{-1} Re[\boldsymbol{\Phi}^*\mathbf{v}].$$

Thus, the bias of the s-domain LS estimate becomes

$$E(\hat{\boldsymbol{\theta}}_{LS} - \boldsymbol{\theta}) = E\{[Re(\boldsymbol{\Phi}^*\boldsymbol{\Phi})]^{-1} Re[\boldsymbol{\Phi}^*\mathbf{v}]\}.$$

It can be seen that the ordinary LS method gives a biased estimate since $\boldsymbol{\Phi}$ and $\mathbf{v}$ are not statistically independent. It should be noted that the matrix $\boldsymbol{\Phi}$ also contains observed noisy data, and $\boldsymbol{\Phi}^*\boldsymbol{\Phi}$ is quadratic in the data and hence $\hat{\boldsymbol{\theta}}_{LS}$ is nonlinear in the data. This s-domain ordinary LS algorithm is often called Levy's method [52]. To reduce the bias introduced in the LS method, s-domain GLS algorithms are presented in the next section.

2.4.2.2 s-domain Generalized Least Squares Algorithms To compensate for the bias of the LS estimate, it is assumed in s-domain GLS algorithms that there possibly exists a $H(j\omega)$ satisfying

$$H(j\omega)v(j\omega) = e(j\omega). \tag{2.36}$$

such that $\{e(j\omega)\}$ is a white complex noise sequence with zero mean. Combining (2.36) and (2.33),

$$H(j\omega)[1 + A(j\omega)]n(j\omega) = e(j\omega).$$

If an appropriate way to obtain the $H(j\omega)$ can be found, then the system equation (2.32) can be modified as

$$H(j\omega)[1 + A(j\omega)]T_M(j\omega) = H(j\omega)B(j\omega) + e(j\omega).$$

At $\omega = \omega_k$

$$H(j\omega_k)T_M(j\omega_k) = -H(j\omega_k)A(j\omega_k)T_M(j\omega_k) + H(j\omega_k)B(j\omega_k) + e(j\omega_k).$$

Defining the follows:

$$\begin{aligned}
\mathbf{e} &= [e(j\omega_1), e(j\omega_2), \cdots, e(j\omega_N)]^T \\
\tilde{\mathbf{y}} &= [T_M(j\omega_1)H(j\omega_1), \cdots, T_M(j\omega_N)H(j\omega_N)]^T && (2.37) \\
\tilde{\boldsymbol{\phi}}_k &= H(j\omega_k)[-(j\omega_k)T_M(j\omega_k), \cdots, -(j\omega_k)^n T_M(j\omega_k), 1, (j\omega_k), \cdots, (j\omega_k)^m]^T && (2.38) \\
\tilde{\boldsymbol{\Phi}} &= [\tilde{\boldsymbol{\phi}}_1, \tilde{\boldsymbol{\phi}}_2, \cdots, \tilde{\boldsymbol{\phi}}_N]^T, && (2.39)
\end{aligned}$$

we have a modified matrix equation

$$\tilde{\mathbf{y}} = \tilde{\boldsymbol{\Phi}}\boldsymbol{\theta} + \mathbf{e}$$

and the new LS estimate is

$$\hat{\boldsymbol{\theta}} = [Re(\tilde{\boldsymbol{\Phi}}^*\tilde{\boldsymbol{\Phi}})]^{-1} Re[\tilde{\boldsymbol{\Phi}}^*\tilde{\mathbf{y}}].$$

Now, it is clear that the new LS estimate of $\boldsymbol{\theta}$ will be consistent if the equation (2.36) is satisfied because the new system model has an uncorrelated residual error vector $\mathbf{e}$.

The remaining problem is how to determine $H(j\omega)$. Two methods will be presented. One is called s-domain GLS since it is similar with the s-domain GLS method using an AR residual model. The other can be referred to as the s-domain counterpart of the z-domain GLS algorithm, GLS(MAp), and will be called s-domain modified GLS (simply M-GLS). This modified GLS algorithm is actually equivalent to the iterative complex LS algorithm presented in [9].

In the s-domain GLS algorithm, the $H(j\omega)$ is modeled as

$$H(j\omega) = 1 + C(j\omega)$$

and thus

$$[1 + C(j\omega)]v(j\omega) = e(j\omega), \tag{2.40}$$

where

$$C(j\omega) = c_1(j\omega) + c_2(j\omega)^2 + \cdots + c_n(j\omega)^p.$$

It is assumed that the system is stable such that the roots of $[1 + C(j\omega)]$ lie left the $j\omega$ axis. The unknown parameters c_i are iteratively estimated by the ordinary LS method. Equation (2.40) is rewritten at $\omega = \omega_k$ as

$$v(j\omega_k) = -C(j\omega_k)v(j\omega_k) + e(j\omega_k)$$

Defining the following:

$$\begin{aligned}
\mathbf{c} &= [c_1, c_2, \cdots, c_p]^T \\
\mathbf{e} &= [e(j\omega_1), e(j\omega_2), \cdots, e(j\omega_N)]^T \\
\mathbf{z}_k &= [-(j\omega_k)v(j\omega_k), \cdots, -(j\omega_k)^p v(j\omega_k)]^T \\
\mathbf{Z} &= [\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_N]^T,
\end{aligned}$$

we have

$$\mathbf{v} = \mathbf{Z}\mathbf{c} + \mathbf{e},$$

and the LS estimate of the parameter c_i is given by

$$\hat{\mathbf{c}} = [Re(\mathbf{Z}^*\mathbf{Z})]^{-1} Re[\mathbf{Z}^*\mathbf{v}].$$

In the second GLS algorithm (M-GLS), the $H(j\omega)$ is modeled as

$$H(j\omega) = 1/[1 + A'(j\omega)],$$

where the coefficients of $A'(j\omega)$ are simply replaced by those obtained from the previous estimate $\hat{\boldsymbol{\theta}}$. Now we have

$$\begin{aligned} v(j\omega) &= [1 + A'(j\omega)]e(j\omega) \\ e(j\omega) &= \frac{1 + A(j\omega)}{1 + A'(j\omega)} n(j\omega). \end{aligned}$$

If this algorithm converges, then $A'(j\omega)$ converges to $A(j\omega)$ and thus, sequence $\{e(j\omega)\}$ approaches $\{n(j\omega)\}$. This algorithm is very simple and efficient since the computational requirement is much less than that of the GLS algorithm where two LS estimations are needed every iteration.

The iterative procedure for the s-domain GLS and M-GLS methods are as follows:

Step 1 Set $H(j\omega_k) = 1$ for $i = 1, 2, \ldots, N$.

Step 2 Form $\tilde{\boldsymbol{\Phi}}$ and $\tilde{\mathbf{y}}$ according to equation (2.37)-(2.39).

Step 3 Obtain the LS estimate

$$\hat{\boldsymbol{\theta}} = [Re(\tilde{\boldsymbol{\Phi}}^* \tilde{\boldsymbol{\Phi}})]^{-1} Re[\tilde{\boldsymbol{\Phi}}^* \tilde{\mathbf{y}}].$$

Step 4 Obtain $H(j\omega)$

GLS Generate $v(j\omega)$ from (2.32) using the previous estimate $\hat{\boldsymbol{\theta}}$, and compute the LS estimate of $\mathbf{c}$

$$\hat{\mathbf{c}} = [Re(\mathbf{Z}^* \mathbf{Z})]^{-1} Re[\mathbf{Z}^* \mathbf{v}]$$

and then compute $H(j\omega_k)$ for $k = 1, 2, \ldots, N$ using

$$H(j\omega_k) = 1 + C(j\omega_k).$$

M-GLS Obtain coefficients a_i from the previous estimate $\hat{\boldsymbol{\theta}}$ and compute $A'(j\omega_k)$, and then compute $H(j\omega_k)$ for $k = 1, 2, \ldots, N$ using

$$H(j\omega_k) = 1/[1 + A'(j\omega_k)].$$

Step 5 Go to Step 2 and repeat until convergence is obtained.

2.4.2.3 Simulation Results In this section the s-domain LS, GLS, and M-GLS algorithms are compared through extensive simulations based upon the bias and consistency of their estimates. The test system is a second-order lowpass notch filter which has the following ideal transfer function:

$$T(s) = \frac{0.419080 + 0.176000s^2}{1.0 + 0.101413s + 1.005084s^2}.$$

The ideal parameter vector is thus

$$\boldsymbol{\theta} = \begin{bmatrix} a_1 \\ a_2 \\ b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 0.101413 \\ 1.005084 \\ 0.419080 \\ 0 \\ 0.176000 \end{bmatrix}. \tag{2.41}$$

The noisy observed data $T_M(j\omega_k)$ are generated from the ideal frequency response data $T(j\omega_k)$ which are added by a random complex noise $n(j\omega_k)$ where $Re[n]$ and $Im[n]$ are uncorrelated and uniformly distributed with zero mean and variance σ_n^2. The frequency response data are sampled at equally spaced normalized angular frequencies from 0 to 2.

The solutions of the s-domain LS problems as well as the z-domain LS problems of the previous section are computed using the singular value decomposition (SVD) method. The SVD method fixes the roundoff problem from which other direct methods using LU decomposition and Gauss-Jordan elimination usually suffer. Moreover, the SVD method can also cure the ill-condition problems of the matrix to be inverted [54].

One typical simulation results with $\sigma_n = 0.1732$ are given in Fig. 2.14 where the convergence rate of GLS and M-GLS algorithms at two different data sizes are shown. The iteration number 0 corresponds to the results of the LS algorithm. The y-axis is the sum of squared estimation error, i.e., $\boldsymbol{\theta}_e^T\boldsymbol{\theta}_e$. The estimation error vector $\boldsymbol{\theta}_e$ is defined as before, i.e., $\boldsymbol{\theta}_e = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$. Both the convergence rate of GLS and M-GLS are very fast and converge after two iterations which is faster than that of z-domain GLS algorithms. It can also be clearly seen that the GLS and M-GLS algorithms give much better results than those of the LS algorithm. In this simulation very significant noisy data ($\sigma_n = 0.1732$) are used to more clearly differentiate the results of the three algorithms. The significantly contaminated frequency response data and the ideal gain response of the notch filter are shown in Fig. 2.15. The identified gain responses are shown in Fig. 2.16 with the actual response for comparison. Fig. 2.16(b) is a magnified one of Fig. 2.16(a). Even with a very high level of noise, the GLS and M-GLS algorithms give good

results while the LS algorithm results in a poor estimation.

In order to investigate the consistency of those algorithms, the sample mean of the squared estimation error $E_s[\boldsymbol{\theta}_e^T\boldsymbol{\theta}_e]$ was computed from 100 independent simulations (sample size N_s = 100) at different data length from $N = 50$ to 500. In this simulation the standard deviation of the noise was set to 0.029, and the iteration limit for the GLS and M-GLS algorithm was set to 5. The results are shown in Fig. 2.17. An estimate $\hat{\boldsymbol{\theta}}$ of the parameter vector $\boldsymbol{\theta}$ is *consistent* [53],[45] if in the long run, the error vector $\boldsymbol{\theta}_e$ approaches the zero vector or alternatively the mean squared error $E[\boldsymbol{\theta}_e^T\boldsymbol{\theta}_e]$ approaches zero, i.e.,

$$\lim_{N\to\infty} E[\boldsymbol{\theta}_e] = \mathbf{0}.$$

or

$$\lim_{N\to\infty} E[\boldsymbol{\theta}_e^T\boldsymbol{\theta}_e] = 0,$$

where N is the data length. An estimate $\hat{\boldsymbol{\theta}}$ is said to be *unbiased* if the expected value of the estimation error vector $\hat{\boldsymbol{\theta}}_e$ is zero for all N [53],[45], i.e.,

$$E[\boldsymbol{\theta}_e] = \mathbf{0} \qquad \text{for all } N \geq n + m + 1$$

The plots in Fig. 2.17 show that the sample mean squared errors of the GLS and M-GLS estimates decrease much faster than than that of the LS estimate as the data length increases up to 500.

To compare their estimation accuracy and the level of bias, the sample mean and variance of parameter estimation errors have been computed from 500 independent simulations. The absolute values of the sample means of 500 parameter estimation errors, $|E_s[\theta_{ei}]|$, are shown in Fig. 2.18(a), where θ_{ei}'s are the individual elements of the error vector $\boldsymbol{\theta}_e$. Their variances calculated using $(E_s[\theta_{ei}^2] - E_s^2[\theta_{ei}])$ are also shown in Fig. 2.18(b). Although the individual parameters (a_1, a_2, b_0, b_1, b_2) may have different units and thus their means and variances can not be compared together, they were plotted with the same y-axis for convenience. In this simulation the following conditions were used: noise standard deviation $\sigma_n = 0.029$, data length $N = 50$, and iteration limit=5. It can be seen from Fig. 2.18(b) that the estimation accuracy is the best in the M-GLS algorithm. These simulation results indicate that among the three s-domain system ID algorithms, the M-GLS algorithm is the most robust.

2.4.3 s-to-z and z-to-s Transformation

The s-to-z transformation has been one of important parts in digital filter and digital control system design where the well-established information of continuous-time filters and

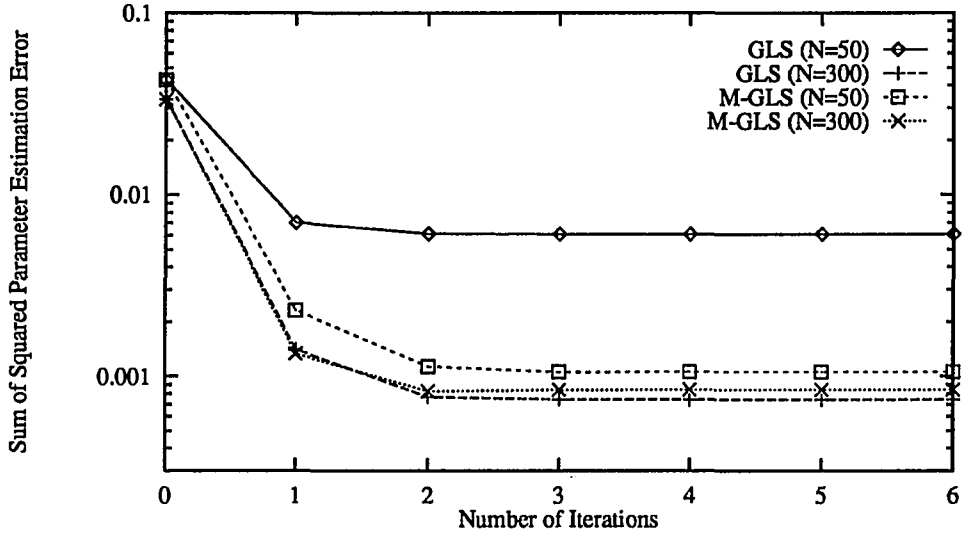


Figure 2.14: Rate of convergence of s-domain system ID algorithms, GLS and M-GLS (y-axis: $\theta_e^T\theta_e$, σ_n=0.1732)

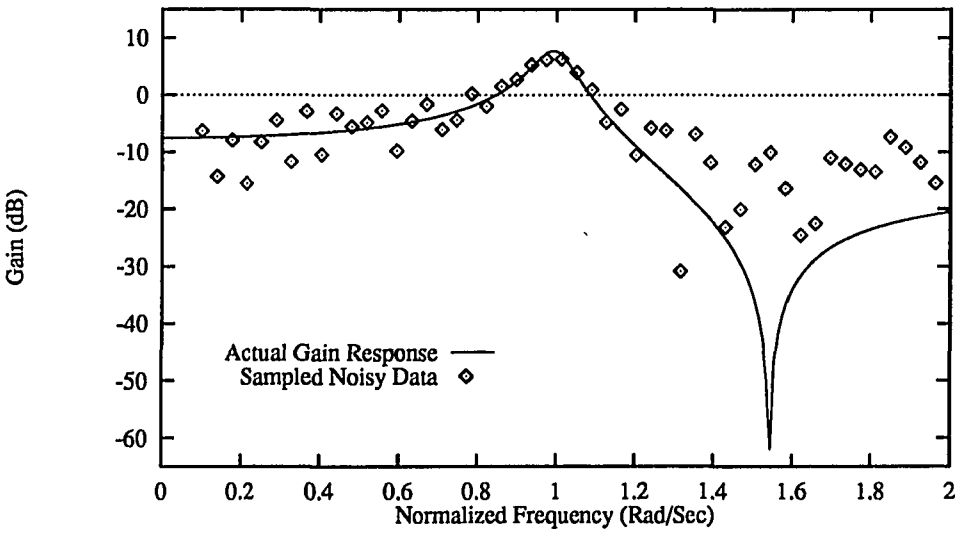
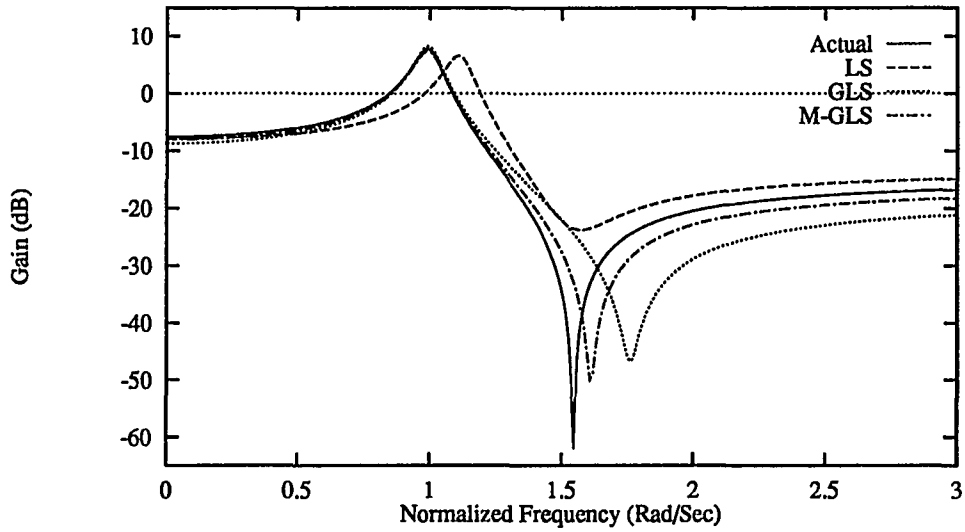
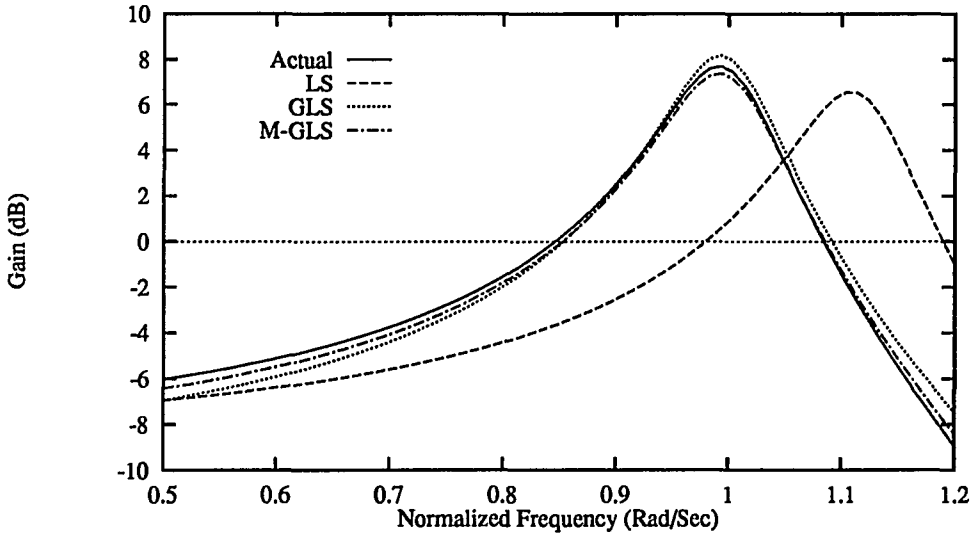


Figure 2.15: Sampled noisy gain response data (σ_n=0.1732) and the ideal gain response of the second-order lowpass notch filter



(a)



(b)

Figure 2.16: (a) The gain responses identified by the s-domain system ID algorithms (b) Magnified view of (a) (σ_n=0.1732, data length N=50, Iteration limit=10)

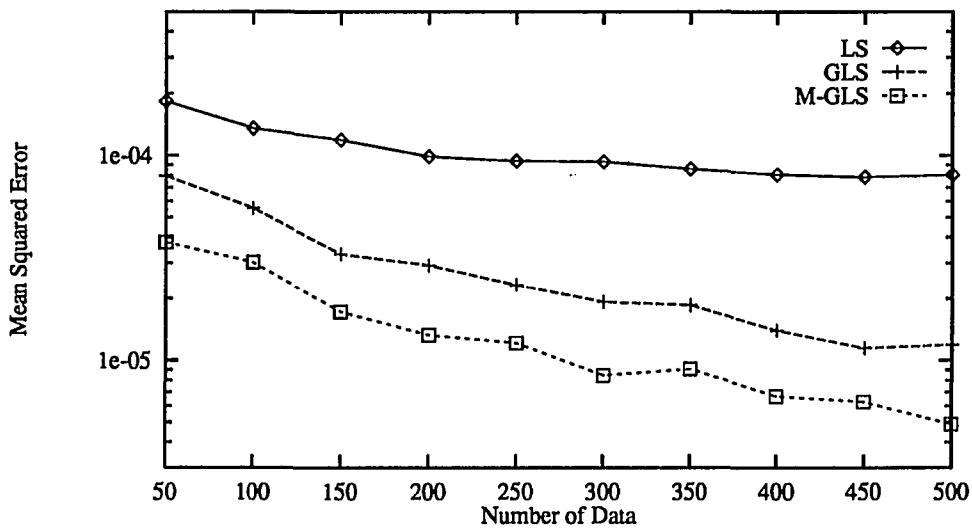
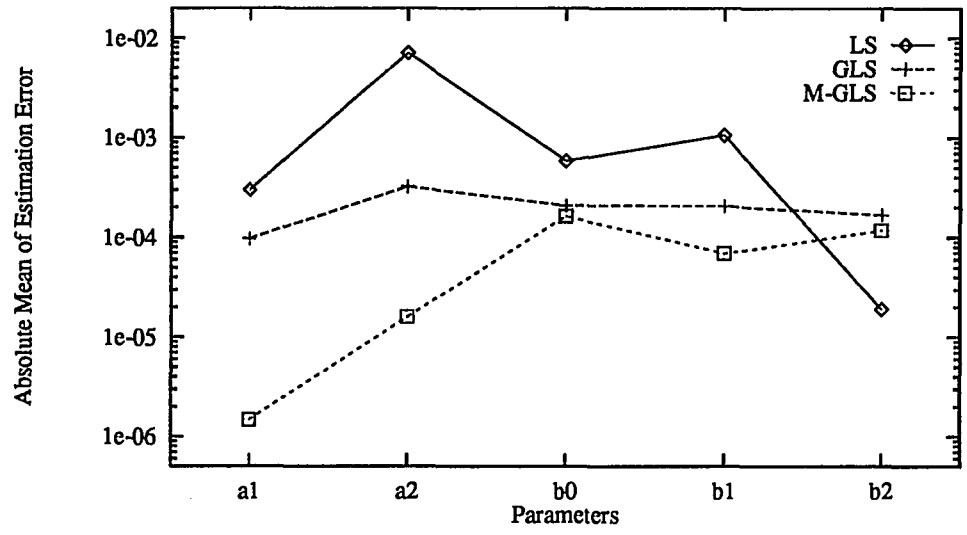
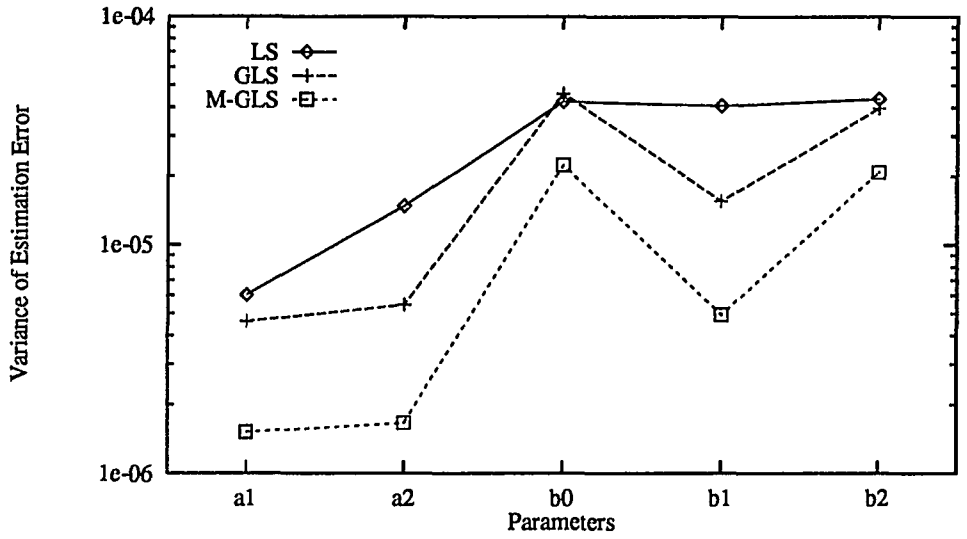


Figure 2.17: Sample mean of the squared parameter estimation error ($E_s[\theta_e^T \theta_e]$) versus number of data computed from 100 independent simulations (σ_n=0.029, Iteration limit=5)



(a)



(b)

Figure 2.18: (a) Absolute value of the sample mean of the parameter estimation errors ($|E_s[\theta_{ei}]|$) (b) Sample variance of the estimation errors ($E_s[\theta_{ei}^2] - E_s^2[\theta_{ei}]$), computed from 500 independent simulations (σ_n=0.029, data length N=50, Iteration limit=5)

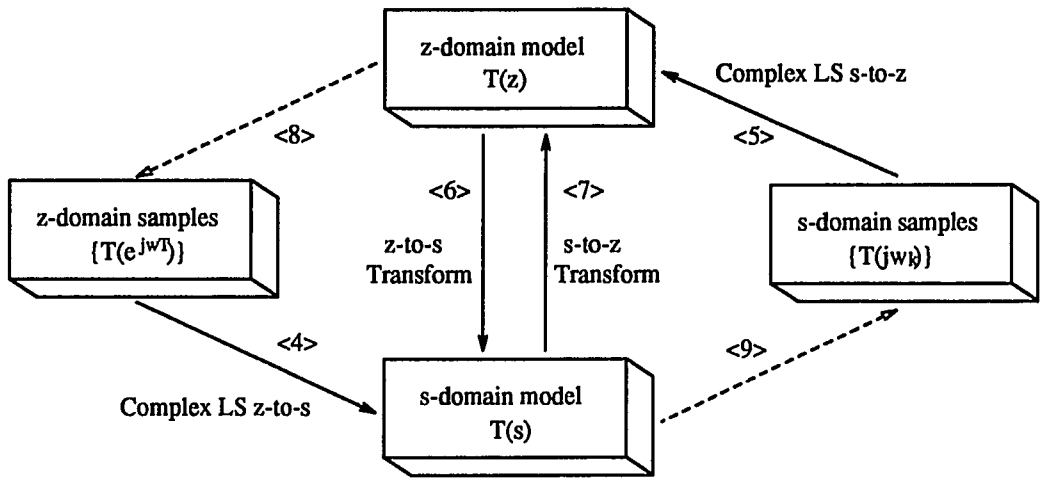


Figure 2.19: z-to-s and s-to-z transformation methods

continuous control systems is utilized. The z-to-s transformation is also required for many areas, especially for indirect continuous-time system identification (see Fig. 2.7). Many methods such as numerical integration, pole-zero mapping, and hold equivalents, have been reported in the literature [53, 60, 59] for both transformations.

In this section a new approach to the transformation problem is presented. As depicted in Fig. 2.19, complex least squares algorithms can be applied to solve the problems. The s-to-z transformation $< 7 >$ can be decomposed into two steps through path $< 9 > - < 5 >$ where the frequency samples $\{T(j\omega_k)\}$ are directly calculated from the known s-domain transfer function $T(s)$, and then the LS algorithm is used to estimate the parameters of the equivalent z-domain model from the frequency samples. This approach will be called a complex LS s-to-z transformation method. The z-to-s transformation $< 6 >$ can also be decomposed into $< 8 > - < 4 >$ where the frequency samples $\{T(e^{j\omega_k T})\}$ can be obtained from the known z-domain transfer function $T(z)$. These frequency samples can be fed into the LS algorithm to estimate the parameters of a s-domain model such that the estimated s-domain model can approximate the z-domain transfer function as well as possible. This approach will be called a complex LS z-to-s transformation method of which the algorithm is actually equivalent to the s-domain system ID algorithm presented in Section 2.4.2.

2.4.3.1 Complex LS s-to-z Transformation Method The problem of s-to-z transformation is that given a continuous-time transfer function $T(s)$, determine the discrete-time transfer function $T(z^{-1})$,

$$T(z^{-1}) = \frac{B(z^{-1})}{1 + A(z^{-1})} \tag{2.42}$$

such that they have approximately the same characteristics. Equation (2.42) can be rewritten as

$$[1 + A(e^{-j\omega T})]T(e^{-j\omega T}) = B(e^{-j\omega T}) \tag{2.43}$$

where T is the sampling period. From the given $T(s)$, we can calculate the complex frequency response data $T(j\omega_k)$ for $k = 1, 2, 3, \ldots, N$. Since we want $T(j\omega)$ and $T(e^{-j\omega T})$ to be as close as possible in a frequency range of interest, by replacing $T(e^{-j\omega T})$ in equation (2.43) with $T(j\omega)$ we have for $\omega = \omega_k$

$$T(j\omega_k) = -A(e^{-j\omega_k T})T(j\omega_k) + B(e^{-j\omega_k T}) + v(j\omega_k)$$

where $v(j\omega_k)$ is the residual which should be minimized. For N different frequencies, we have the following matrix equation:

$$\mathbf{y} = \mathbf{\Phi}\boldsymbol{\theta} + \mathbf{v}$$

where

$$\begin{aligned} \mathbf{y} &= [T(j\omega_1), T(j\omega_2), \cdots, T(j\omega_N)]^T \\ \mathbf{v} &= [v(j\omega_1), v(j\omega_2), \cdots, v(j\omega_N)]^T \\ \boldsymbol{\theta} &= [a_1, \cdots, a_n, b_0, \cdots, b_m]^T. \end{aligned}$$

and

$$\mathbf{\Phi} = \begin{bmatrix} -e^{-j\omega_1 T}T(j\omega_1) & \cdots & -e^{-j\omega_1 nT}T(j\omega_1) & 1 & e^{-j\omega_1 T} & \cdots & e^{-j\omega_1 mT} \\ -e^{-j\omega_2 T}T(j\omega_2) & \cdots & -e^{-j\omega_2 nT}T(j\omega_2) & 1 & e^{-j\omega_2 T} & \cdots & e^{-j\omega_2 mT} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ -e^{-j\omega_N T}T(j\omega_N) & \cdots & -e^{-j\omega_N nT}T(j\omega_N) & 1 & e^{-j\omega_N T} & \cdots & e^{-j\omega_N mT} \end{bmatrix}.$$

Now we can obtain a LS estimate

$$\hat{\boldsymbol{\theta}}_{LS} = [Re(\mathbf{\Phi}^*\mathbf{\Phi})]^{-1} Re[\mathbf{\Phi}^*\mathbf{y}]. \tag{2.44}$$

The elements of $\boldsymbol{\theta}$ are the coefficients of the z-domain model. The estimation can be improved by using an iterative procedure similar with the M-GLS algorithm presented in the previous section.

To investigate the performance of the proposed method, a second-order s-domain filter of which the transfer function is

$$T(s) = \frac{1}{1+s+s^2},$$

was transformed, and the results were compared with those obtained by the well-known bilinear transformation with prewarping. The transformed z-domain gain and phase responses and their absolute deviations from the s-domain responses are shown in Fig. 2.20 and Fig. 2.21 for $\omega_s/\omega_p = 20\pi$ and for $\omega_s/\omega_p = 2\pi$, respectively, where ω_s is the sampling frequency and $\omega_p = 1$ is the pole frequency.

The critical frequency for prewarping in the bilinear transformation was set to ω_p. A set of 100 frequency samples were uniformly selected from the normalized frequency range of 0 to 0.8 for the complex LS s-to-z transformation and the iteration limit was set to 5. The transformed z-domain transfer functions are as follows:

- Prewarped Bilinear Transformation with $T = 0.1$

$$T_1(z^{-1}) = \frac{2.37915764e-03 + 4.75831529e-03z^{-1} + 2.37915764e-03z^{-2}}{1.00000000e+00 - 1.89539638e+00z^{-1} + 9.04913013e-01z^{-2}}$$

- Complex LS s-to-z Transformation with $T = 0.1$

$$T_2(z^{-1}) = \frac{8.31715104e-04 + 7.92425606e-03z^{-1} + 7.52360402e-04z^{-2}}{1.00000000e+00 - 1.89532901e+00z^{-1} + 9.04837342e-01z^{-2}} \qquad (2.45)$$

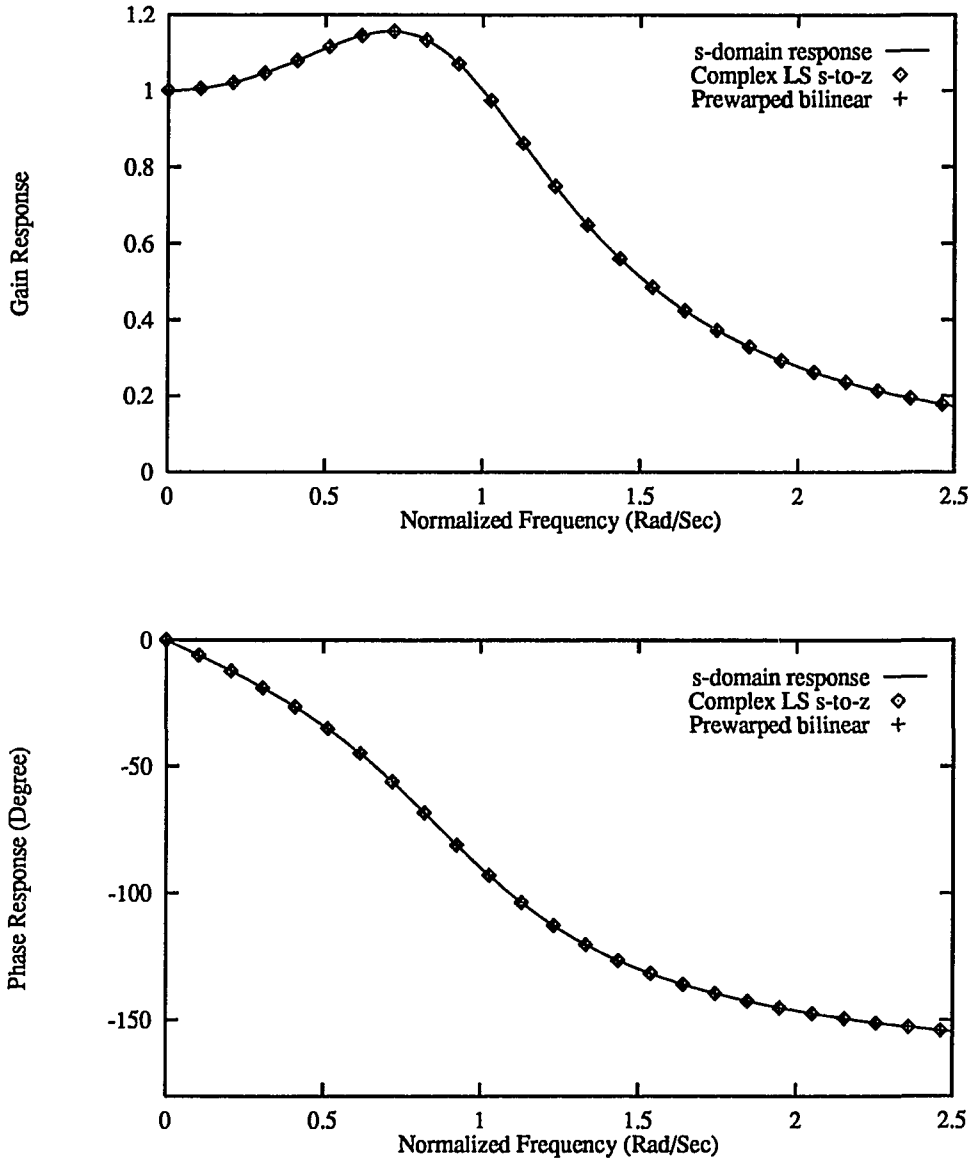
- Prewarped Bilinear Transformation with $T = 1.0$

$$T_3(z^{-1}) = \frac{1.61781590e-01 + 3.23563180e-01z^{-1} + 1.61781590e-01z^{-2}}{1.00000000e+00 - 7.60595211e-01z^{-1} + 4.07721571e-01z^{-2}}$$

- Complex LS s-to-z Transformation with $T = 1.0$

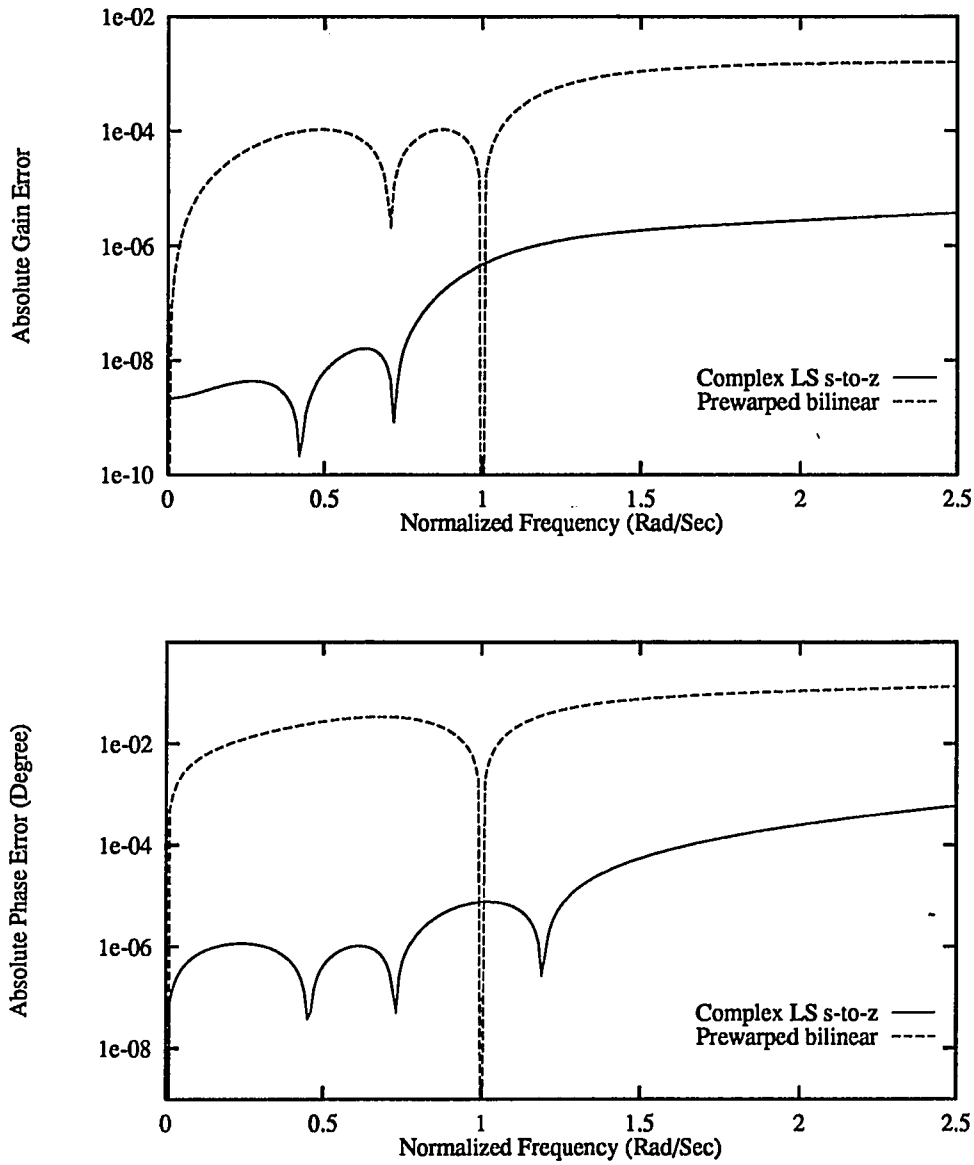
$$T_4(z^{-1}) = \frac{7.41585933e-02 + 4.88408867e-01z^{-1} + 1.75869912e-02z^{-2}}{1.00000000e+00 - 7.83985434e-01z^{-1} + 3.64119370e-01z^{-2}} \qquad (2.46)$$

It can be seen that the complex LS s-to-z transformation gives much better results than the prewarped bilinear transformation in most frequencies of interest except for dc and ω_p. Note that ω_p was the critical frequency for prewarping. At a high sample rate $\omega_s/\omega_p = 20\pi$ ($T = 0.1$), both methods give fairly good results. However, at a low sample rate $\omega_s/\omega_p = 2\pi$ ($T = 1.0$), their performances are degraded. It has been observed from extensive computation that in the complex LS s-to-z method, the performance improvement with iteration is almost negligible, and the number of the frequency samples also has little effect on the performance. Except for the sample rate, the factor that can affect the performance is the frequency sampling



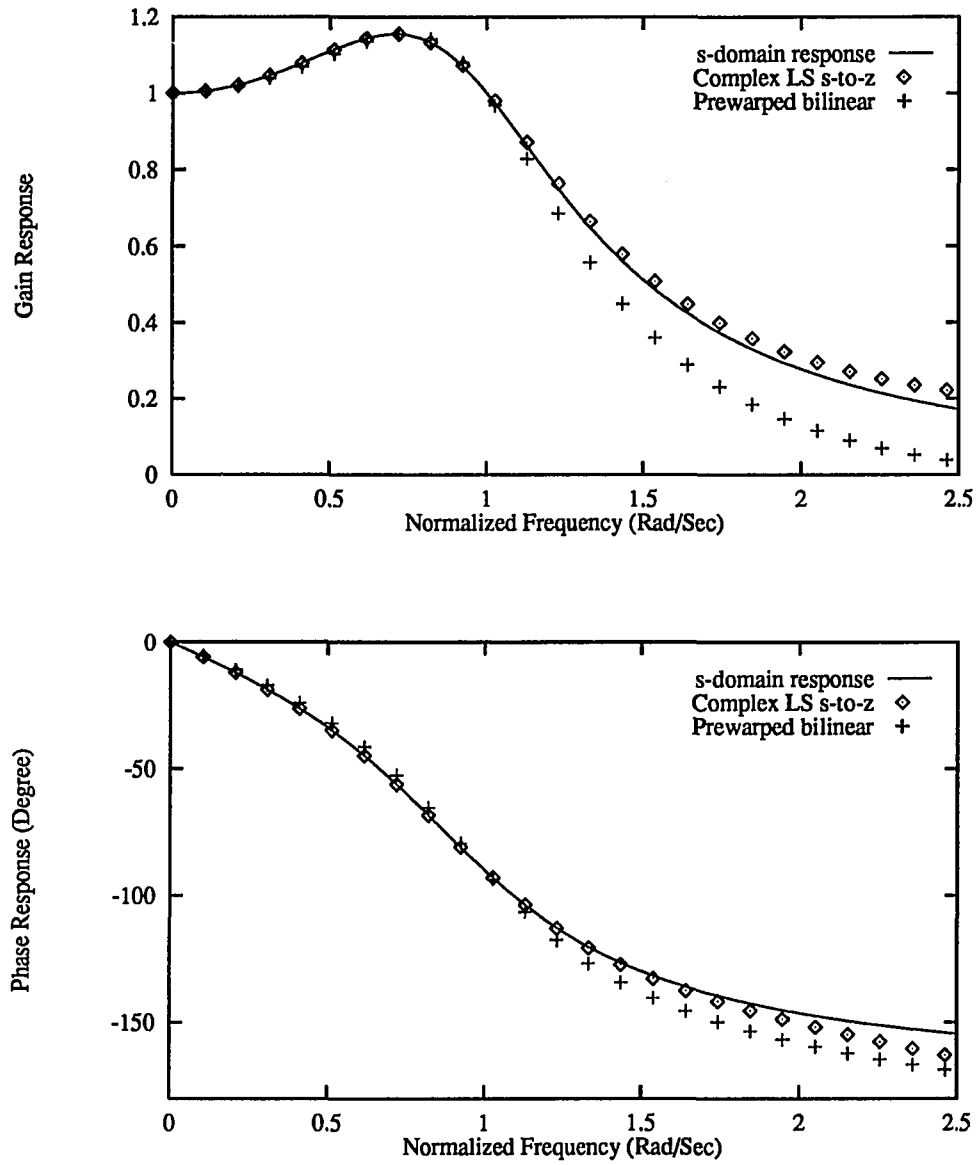
(a)

Figure 2.20: Response of s-domain second-order lowpass filter and z-domain equivalents for $\omega_s/\omega_p = 20\pi$ (a) Gain and phase responses



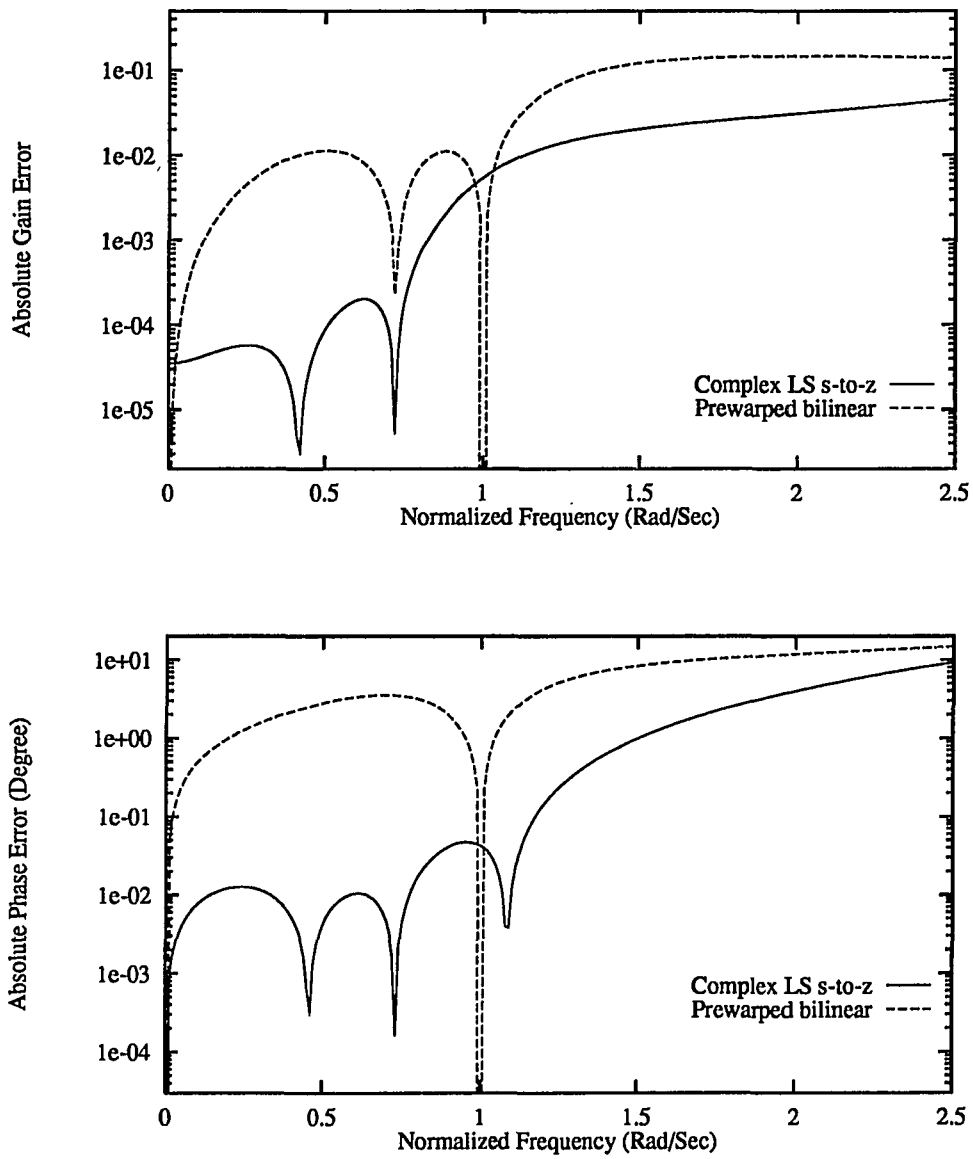
(b)

Figure 2.20: (continued) (b) Absolute gain and phase errors



(a)

Figure 2.21: Response of s-domain second-order lowpass filter and z-domain equivalents for $\omega_s/\omega_p = 2\pi$ (a) Gain and phase responses



(b)

Figure 2.21: (continued) (b) Absolute gain and phase errors

range. Thus, the frequency sampling range should be carefully selected such that it can cover the frequency range of interest.

The complex LS s-to-z method is very attractive because it performs very accurate transformation. In the frequency range of interest, it offers improvements in accuracy by well over two decades for $\omega_s/\omega_p = 20\pi$ and one decade for $\omega_s/\omega_p = 2\pi$ compared to the popular bilinear transformation method.

2.4.3.2 Complex LS z-to-s Transformation Method The problem of z-to-s transformation is that given a discrete-time transfer function $T(z^{-1})$, determine the continuous-time transfer function such that they have approximately the same characteristics. For this problem, we can use the s-domain system ID algorithm, M-GLS, presented in the previous section with the frequency sample data directly computed from the given z-domain transfer function. The parameter estimation of a model $T(s)$ can be obtained at each iteration from equation (2.44) where

$$\begin{aligned} \mathbf{y} &= [T(e^{-j\omega_1 T}), T(e^{-j\omega_2 T}), \cdots, T(e^{-j\omega_N T})]^T \\ \boldsymbol{\theta} &= [a_1, \cdots, a_n, b_0, \cdots, b_m]^T. \end{aligned}$$

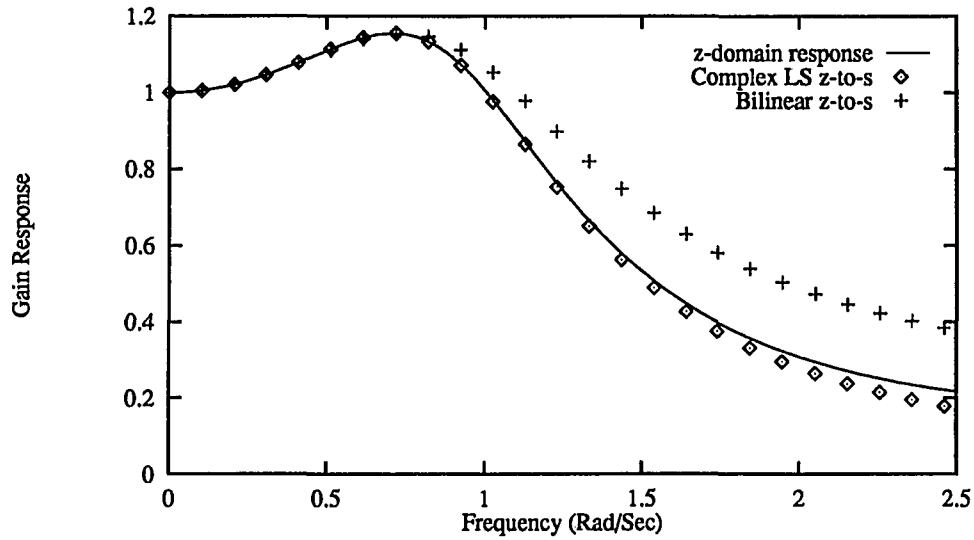
and

$$\Phi = \begin{bmatrix} -(j\omega_1)T(e^{-j\omega_1 T}) & \cdots & -(j\omega_1)^n T(e^{-j\omega_1 T}) & 1 & (j\omega_1) & \cdots & (j\omega_1)^m \\ -(j\omega_2)T(e^{-j\omega_2 T}) & \cdots & -(j\omega_2)^n T(e^{-j\omega_2 T}) & 1 & (j\omega_2) & \cdots & (j\omega_2)^m \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ -(j\omega_N)T(e^{-j\omega_N T}) & \cdots & -(j\omega_N)^n T(e^{-j\omega_N T}) & 1 & (j\omega_N) & \cdots & (j\omega_N)^m \end{bmatrix}.$$

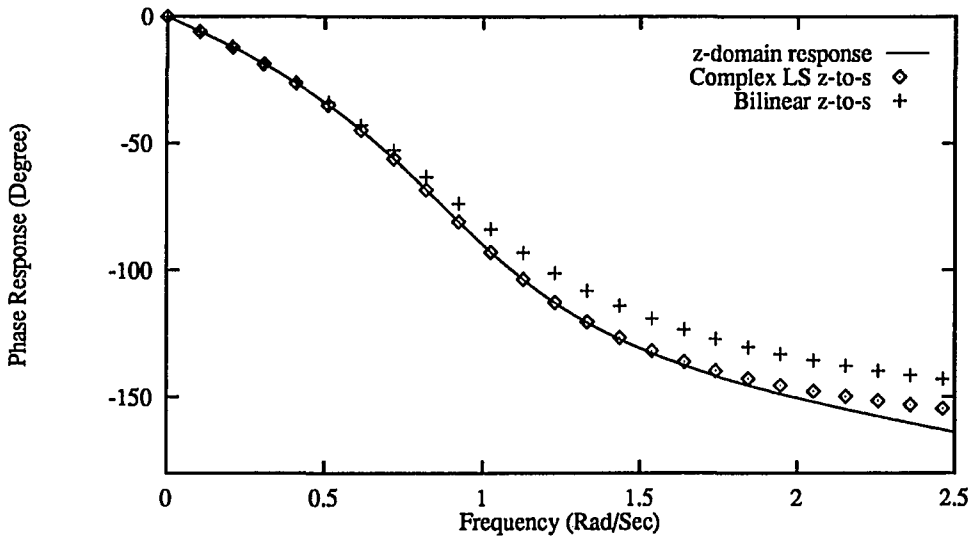
The elements of $\boldsymbol{\theta}$ are the coefficients of the s-domain model.

This complex LS z-to-s method is applied to the z-domain transfer function in (2.45) and (2.46) and the results are compared with those obtained from the bilinear transformation. The transformed s-domain gain and phase responses are shown in Fig. 2.22 for $\omega_s/\omega_p = 20\pi$. In Fig. 2.23 the absolute gain and phase errors are shown for $\omega_s/\omega_p = 2\pi$. A set of 10 frequency samples were uniformly selected from the normalized frequency range of 0 to 1.0 for the complex LS s-to-z transformation and the iteration limit was set to 1.

As in the complex LS s-to-z transformation, the performance of the complex LS s-to-z method has little to do with the iteration number and the number of the frequency samples but gives much better results than the bilinear transformation. This indicates that the complex

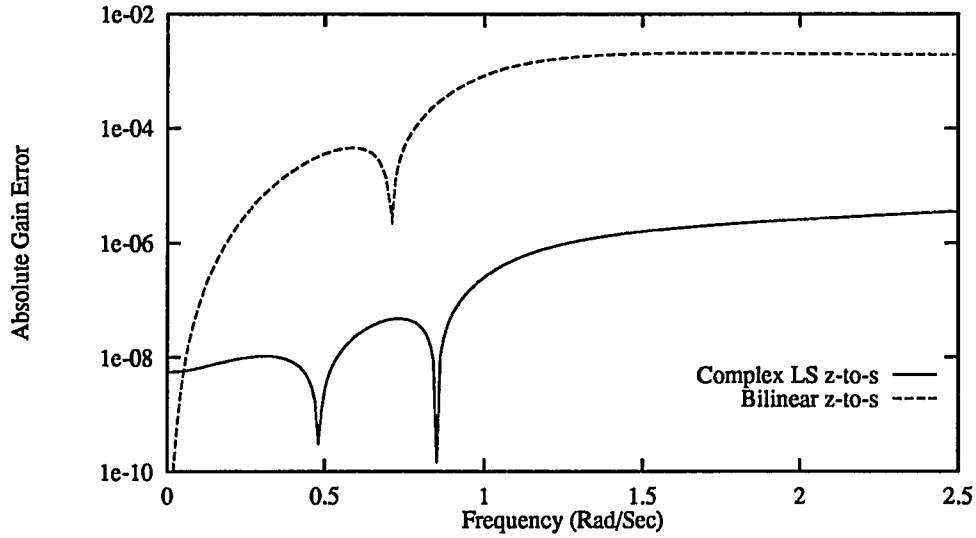


(a)

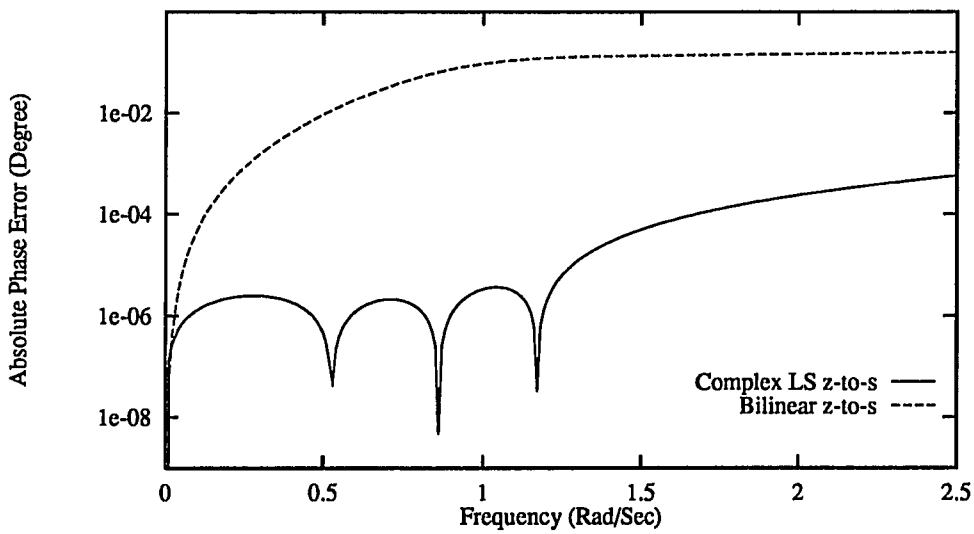


(b)

Figure 2.22: Response of z-domain second-order lowpass filter and s-domain equivalents for $\omega_s = 2\pi$. (a) Gain responses (b) Phase responses



(a)



(b)

Figure 2.23: Accuracy comparison between Complex LS z-to-s transformation and Bilinear z-to-s transformation for $\omega_s = 20\pi$. (a) Absolute gain error (b) Absolute phase error

LS methods can accurately perform the s-to-z or z-to-s transformations with only one matrix inversion and with a few frequency data. Of course, the number of frequency samples must be greater than or equal to $m+n+1$ where n and m are the order of denominator and numerator, respectively.

In section 2.4.1 several LS methods have been discussed for estimating the parameters of a z-domain model from the input-output sample data. The next step of the indirect method 2 is thus to estimate a s-domain model from the identified z-domain model. The complex LS z-to-s method can thus be applied to this problem very well.

2.4.4 Frequency Response Measurement Methods

Frequency response measurements at discrete frequencies are required for continuous-time system identifications (see Fig. 2.7 path $< 3 >$). Accurate frequency response measurements are crucial, since the accuracy of frequency response measurements directly affects the accuracy of the continuous-time system identification through the indirect method 3. The measurement circuit implementation cost as well as the accuracy should be taken into account because the cost and the accuracy are usually in a trade-off relation.

Two frequency response measurement algorithms have been reported in literature. They use sinusoidal inputs and collect input and output time-domain samples for further interpretation. One uses FFT algorithms [53], and the other uses least squares algorithms [67] based on a first-order moving average (MA) model to estimate the frequency responses. A least squares (LS) algorithm based on a first-order auto-regressive moving average (ARMA) model has been presented and compared with the two algorithms mentioned above through extensive Monte Carlo based simulations [4]. The FFT method requires a large number of consecutive samples and thus, requires high-speed A/D converters and sample-and-hold circuits. In contrast, in the least squares methods each data set contains only a few consecutive input-output samples, and each set can be grabbed randomly or asynchronously. Thus, the least squares method requires lower cost for data acquisition hardware implementation than the FFT method.

If a linear continuous-time system is excited by a sinusoidal input

$$x(t) = Acos(\omega_o t), \tag{2.47}$$

then the output, in the steady state, can be described by

$$y(t) = GAcos(\omega_o t + \phi),$$

where G and ϕ are the gain and phase responses of the system at frequency ω_o. The general frequency response measurement problem is to find G and ϕ at a finite number of frequencies. Since experimentally obtained data are inevitably contaminated by some noise and system nonlinear effects, a reasonable approach is to find G and ϕ which best fit the given data.

2.4.4.1 FFT Method In the FFT method [53] the input and output are recorded for N_f sample sets at sampling frequency $f_s = 1/T_s$ from which the frequency response at one point is calculated. Each set contains two samples. N_f must be a power of 2. The output estimate can be defined as

$$y(kT_s) = G_c cos(\omega_o kT_s) + G_s sin(\omega_o kT_s),$$

where

$$G_c = GAcos(\phi) \tag{2.48}$$

$$G_s = -GAsin(\phi). \tag{2.49}$$

It can be easily seen from (2.48) and (2.49) that if G_c ,G_s and A are found, then the gain and phase responses can be calculated by

$$G = \frac{1}{A}\sqrt{G_c^2 + G_s^2}$$

$$\phi = -tan^{-1}\left(\frac{G_s}{G_c}\right).$$

The estimated G_c and G_s which best fit to the data in the least squares sense are closely related to the DFT/FFT of $y(kT_s)$ if the test frequencies are selected to be $\omega_l = 2\pi l/(N_f T_s)$ for integer l, $l = 1, 2, \ldots$. The DFT/FFT of $y(kT_s)$ is

$$Y_n = FFT(y) = \sum_{k=0}^{N_f-1} GAcos\left(\frac{2\pi lk}{N_f} + \phi\right) e^{-j(2\pi nk)/N_f}$$

$$= \begin{cases} \frac{N_f}{2}(G_c - jG_s), & n = l \\ 0, & n \neq l \end{cases}$$

The DFT/FFT of the input $x(kT_s)$ is given by

$$X_n = FFT(x) = \sum_{k=0}^{N_f-1} Acos\left(\frac{2\pi lk}{N_f}\right) e^{-j(2\pi nk)/N_f}$$

$$= \begin{cases} \frac{N_f A}{2}, & n = l \\ 0, & n \neq l \end{cases}$$

For $n = l$, the ratio of Y_l to X_l is given by

$$\begin{aligned} \frac{Y_l}{X_l} &= \frac{G_c - jG_s}{A} \\ &= Ge^{j\phi}. \end{aligned}$$

Thus, in the FFT method the gain and phase responses at $\omega_l = 2\pi l/(N_f T_s)$ are obtained by the ratio of the FFT of $y(kT_s)$ to the FFT of $x(kT_s)$ for $n = l$.

The FFT method estimates the frequency responses accurately in a fast way by the effective FFT algorithm, but using the FFT on $y(kT_s)$ and $x(kT_s)$ for only one frequency point is somewhat inefficient. This inefficiency can be alleviated by using a linearly frequency sweeping signal [40] or a chirp signal [53] of which the frequency changes from a starting value to a final value so as to estimate the frequency responses at several frequencies at once. However, the accuracy might be degraded at fixed N_f compared to that of the original method. Since the FFT method uses a large number of consecutive data, a fast data acquisition system is required for high frequency measurements to avoid the aliasing problem.

2.4.4.2 LS Method with a MA Model In this algorithm [67] a first-order MA model is used to determine the gain and phase responses using the LS algorithm. If the input $x(t)$ is sampled at $t = t_o$ and $t = t_o - T_s$, and the output $y(t)$ at $t = t_o$, then the relationship between the output and input samples can be given by

$$y(t_o) = b_0 x(t_o) + b_1 x(t_o - T_s), \tag{2.50}$$

where

$$\begin{aligned} b_0 &= G[cos\phi + sin\phi cot(\omega_o T_s)] \\ b_1 &= -G\frac{sin\phi}{sin(\omega_o T_s)}. \end{aligned}$$

Equation (2.50) is valid for all t_o provided $\omega_o T_s \neq n\pi$ for integer n. By taking the Fourier Transformation on (2.50) the frequency response at ω_o is given by

$$Ge^{j\phi} = b_0 + b_1 e^{-j\omega_o T_s}.$$

If there is no measurement error, only two data sets are enough to determine b_0 and b_1 and thus, the frequency response exactly. Taking into account the measurement errors, a number of data sets can be used to perform the LS algorithm. This gives an estimation of b_0 and b_1 which best fit to the given data. The estimated coefficient vector can be described by a LS solution

$$\hat{\mathbf{c}} = [\mathbf{A}^T\mathbf{A}]^{-1}[\mathbf{A}^T\mathbf{y_0}], \tag{2.51}$$

where $\hat{\mathbf{c}} = [b_0, b_1]^T$ and $\mathbf{A} = [\mathbf{x}_0, \mathbf{x}_{-1}]$, and $\mathbf{x}_{-1}$ is a sampled input vector, and $\mathbf{x}_0$ and $\mathbf{y}_0$ are the input and output data vectors sampled with a delay T_s. Vectors, $\mathbf{x}_0$, $\mathbf{x}_{-1}$ and $\mathbf{y}_0$ have dimension N_m, where N_m is the number of data sets, and each data set contains 3 samples.

In this method each data set contains only two consecutive input samples and one output sample. Since the data sets can be collected randomly or asynchronously, the data acquisition system can be implemented with three fast sample-and-hold circuits (two for input and one for output) and a slow A/D converter.

2.4.4.3 LS Method with an ARMA Model To utilize a first-order ARMA model for the LS frequency response measurement, one more output samples must be added to the data set of the MA model. The input $x(t)$ and output $y(t)$ are sampled at $t = t_o$ and $t = t_o - T_s$ by four sample-and-hold circuits, and these four samples constitute one data set. Once all four samples are converted by an A/D converter which does not have to be fast, another data set is sampled with time delay T_{int}. Of course, the time interval, T_{int}, must be selected to be long enough for the A/D converter to finish conversion of four sampled and held data. T_{int} does not have to be the same for the whole data acquisition period. The relationship between output samples and input samples for any one data set can be readily obtained as follows by applying basic trigonometric identities to (2.47) at $t = t_o - T_s$:

$$y(t_o) = a_1 y(t_o - T_s) + b_0 x(t_o) + b_1 x(t_o - T_s), \tag{2.52}$$

where

$$\begin{aligned} a_1 &= 1/cos(\omega_o T_s) \\ b_0 &= Gcos(\omega_o T_s + \phi)/cos(\omega_o T_s) \\ b_1 &= -Gcos\phi/cos(\omega_o T_s). \end{aligned}$$

Equation (2.52) is valid for all t_o provided $\omega_o T_s \neq (2n+1)\pi/2$ for integer n. The frequency response at ω_o is given by

$$Ge^{j\phi} = \frac{b_0 + b_1 e^{-j\omega_o T_s}}{1 + a_1 e^{-j\omega_o T_s}}.$$

The LS solution can be obtained from equation (2.51), where $\hat{\mathbf{c}}_a = [a_1, b_0, b_1]^T$ and $\mathbf{A} = [\mathbf{y}_{-1}, \mathbf{x}_0, \mathbf{x}_{-1}]$, and $\mathbf{x}_{-1}$ and $\mathbf{y}_{-1}$ are sampled input and output data vectors, and $\mathbf{x}_0$ and $\mathbf{y}_0$ are the input and output data vectors sampled with a delay T_s. The vectors have dimension N_a, where N_a is the number of data sets, and each data set contains 4 samples. In this method the data acquisition system can be implemented with one more sample-and-hold circuit added to the system for the MA model.

If the signals are sampled at the same rate for all data sets, i.e., T_{int}=T_s, as in the FFT method, then we can increase the number of data sets for the fixed total sample number. One data set can be obtained at every sample time since one input and output sample set contributes to two data sets as previous data and as current data. This is the case as the general time-domain LS system identification methods do, where for the nth-order ARMA model $N-n+1$ data sets can be obtained with N input and output samples. For our first-order case, $2N_a$ data sets can thus be used for the LS problem with increased dimension $2N_a$. This will lead to improved accuracy, but more cost for data acquisition will be required to grab a large number of consecutive data.

Since the cost and the accuracy are in a trade-off relation, one of the two strategies can be selected according to which has a higher priority. This may be one advantage of the ARMA method over the MA method because in the MA method the number of data sets can not be increased that much through sampling with $T_{int} = T_s$, or even though it can be increased by adding one more sample-and-holder, one output sample can contribute to only one data set.

2.4.4.4 Simulation Results The three frequency response measurement algorithms are compared through Monte Carlo based simulations. The nonidealities of the system under test, the data acquisition system, and the excitation signals are included in this simulation. The input-output measurement noise associated with the data acquisition process such as quantization noise and system noise are modeled as uniformly distributed additive random numbers. If the measurement noise is distributed as $U(-\epsilon_n, \epsilon_n)$, and the input signal amplitude is A, then the mean square values of the signal and the noise are

$$\sigma_s^2 = \frac{A^2}{2} \quad \text{and} \quad \sigma_n^2 = \frac{\epsilon_n^2}{3}.$$

The input signal to noise ratio (SNR) is then defined as

$$\begin{aligned} SNR &= 10log_{10}\left(\frac{\sigma_s^2}{\sigma_n^2}\right) \\ &= 10[log_{10}(1.5) - 2log_{10}(A/\epsilon_n)]. \end{aligned}$$

The SNR will be about 40dB for $\epsilon_n = 0.01$ and $A = 1$. The excitation signal nonlinearity and the system nonlinearity are approximated by the second harmonic distortion, and the higher-order distortions are neglected. The total harmonic distortion (THD) of the excitation input signal and the output signal were set to -40dB.

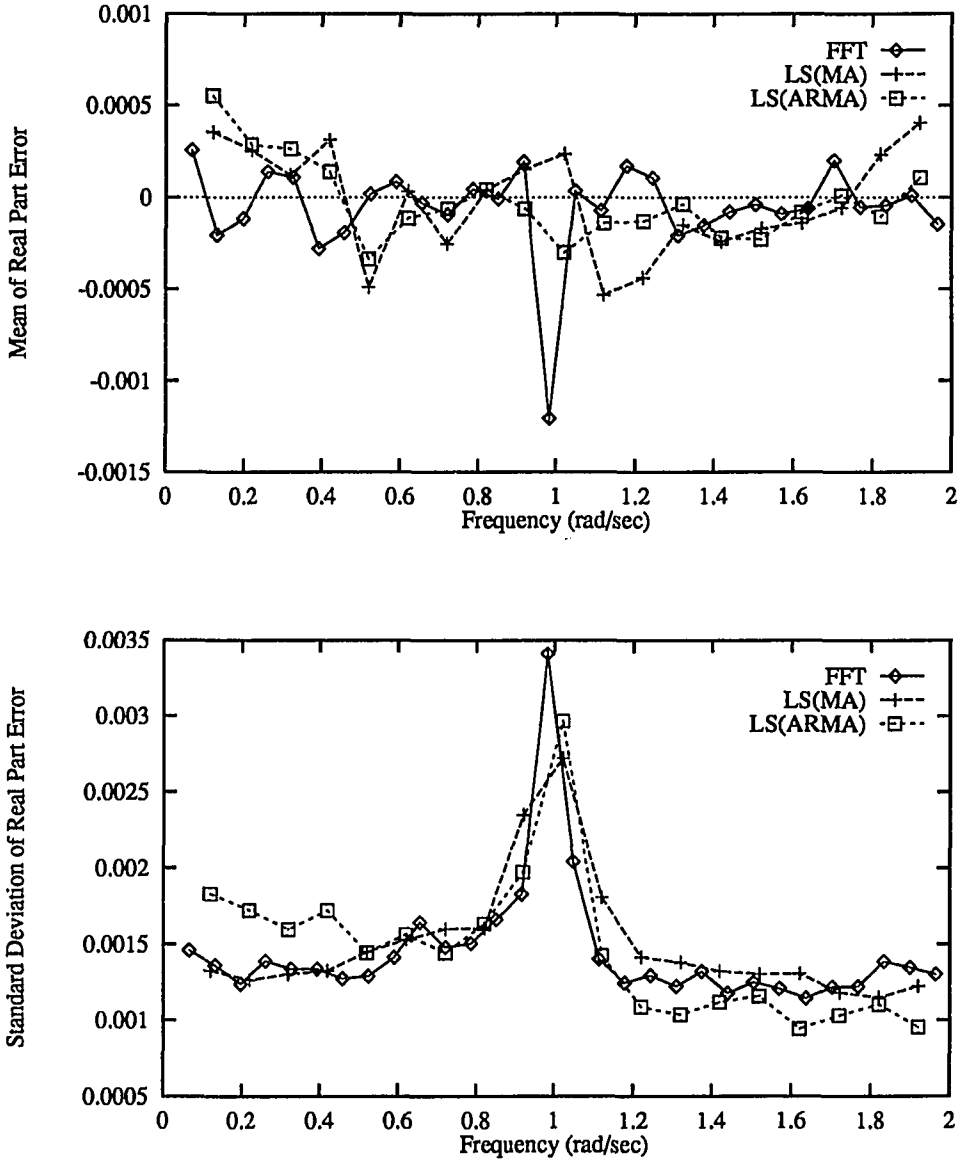
The second-order lowpass notch filter used in Section 2.4.2.3 has been selected again to compare the frequency response measurement algorithms. Fig. 2.24 shows the statistics obtained from 100 independent measurements using the three measurement algorithms. The

mean and standard deviation of the measured real and imaginary part errors are shown in Fig 2.24 (a) and (b), respectively. The input signal to noise ratio (SNR) was set to 40dB. If it is assumed that the measurement noise is dominated by the quantization errors of A/D converters, then the 40dB SNR corresponds to about 7-bit resolution for ± 1V reference levels. The number of data sets was set to 64. The sampling period T_s was chosen to be 1.5sec. The simulation results show that the three algorithms have similar performance. More extensive simulation results can be found in [4]. For a simple test lowpass filter the LS methods have shown accurate measurement results at reasonable noise environment. For 40dB SNR and -40dB THD, the gain error less than 1% and the phase error less than 1° were obtained on the normalized frequency range from dc to 2 (rad/sec). Thus, these LS methods can be well applied to the s-domain system ID algorithms. It has been demonstrated that the M-GLS (or ICLS) algorithm can achieve good results for 1% measurement errors [9].

Although the FFT method has an advantage that it can measure a set of frequency response data with a smaller number of input-output samples using a frequency sweeping signal, for high frequency response measurement it requires higher cost for implementation of data acquisition systems than the LS methods because it needs a large number of consecutive samples. In contrast, the LS methods can be applicable with a low-cost data acquisition system as mentioned earlier. Since the LS method based on a first-order ARMA model which is a more general model, has more fiexibility associated with the number of data sets available from a fixed number of input-output samples as mentioned before, and it is more insensitive to the system nonlinearity which has been demonstrated in [4], it can be chosen to serve the second path $< 3 >$ of the indirect continuous-time system identification method 3.

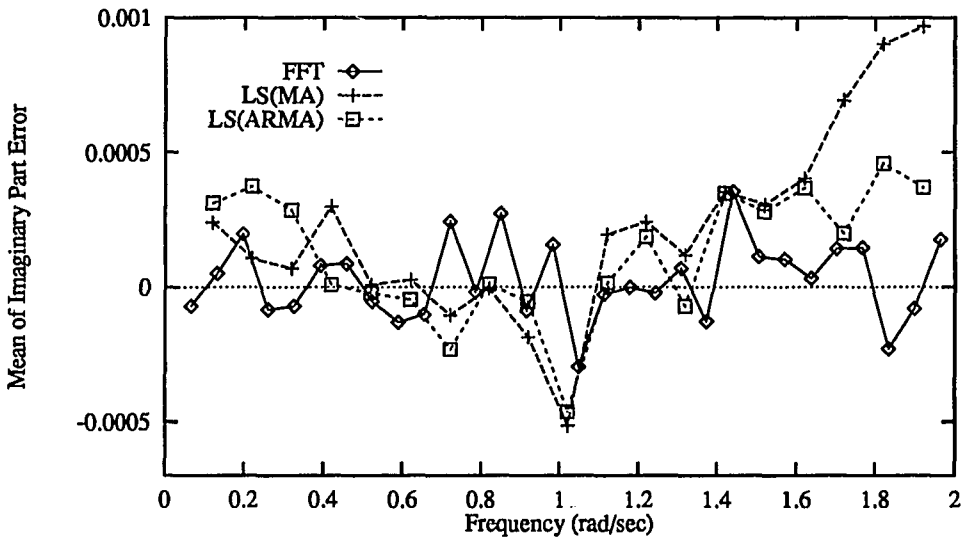
2.4.5 Continuous-Time Filter Identification

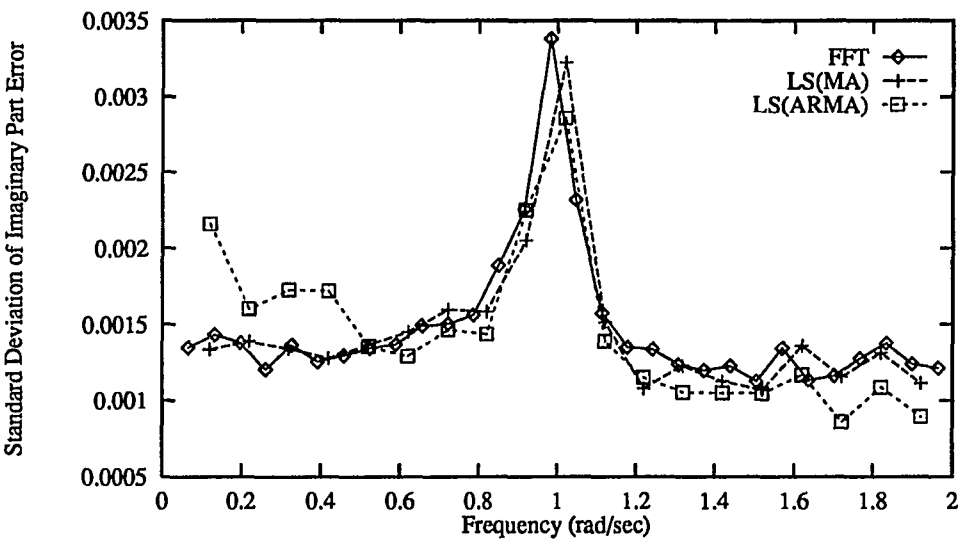
In the previous sections we have discussed LS methods associated with general system identification problems. It has been shown that for the continuous-time filter parameter identification, there are two possible indirect methods, i.e., Method 2 and 3, where the problem can be decomposed into two simple steps. In Method 2, the generalized LS algorithm using an AR noise model can be used for the first step since its better accuracy and stability have been demonstrated for various actual noise models. The complex LS z-to-s transformation method can be chosen to serve for the second part of Method 2 because of its superiority to the well-known bilinear transformation in accuracy. Method 2 can thus accurately estimate the parameters of continuous-time filters if the test input is a persistently exciting signal. A linear system is said to be identifiable if the system is stable and the input test signal is persistently



(a)

Figure 2.24: Performance comparison of frequency response measurement algorithms. Mean and standard deviation computed from 100 independent simulation results (a) Real part error





(b)

Figure 2.24: (continued) (b) Imaginary part error

exciting. If the input test signal is persistently exciting of order $(n+m+1)$, the matrix $\Phi^T\Phi$ will be nonsingular and thus the least squares solution, $\hat{\theta} = (\Phi^T\Phi)^{-1}\Phi^T\mathbf{y}$, does exist. The pseudo-random binary sequence (PRBS) or the linearly frequency sweeping signal can be used for a persistently exciting test input signal for the z-domain system ID which is the first step of Method 2. The PRBS which serves as a practical white noise signal can be generated from a shift register circuit cascaded with a lowpass filter [55]. The linearly frequency sweeping signal can be generated by applying a ramp or triangle signal to a voltage controlled oscillator circuit [40].

Although Method 2 is one of the promising approaches to the continuous-time filter identification, it has a shortcoming when applied to the filters operating at high frequencies because its z-domain system ID algorithm requires a large number of consecutive samples. Therefore, a fast and thus high-cost data acquisition circuit is required. For this reason, Method 3 will be selected as the continuous-time filter ID method. In Method 3 the frequency response data can be first measured using the LS algorithm based on a first-order ARMA model, LS(ARMA), and secondly, the parameters of the filter are estimated using the s-domain system ID algorithm, M-GLS or iterative complex least squares (ICLS) algorithm which has been shown to be robuster than other algorithms discussed in Section 2.4.2. In this section, the combination of the two parts of Method 3, LS(ARMA) and ICLS, are investigated.

In Section 2.4.2, it has been demonstrated that the ICLS (or M-GLS) algorithm is very robust when the frequency response data are assumed to be corrupted with an independent uniform noise sequence for both real and imaginary parts. If the frequency response data are measured from the LS(ARMA) algorithm using input-output samples, the measurement noise may not be independent. To investigate this, the lowpass notch filter used in Section 2.4.2.3 is tested again. The identified results using the data measured by the LS(ARMA) algorithm are shown in Fig. 2.25 and Fig. 2.26.

In this simulation the number of data sets for each frequency response measurement was set to 50, and the frequency response data length for the s-domain system ID was also set to 50. The THD of the input and output signals was set to -40dB. The SNR that has been defined in Section 2.4.4.4 was set to 40dB for Fig 2.25. The absolute mean ($|E_s[\theta_{ei}]|$) and variance ($E_s[\theta_{ei}^2] - E_s^2[\theta_{ei}]$) of estimation error computed from 100 independent identifications are shown in Fig. 2.25, where θ_{ei}'s are the individual elements of the parameter estimation error vector $\boldsymbol{\theta}_e$ $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$. Compared with the results of Fig. 2.18 which were obtained from ideal frequency response data corrupted with independent uniform noise, the bias reduction schemes, GLS or M-GLS, do not offer substantial improvements over the LS algorithm if the data obtained from

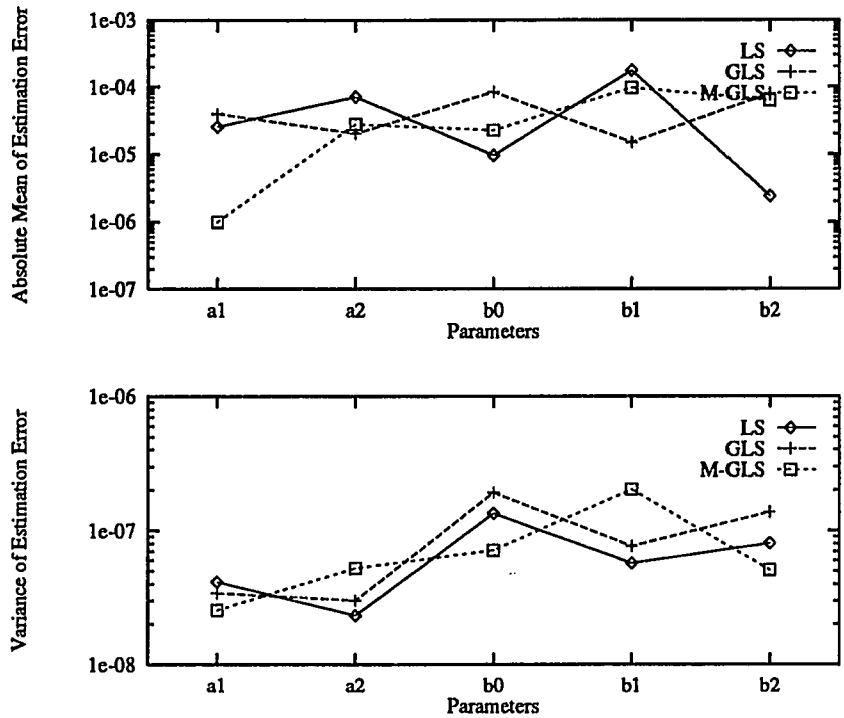


Figure 2.25: Absolute mean ($|E_s[\theta_{ei}]|$) and variance ($E_s[\theta_{ei}^2] - E_s^2[\theta_{ei}]$) of estimation error computed from 100 independent identifications. The frequency response data were measured from the LS(ARMA) algorithm (SNR=40dB)

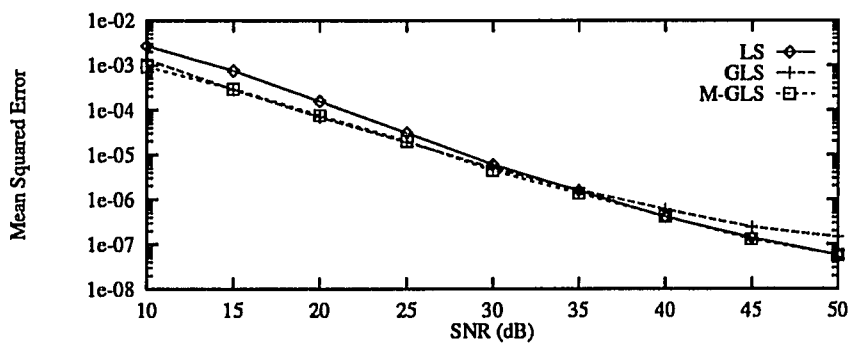


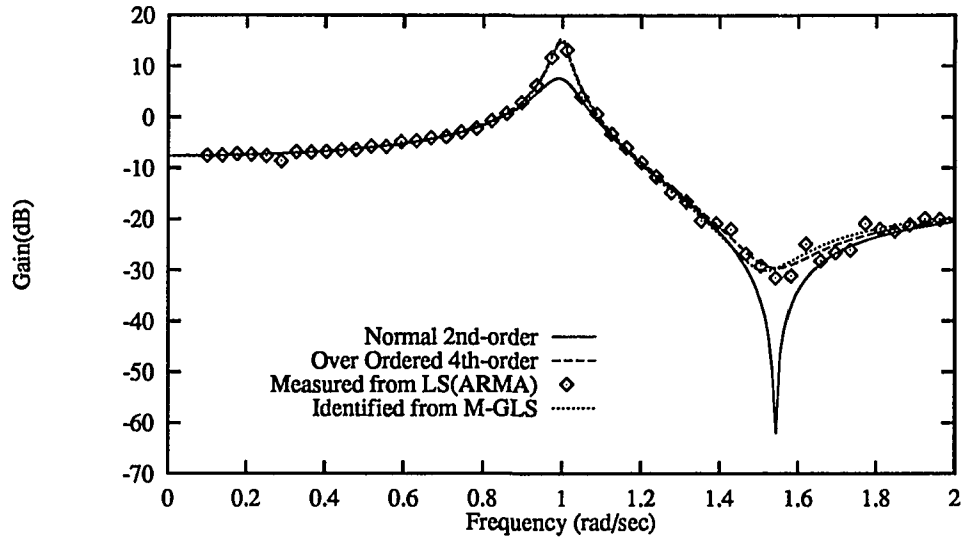
Figure 2.26: Performance comparison of s-domain system ID algorithms using the frequency response data measured from the LS(ARMA) algorithm. Mean squared estimation error ($E_s[\theta_e^T \theta_e]$) versus SNR, computed from 100 independent identifications

the frequency measurement algorithm are used. This is due to the correlation of the frequency response measurement errors. The mean squared estimation errors ($E_s[\theta_e^T\theta_e]$) versus SNR are shown in Fig. 2.26. At low noise environment the accuracy improvement by the M-GLS algorithm over the LS algorithm is almost negligible. At high noise environment, however, the M-GLS still shows better estimation results than the LS algorithm.

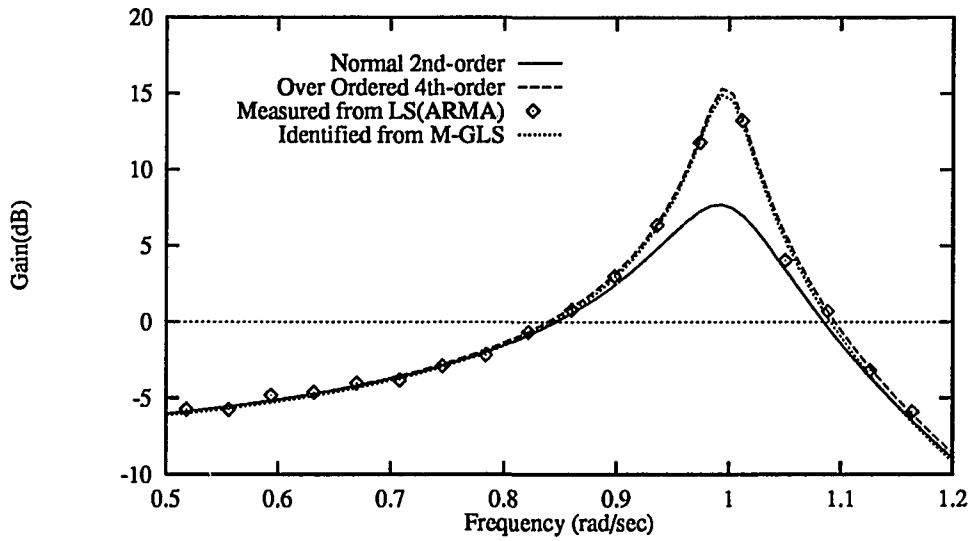
Fig. 2.25 shows that the system ID method using the LS(ARMA) and ICLS algorithms can give very accurate identification results with a medium resolution A/D converter (SNR=40dB) and an inexpensive excitation system (THD=−40dB). For example, the sample mean and variation of 100 estimation errors are 0.036% and 28.8 ppm, respectively, for the parameter b_2 of which the ideal value is given in equation (2.41). Since tuning accuracy depends on the ID accuracy, tuned filters could maintain the ID accuracy if filter adjustments were performed correctly. However, the tuning accuracy will be degraded by the limited resolution of the filter control circuit (D/A converters) and the filter parasitic effects.

So far, the parasitic effects associated with the system to be identified has been neglected so that the order of the system to be identified and the order of the model have been the same. However, the actual continuous-time filter usually has a higher order due to the parasitic poles and zeros as discussed in Section 2.3. In our case, the ID problem of the continuous-time filters is thus to identify the over-ordered system with a reduced-order model because the system ID model should have the same order as the desired one for a simple filter adjustment procedure as mentioned before. Note that zero error in ID is not possible, and accuracy degradation is expected because the ID is now approximating the over-ordered system. The accuracy degradation can be avoided if a higher-order model is used for identification of the over-ordered system. Actually, it has been found from simulations that adding one more pole to the ideal model leads to much better ID results. However, increasing the order of the system ID model will result in a much complicated filter tuning/adjustment procedure due to loss of the independent adjustability of the filter transfer function coefficients. Although they are complicated, there may exist many ways to map the more accurate ID results obtained by using a higher-order model to the control parameter values required for filter adjustment. However, we will not deal with this approach here and will use ideal system ID models in order to maintain the filter adjustment procedure very simple.

To investigate the effects of the over-ordering problem on the ID accuracy, the actual over-ordered 4th-order notch filter has been identified using an ideal 2nd-order model. The gain and phase responses of the ideal 2nd-order filter and the actual over-ordered 4th-order filter are depicted in Fig. 2.27, where the over-ordering factor ω_o/ω_p is 0.04.

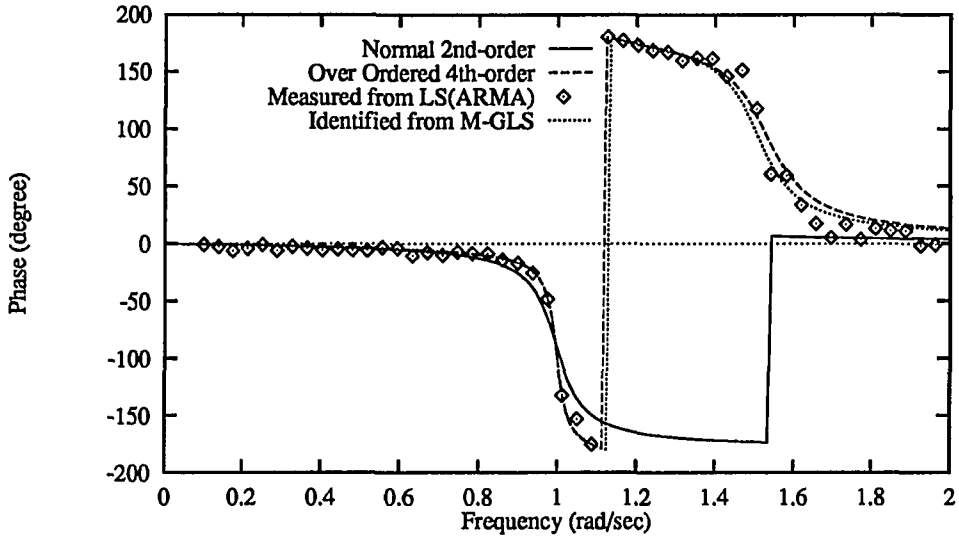


(a)



(b)

Figure 2.27: Identified results of the over-ordered 4th-order system using a 2nd-order model (a) Gain responses (b) Magnified view of (a) (SNR=20dB, $\omega_o/\omega_p = 0.04$)

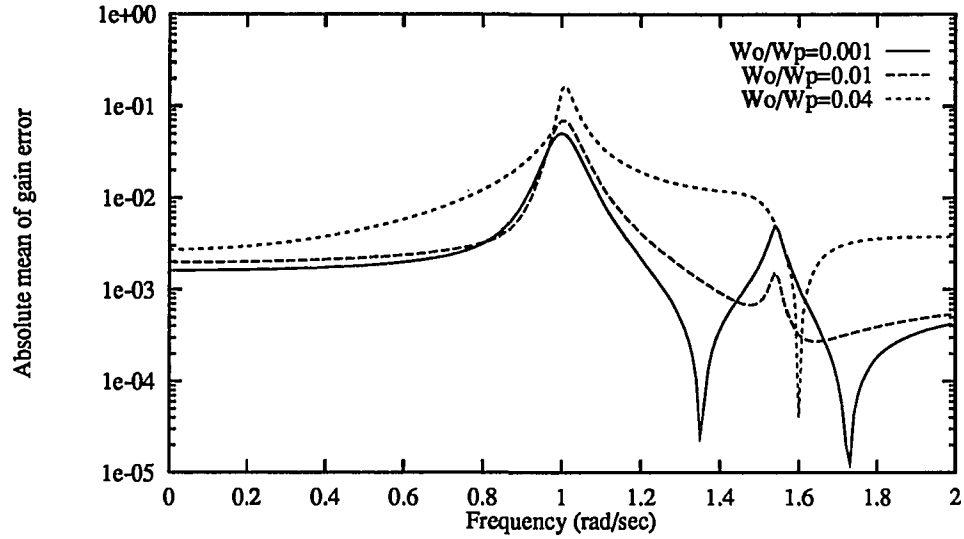


(c)

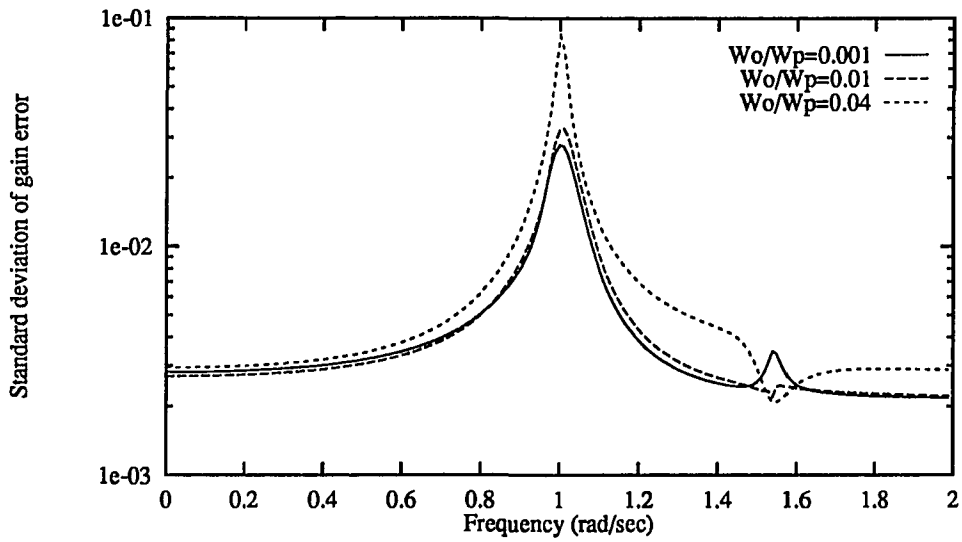
Figure 2.27: (continued) (c) Phase responses

With the same conditions as before, the frequency responses measured from the LS(ARMA) method and the identified responses from the M-GLS algorithm based upon the measured data are also shown in the figure. Even with a significant noisy environment SNR=20dB, the ICLS algorithm along with the LS(ARMA) algorithm can identify the over-ordered system with a good accuracy using a lower-order model although it can be seen in Fig. 2.28 that the ID accuracy is degraded as the over-ordering factor increases.

The reason why a 2nd-order model can approximate an over-ordered 4th-order system accurately as shown in the above simulation results can be analyzed as follows. The 4th-order system which is an over-ordered one of the ideal 2nd-order system due to over-ordering effects addressed in Section 2.3 has parasitic poles typically at much higher frequencies than the system pole frequencies, so the deviations of the system responses from the ideal responses are not severe. Even in the presence of severe deviations due to high over-ordering effects, the iterative complex least squares algorithm can identify the over-ordered system with a good accuracy using an ideal low-order model. The reason seems to be that the ID model has full degrees of freedoms for system ID, i.e., for a second-order case five coefficients can be used to identify the over-ordered system, but most filter types except for allpass filters have missing terms



(a)



(b)

Figure 2.28: Effects of the over-ordered factor (ω_o/ω_p) on the identification accuracy (a) Absolute mean of gain error (b) Standard deviation of gain error computed from 100 independent identifications (SNR=20dB)

in the numerator polynomial of their transfer functions. Therefore the identified coefficients corresponding to the missing terms help to compensate for the errors due to over-ordering effects. For example, the ideal transfer function of a 2nd-order bandpass filter does not contain the constant and s^2 terms in the numerator, but the actual 4th or higher order system will be identified with a 2nd-order model which has both constant and s^2 terms, so these terms can be used for approximating the higher-order system. Thus, this system ID method can be simply applied to our tuning scheme even for higher-frequency and high-Q applications.

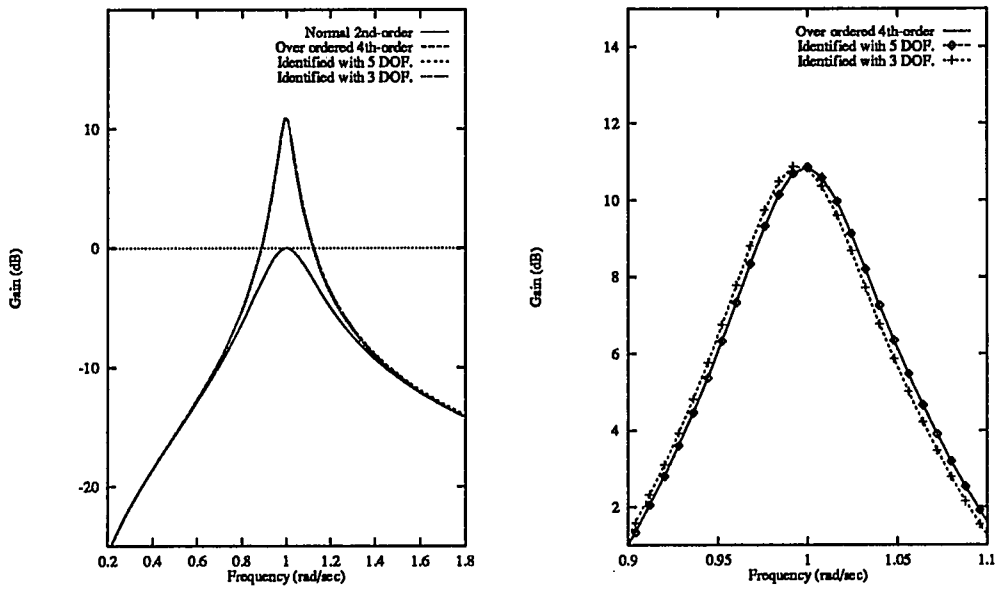
The digitally programmable continuous-time filter under test, however, does not offer the adjustability of s^2 term in the numerator, so the model for system ID will have only 4 degrees of freedom. Moreover, for a bandpass filter only 3 degrees of freedom can be used to identify the over-ordered system because the constant term of the numerator does not have adjustability for a small value due to the restricted g_m adjustable range of the programmable filter . This will degrade the system ID results. To investigate this problem the over-ordered 4th-order bandpass filter with the over-ordering factor of 0.1 has been identified using two second-order models: a 5 degree-of-freedom (DOF) model and a 3 degree-of-freedom model without the s^2 term and the constant term in the numerator. The identified results are shown in Fig. 2.29. It can be seen in the magnified plots that the identification using a 3 DOF model gives degraded results. The degradation can not be improved even with noiseless measurements because it is mainly due to the model error.

2.5 Adjustment

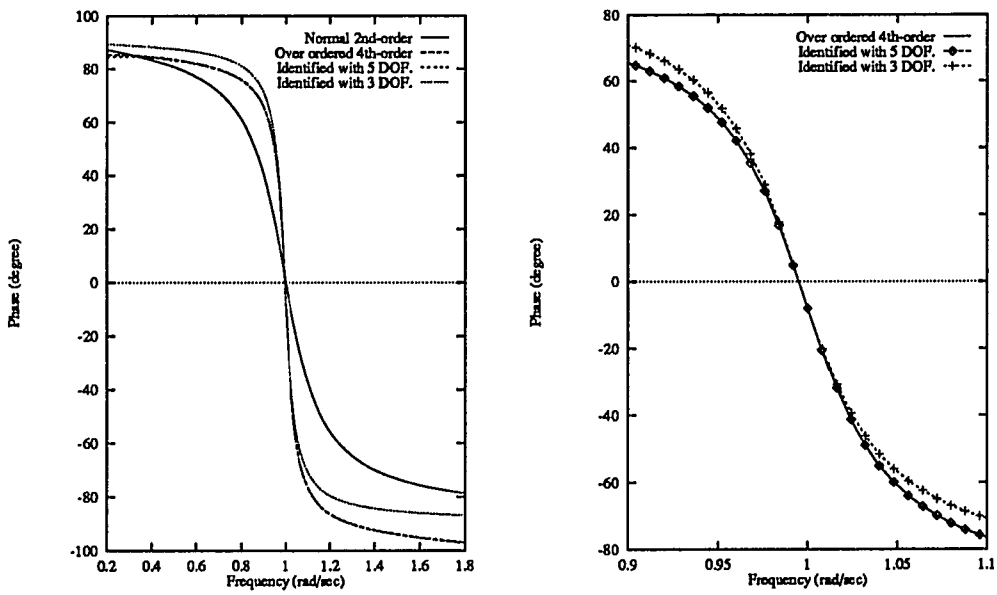
In the previous section, various system ID methods have been discussed which estimate the continuous-time filter model using time-domain input-output samples. It has been shown that the combination of the s-domain ID algorithm, ICLS, and the frequency response measurement algorithm, LS(ARMA), can serve as a good continuous-time filter ID method even in the presence of parasitics. Now, the remaining part of the tuning scheme is the filter adjustment using the identified results as shown in Fig. 2.1. The procedure of obtaining the component correction vector ΔG can be very simple thanks to the independent or sequential transfer function coefficient adjustability of the digitally programmable continuous-time filter.

2.5.1 Tuning Algorithm

The rudiment of the adjustment or tuning algorithm is to estimate the process dependent parameters of the OTAs based upon system ID results and to determine the control parameter



(a)



(b)

Figure 2.29: Identified (a) Gain and (b) Phase responses of the over-ordered 4th-order system with 2nd-order models: a 5 degree-of-freedom model and a 3 degree-of-freedom model (SNR=40dB, $\omega_o/\omega_p = 0.1$)

values for filter adjustment. In the digitally programmable continuous-time filter, the linearized OTA proposed by Nedungadi [42] was used for the input transconductance stage to achieve better linearity [39]. The schematic diagram of the OTA is shown in Fig. 2.30.

From [42], the control mechanism relating the OTA transconductance gain g_{mi} to its control voltage V_{ci} and current mirror gain M_i for $i = 1, \ldots, 5$, is characterized by the linear equation

$$g_{mi}(V_{ci}, M_i) = 2\sqrt{\frac{k_i k_b}{3}}(V_{ci} - V_{ss} - V_{Ti})M_i$$

where k_i and k_b are the process dependent constants corresponding to input pair transistors and bias transistors respectively, V_{Ti} is threshold voltage, M_i is the controllable output stage mirror gain, and V_{ci} is the tail voltage which is the output of a D/A converter. The equation can be rewritten as,

$$g_{mi}(V_{ci}, M_i) = M_i m_i(k_i, k_b)[V_{ci} + n_i(V_{Ti})] \tag{2.53}$$

where

$$m_i(k_i, k_b) = 2\sqrt{\frac{k_i k_b}{3}} \tag{2.54}$$

$$n_i(V_{Ti}) = -(V_{ss} + V_{Ti}) \tag{2.55}$$

The m_i and n_i are the process dependent parameters and the V_{ci} and M_i are the control parameters for filter tuning. The OTA tail bias voltage V_{ci} is used for smaller (fine) adjustment while the output current mirror gain Mi is used for more significant (coarse) adjustment. Thus, the transconductance g_{mi} of $i'th$ OTA of each biquad can be controlled by changing V_{ci} and M_i. The basic idea of the tuning algorithm is to calculate the control parameters V_{ci} and M_i for filter adjustment such that the identified g_{mi} of each OTA becomes close to its design (nominal) value. The identified g_{mi} of each OTA can be obtained from identified coefficients through the relations (2.4) to (2.7).

The tuning procedure is divided into three parts: initial implementation, first iteration, and subsequent iterations. The whole tuning procedure flow chart is shown in Fig. 2.31. In the initial implementation, the initial parameters m_i and n_i for $i = 1, 2, \ldots, 5$ of each biquad are set to their design values, and the initial control parameters V_{ci} and M_i are calculated. These control parameter values are used for initial implementation of the filter. Each iteration consists of four steps as follows:

1. System identification using the system ID method
2. Estimation of process parameters m_i and n_i from the identified transfer function coefficients

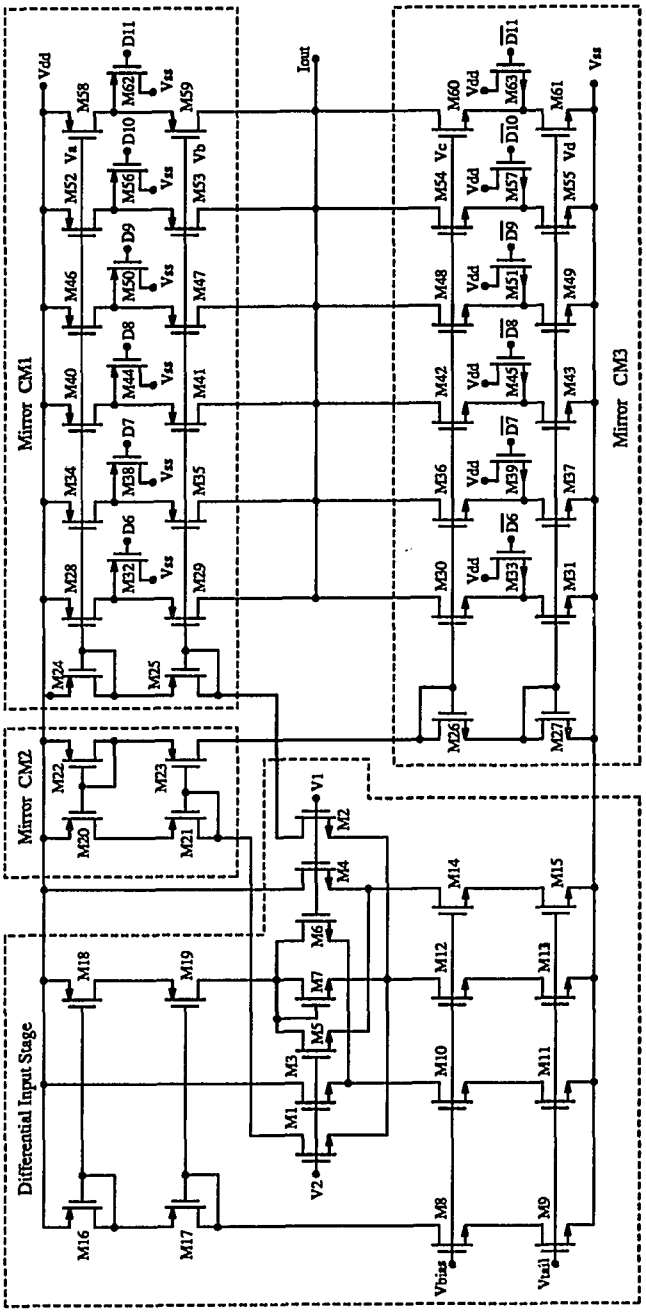


Figure 2.30: Schematic diagram of the OTA

Table 2.2: Expressions for $m_i^{(1)}$, $n_i^{(1)}$ and $V_{ci}^{(k)}$

i	$m_i^{(1)}$	$n_i^{(1)}$	$V_{ci}^{(k)}$
1	$\frac{(b_0^{(0)}/\sqrt{a_0^{(0)}})C}{M_1^{(0)}(V_{c1}^{(0)}+n_1^{(0)})}$	$n_1^{(0)}$	$\frac{\bar{b_0}C}{\sqrt{\bar{a_0}}m_1^{(k)}M_1^{(k)}} - n_1^{(k)}$
3	$\frac{\sqrt{a_0^{(0)}}C}{M_3^{(0)}(V_{c3}^{(0)}+n_3^{(0)})}$	$n_3^{(0)}$	$\frac{\sqrt{\bar{a_0}}C}{m_3^{(k)}M_3^{(k)}} - n_3^{(k)}$
4	$\frac{(b_1^{(0)}+\sqrt{a_0^{(0)}})C}{M_4^{(0)}(V_{c4}^{(0)}+n_4^{(0)})}$	$n_4^{(0)}$	$\frac{(\bar{b_1}+\sqrt{\bar{a_0}})C}{m_4^{(k)}M_4^{(k)}} - n_4^{(k)}$
5	$\frac{a_1^{(0)}C}{M_5^{(0)}(V_{c5}^{(0)}+n_5^{(0)})}$	$n_5^{(0)}$	$\frac{\bar{a_1}C}{m_5^{(k)}M_5^{(k)}} - n_5^{(k)}$

3. Calculation of control parameters V_{ci} and M_i from the estimated m_i and n_i

4. Adjustment using the obtained control parameters

The estimation formulas of m_i and n_i for $i = 1, 3, 4, 5$ are summarized in Table 2.2 and 2.3 with special conditions $B_{bp} = 1$, and $B_{lp} = 0$. It is assumed that $C_6 = C_7 = C$ and $g_{m2} = g_{m3}$, so the expressions for m_2 and n_2 are the same as those for m_3 and n_3. At each iteration the control voltages can be calculated by using the equations shown in Table 2.2. If the control voltage V_{ci} exceeds a specified range, the current mirror gain M_i should be adjusted to keep the control voltage within the controllable range.

In the tables, the followings are should be noticed:

- $a_0^{(k)}, a_1^{(k)}, b_0^{(k)}$, and $b_1^{(k)}$: identified transfer function coefficients at (k+1)'th iteration
- $\bar{a_0}, \bar{a_1}, \bar{b_0}$, and $\bar{b_1}$: design (ideal) transfer function coefficients
- $V_{ci}^{(k)}$: control voltage of g_{mi} at k'th iteration
- $M_i^{(k)}$: current mirror gain of g_{mi} at k'th iteration
- $V_{ci}^{(0)}, M_i^{(0)}, m_i^{(0)}, n_i^{(0)}$, and C : design values
- $p1 = k - 1$ and $p2 = k - 2$

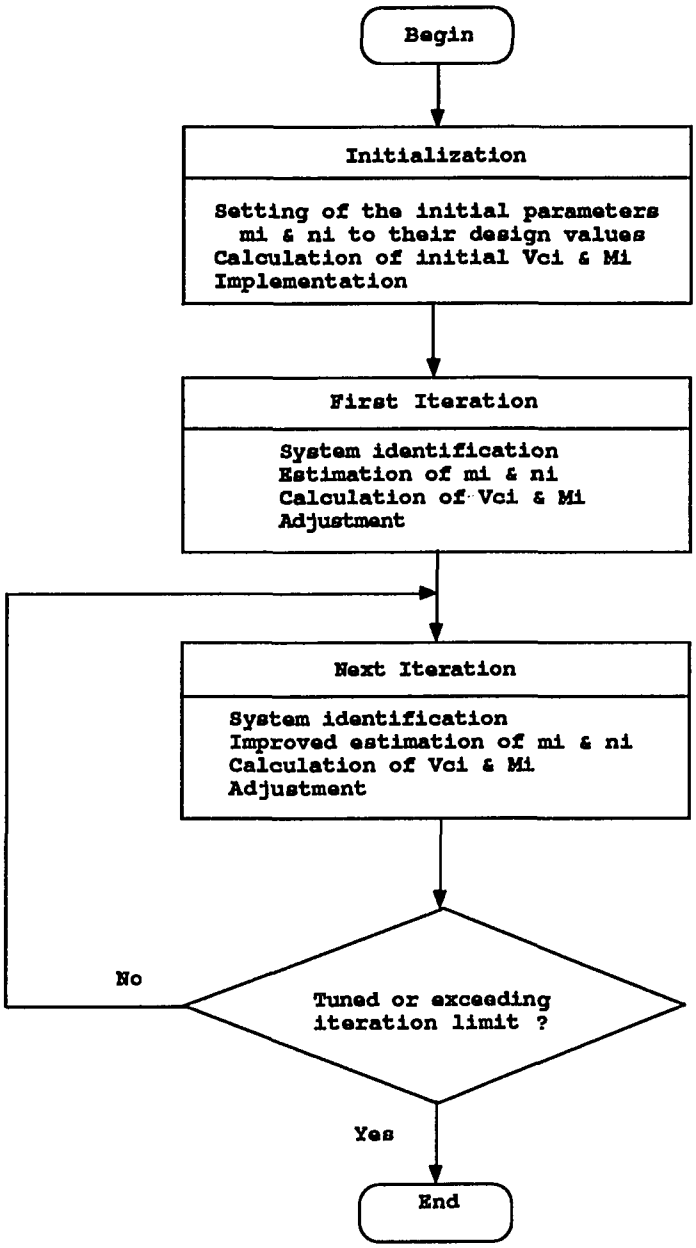


Figure 2.31: Tuning procedure flow chart

Table 2.3: Expressions for $m_i^{(k)}$ and $n_i^{(k)}$

i	$m_i^{(k)}$
1	$\frac{C[b_0^{(p1)}/(M_1^{(p1)}\sqrt{a_0^{(p1)}})-b_0^{(p2)}/(M_1^{(p2)}\sqrt{a_0^{(p2)}})]}{V_{c1}^{(p1)}-V_{c1}^{(p2)}}$
3	$\frac{C(\sqrt{a_0^{(p1)}}/M_3^{(p1)}-\sqrt{a_0^{(p2)}}/M_3^{(p2)})}{V_{c3}^{(p1)}-V_{c3}^{(p2)}}$
4	$\frac{C[(b_1^{(p1)}+\sqrt{a_0^{(p1)}})/M_4^{(p1)}-(b_1^{(p2)}+\sqrt{a_0^{(p2)}})/M_4^{(p2)}]}{V_{c4}^{(p1)}-V_{c4}^{(p2)}}$
5	$\frac{C(a_1^{(p1)}/M_5^{(p1)}-a_1^{(p2)}/M_5^{(p2)})}{V_{c5}^{(p1)}-V_{c5}^{(p2)}}$
i	$n_i^{(k)}$
1	$\frac{M_1^{(p2)}\sqrt{a_0^{(p2)}}b_0^{(p1)}V_{c1}^{(p2)}-M_1^{(p1)}\sqrt{a_0^{(p1)}}b_0^{(p2)}V_{c1}^{(p1)}}{M_1^{(p1)}\sqrt{a_0^{(p1)}}b_0^{(p2)}-M_1^{(p2)}\sqrt{a_0^{(p2)}}b_0^{(p1)}}$
3	$\frac{M_3^{(p2)}\sqrt{a_0^{(p1)}}V_{c3}^{(p2)}-M_3^{(p1)}\sqrt{a_0^{(p2)}}V_{c3}^{(p1)}}{M_3^{(p1)}\sqrt{a_0^{(p2)}}-M_3^{(p2)}\sqrt{a_0^{(p1)}}}$
4	$\frac{M_4^{(p2)}(b_1^{(p1)}+\sqrt{a_0^{(p1)}})V_{c4}^{(p2)}-M_4^{(p1)}(b_1^{(p2)}+\sqrt{a_0^{(p2)}})V_{c4}^{(p1)}}{M_4^{(p1)}(b_1^{(p2)}+\sqrt{a_0^{(p2)}})-M_4^{(p2)}(b_1^{(p1)}+\sqrt{a_0^{(p1)}})}$
5	$\frac{M_5^{(p2)}a_1^{(p1)}V_{c5}^{(p2)}-M_5^{(p1)}a_1^{(p2)}V_{c5}^{(p1)}}{M_5^{(p1)}a_1^{(p2)}-M_5^{(p2)}a_1^{(p1)}}$

The following is a detailed procedure to control g_{m5} associated with adjusting the coefficient a_1 of each biquad.

1. Set the initial control parameter g_{m5} and $C_7(=C)$ to their design values. From (2.54) and (2.55), the parameters $m_5^{(0)}$ and $n_5^{(0)}$ at the first iteration are given by

$$\begin{aligned} m_5^{(0)} &= 2\sqrt{\frac{k_{i5}^{(0)} k_{b5}^{(0)}}{3}} \\ n_5^{(0)} &= -(V_{ss} + V_{T5}^{(0)}) \end{aligned}$$

where $K_5'^{(0)}, W_5^{(0)}, L_5^{(0)}$, and $V_{T5}^{(0)}$ are the nominal values. Also, set the initial current mirror gain $M_5^{(0)}$ to a proper value so that it may not exceed the specified range in which the good linearity of the transconductance is kept. Then, from (2.4) and (2.53), the control voltage $V_{c5}^{(0)}$ at the initial implementation becomes,

$$V_{c5}^{(0)} = \frac{\bar{a_1} C}{m_5^{(0)} M_5^{(0)}} - n_5^{(0)}$$

where C and $\bar{a_1}$ are the nominal values.

2. Obtain the identified coefficient $a_1^{(0)}$ from system identification of the physical filter.

3. At the first iteration, it will be assumed that $n_5^{(1)} = n_5^{(0)}$ since we have only one equation for two unknowns. From (2.4) and (2.53),

$$a_1 = \frac{1}{C} m_5 M_5 (V_{c5} + n_5) \tag{2.56}$$

Thus, we may approximate the estimate for m_5 by

$$m_5^{(1)} = \frac{a_1^{(0)} C}{M_5^{(0)} (V_{c5}^{(0)} + n_5^{(0)})}$$

4. Calculate the control parameters V_{c5} and M_5 from

$$\begin{aligned} M_5^{(1)} &= M_5^{(0)} \\ V_{c5}^{(1)} &= \frac{\bar{a_1} C}{m_5^{(1)} M_5^{(1)}} - n_5^{(1)} \end{aligned}$$

and test if $V_{c5}^{(1)}$ exceeds the specified range. If it does, then calculate new $V_{c5}^{(1)}$ and $M_5^{(1)}$.

5. Obtain the identified coefficient $a_1^{(1)}$ from identification of the actual filter with the updated control voltages and mirror gains.

6. To obtain $m_5^{(2)}$ and $n_5^{(2)}$, observe from (refetad4) that

$$\begin{aligned} a_1^{(0)} &= \frac{1}{C} m_5 M_5^{(0)} (V_{c5}^{(0)} + n_5) \\ a_1^{(1)} &= \frac{1}{C} m_5 M_5^{(1)} (V_{c5}^{(1)} + n_5) \end{aligned}$$

Solving these two equations simultaneously, obtain the new estimates of m_5 and n_5.

$$\begin{aligned} m_5^{(2)} &= \frac{C(a_1^{(1)}/M_5^{(1)} - a_1^{(0)}/M_5^{(0)})}{V_{c5}^{(1)} - V_{c5}^{(0)}} \\ n_5^{(2)} &= \frac{M_5^{(0)} a_1^{(1)} V_{c5}^{(0)} - M_5^{(1)} a_1^{(0)} V_{c5}^{(1)}}{M_5^{(1)} a_1^{(0)} - M_5^{(0)} a_1^{(1)}} \end{aligned}$$

7. Calculate the new control parameters V_{c5} and M_5 from

$$\begin{aligned} M_5^{(2)} &= M_5^{(1)} \\ V_{c5}^{(2)} &= \frac{\bar{a_1} C}{m_5^{(2)} M_5^{(2)}} - n_5^{(2)} \end{aligned}$$

and check again $V_{c5}^{(2)}$, and obtain new $V_{c5}^{(2)}$ and $M_5^{(2)}$ if necessary.

8. Test whether the system is tuned and repeat the step 5, 6, and 7 until a tuned system is obtained.

Actually, the similar procedures for g_{m1}, g_{m3}, and g_{m5} are performed simultaneously. If the system model is ideal, this algorithm will converge after two iterations. In reality each OTA has parasitic poles and zeros which make the actual systems have over-ordered transfer functions. Thus, more iterations are needed to get better results. It can be seen that the computational requirement of this adjustment algorithm is very simple.

2.5.2 Tuning Simulation Examples

To investigate the performance of the adjustment/tuning algorithm, several sample filters were tuned by simulation. In the tuning simulation, the following nonidealities were considered: measurement error ($mn\%$), parameter variation ($p\%$), and over-ordering factor (ω_o/ω_p). The manufacturing process parameter variations were simulated with uniform random values of $\pm p\%$. Frequency domain additive measurement errors of $\pm mn\%$ with uniform distribution were directly fed to the s-domain system ID algorithm. Thus, the frequency response measurement algorithm was not used in this simulation. For every simulation, 50 noisy data obtained at equally spaced frequency points were used for the system identification. The iterative complex

least squares (ICLS) method was performed with the iteration limit set to 10. The iteration limit for tuning was also set to 10, and the filter is considered to be tuned if all the control voltages (fine adjustments) are within 9 bit accuracy with respect to the previous control voltages.

2.5.2.1 Over-ordering effects In order to evaluate the effect of the over-ordering problem on this tuning algorithm, a 6th-order elliptic lowpass filter which has a normalized cutoff frequency at 1(rad/sec) and 0.5dB passband ripple was tuned. It consists of three second-order lowpass notch (LPN) filters. Its transfer function is given by

$$T(s) = \prod_{i=1}^{3} \frac{s^2 + B_{0i}}{s^2 + A_{1i}s + A_{0i}}$$

where

i	A_{1i}	A_{0i}	B_{0i}	ω_o	Q_p
1	0.933855	0.611899	4.36790	0.7822	0.8376
2	0.156221	0.934830	1.19243	0.9669	6.1891
3	0.017576	0.990620	1.02486	0.9952	56.628

Three LPN biquads were tuned separately to tune the 6th-order elliptic filter. The tuning results with 1% measurement error and 5% parameter variation and various over-ordering factors are shown in Table 2.4, Fig. 2.32 and Fig. 2.33.

For the third LPN filter, predistortion was performed for every over-ordering case because the filter has very high design Q of 56, while for the first LPN filter no predistortion was performed due to its low Q. From the results it can be seen that this tuning algorithm converges fast and attains good accuracy in the presence of over-ordering (up to $\omega_o/\omega_p = 0.1$) effects. However, the over-ordering factor $\omega_o/\omega_p = 0.2$ leads to a relatively big ripple error at the transition region. The simulation results show that the tuning algorithm can handle very high over-ordering factors up to 0.2, so it can be well applicable to high-frequency and high-Q applications. This kind of over-ordering effects were scarcely handled in the tuning literature.

2.5.2.2 Parameter variations To evaluate the effect of measurement errors and parameter variations on this tuning algorithm, a simple second-order lowpass filter was chosen Its transfer function is

$$T_{LP}(s) = \frac{1}{s^2 + s + 1} \tag{2.57}$$

Tuning results with various parameter variations, 1% measurement error and over-ordering factor $\omega_o/\omega_p = 0.1$ are shown in Table 2.5, and Fig. 2.34. Even when 30% parameter variations were considered, the tuned filter had a maximum gain (A_{max}) within 0.35% of the desired maximum gain and a resonant frequency within 0.01% of the desired frequency. It can thus be seen that this algorithm is not sensitive to the effect of parameter variations.

2.5.2.3 Measurement errors Tuning results with various measurement errors, 5% parameter variation and over-ordering factor $\omega_o/\omega_p = 0.1$ are shown in Table 2.6 and Fig. 2.35. Even in the presence of high measurement errors (up to 5%), this tuning algorithm attained good accuracy and fast convergence rate. However, 10% measurement error resulted in a poor tuned state. Actually, in this case, the tuning algorithm had not converged but was stopped by the iteration limit of ten. This phenomenon is caused by the fact that this model-based tuning algorithm heavily depends on the results of the system identification and the accuracy of the system identification is a function of the accuracy of measurements.

2.5.3 Performance evaluation of the digital tuning scheme

The adjustment algorithm presented in Section 2.5.1 was developed for the digitally programmable continuous-time filter structure [39]. However, the tuning algorithm can be readily extendable to any kinds of digitally controllable filters. Most recently reported high-frequency continuous-time filters contain OTA-C type integrators because of their inherently better high-frequency characteristics compared with other types of integrators such as active RC or MOSFET-C integrators of which the basic building circuits are conventional operational amplifiers. Most OTA structures [13]-[23],[68, 69] have some sort of mechanism for linear control of their transconductance gain, and thus, these OTAs can be easily incorporated to build digitally programmable/tunable biquadratic structures. It has been shown in Section 2.4.5 that the biquadratic structure under test leads to degraded system ID results in the presence of significant parasitic effects because it does not offer the overall gain adjustability and small value adjustability of the constant term of the transfer function numerator polynomial.

In this section a more robust biquadratic structure is tested to evaluate the performance of the digital tuning scheme more generally. Small value adjustability of the numerator constant term and the gain adjustability of the biquad are added to the biquadratic structure under test such that 5 degrees of freedom models can be used for identification of notch biquads and 4 degrees of freedom models for lowpass or bandpass biquads. Many biquad structures satisfying these conditions may be possible. One simple example structure is depicted in Fig. 2.36. Its

Table 2.4: Tuning results of the 6th-order lowpass elliptic filter with ω_o/ω_p over-ordering factors ($mn = 1.0\%$, $p = 5.0\%$)

	LPN1	LPN2	LPN3	6th-order Elliptic Filter	
ω_o/ω_p	No. of iteration	No. of iteration	No. of iteration	Passband ripple (dB)	3dB band error (%)
0.01	2	2	2*	0.501	0.30
0.05	5	6	4*	0.634	0.34
0.1	6	4*	4*	0.556	0.32
0.2	7	7*	5*	1.357	0.60

* Predistortion was performed

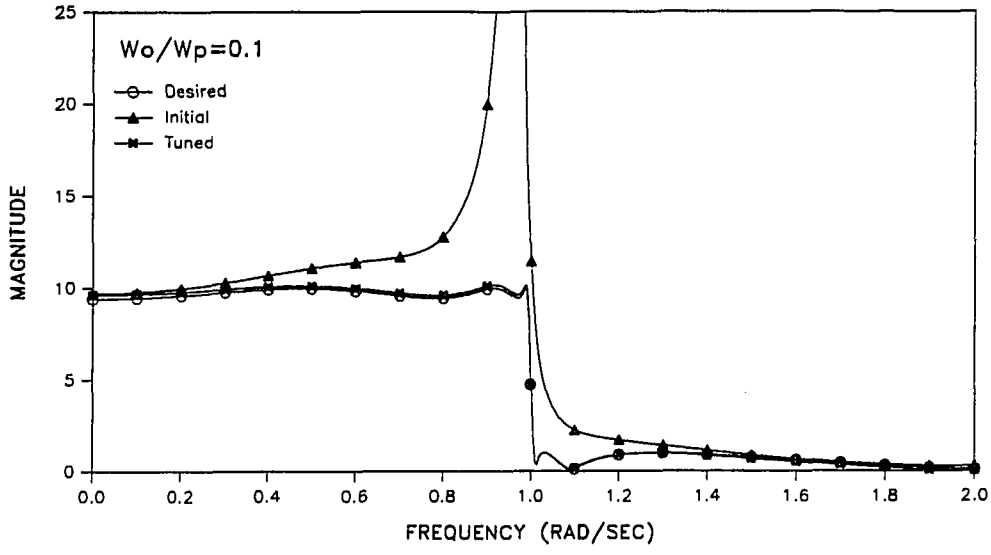
Table 2.5: Tuning results of a simple lowpass filter with $p\%$ parameter variations ($mn = 1.0\%$, $\omega_o/\omega_p = 0.1$)

Parameter Variations (p)	No. of Iteration	DC Gain A_0 (%)	Max. Gain A_{max} (%)	ω_{max} (%)	ω_o (%)
1%	7	0.005	0.27	0.13	0.10
5%	9	0.002	0.79	0.44	0.80
10%	4	0.232	0.32	0.02	0.60
20%	4	0.096	0.10	0.27	0.10
30%	6	0.937	0.35	0.72	0.01

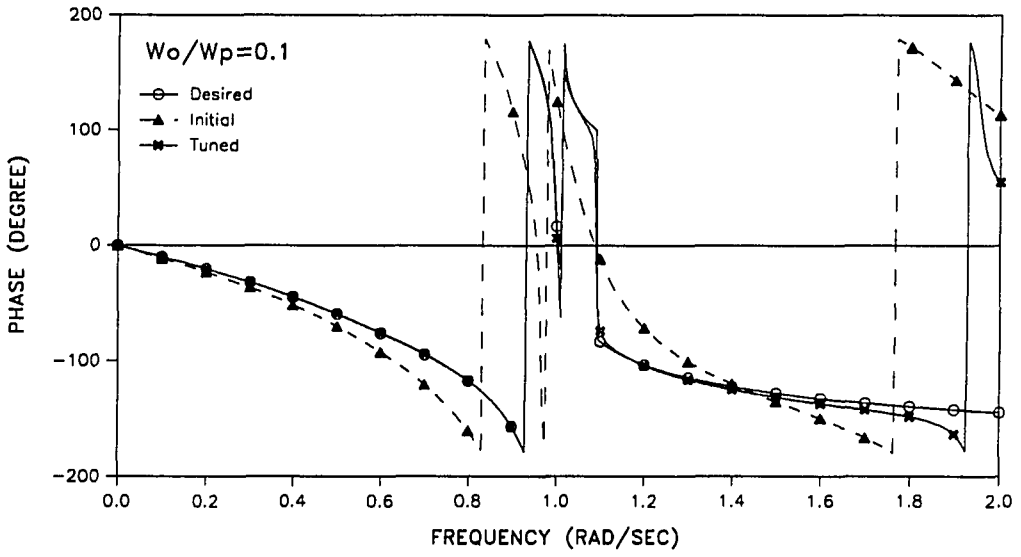
Table 2.6: Tuning results of a simple lowpass filter with $mn\%$ measurement errors ($p = 5.0\%$, $\omega_o/\omega_p = 0.1$)

Measurement Error (mn)	No. of Iteration	DC Gain A_0 (%)	Max. Gain A_{max} (%)	ω_{max} (%)	ω_o (%)
0.1%	7	0.186	0.31	0.30	0.4
1%	9	0.002	0.79	0.44	0.8
5%	7	0.331	1.18	0.30	0.4
10%	10*	0.349	5.29	1.68	0.8

* Iteration limit was exceeded

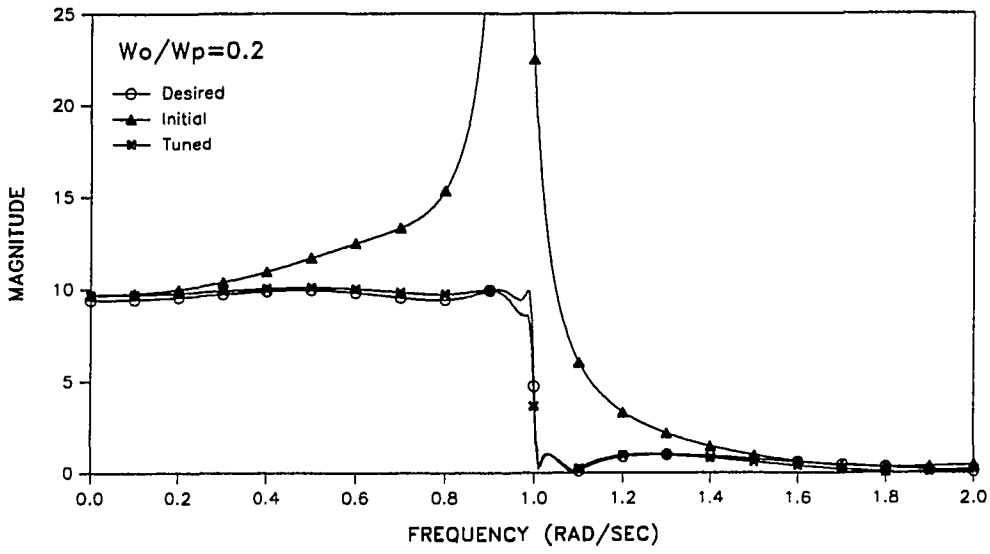


(a)

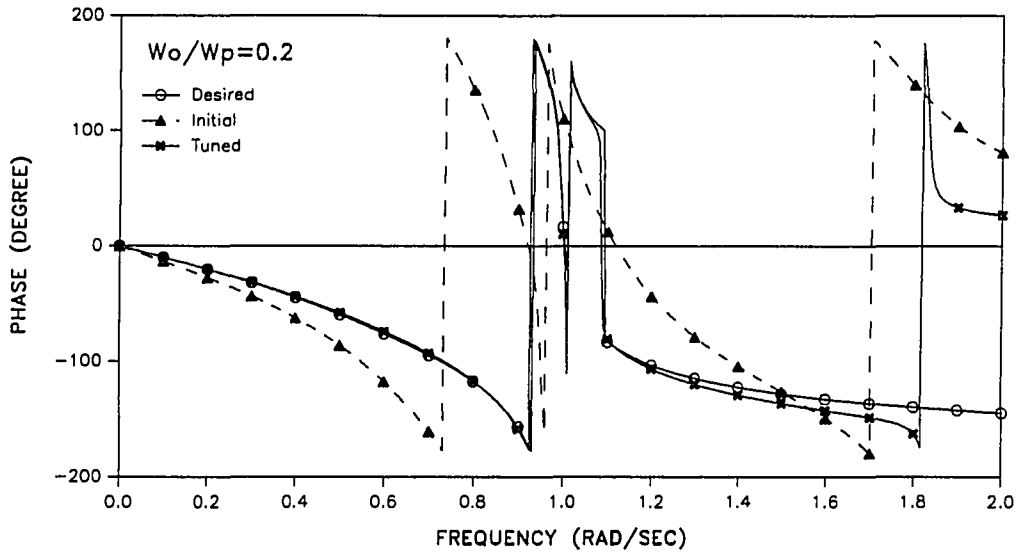


(b)

Figure 2.32: Tuning results of the 6th-order elliptic lowpass filter (a) Magnitude response (b) Phase response (Over-ordering factor $\omega_o/\omega_p = 0.1$, Measurement errors $mn = 1\%$, and Parameter variations $p = 5\%$)

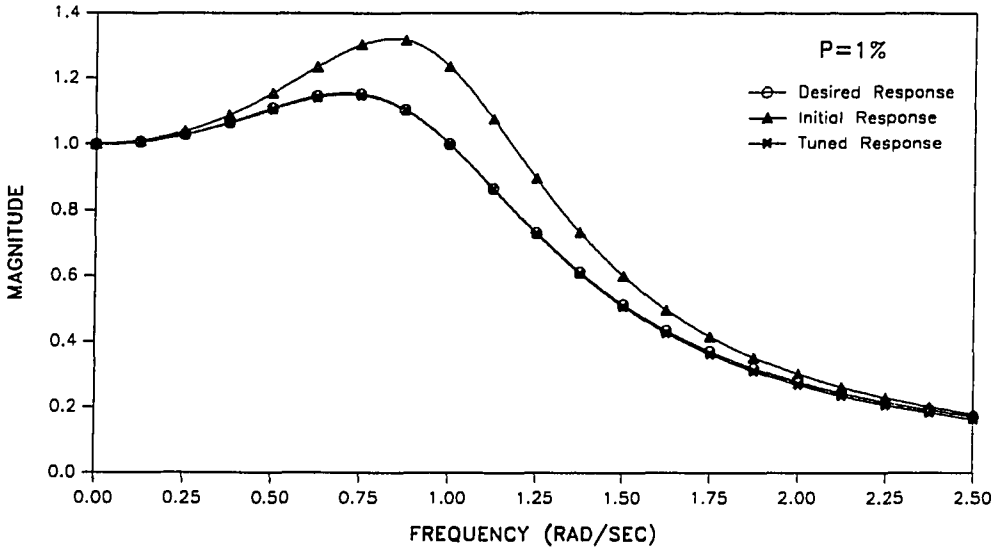


(a)

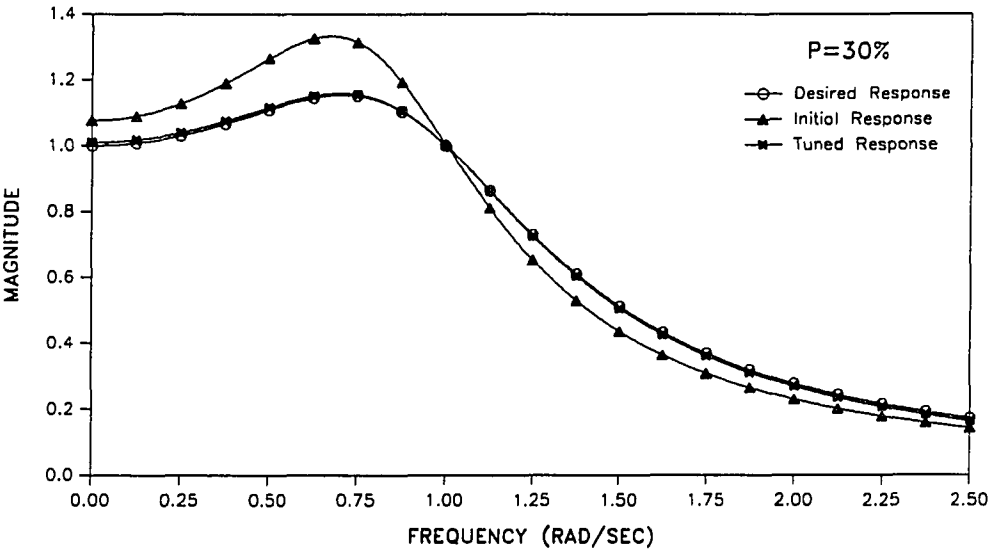


(b)

Figure 2.33: Tuning results of the 6th-order elliptic lowpass filter (a) Magnitude response (b) Phase response (Over-ordering factor $\omega_o/\omega_p = 0.2$, Measurement errors $mn = 1\%$, and Parameter variations $p = 5\%$)

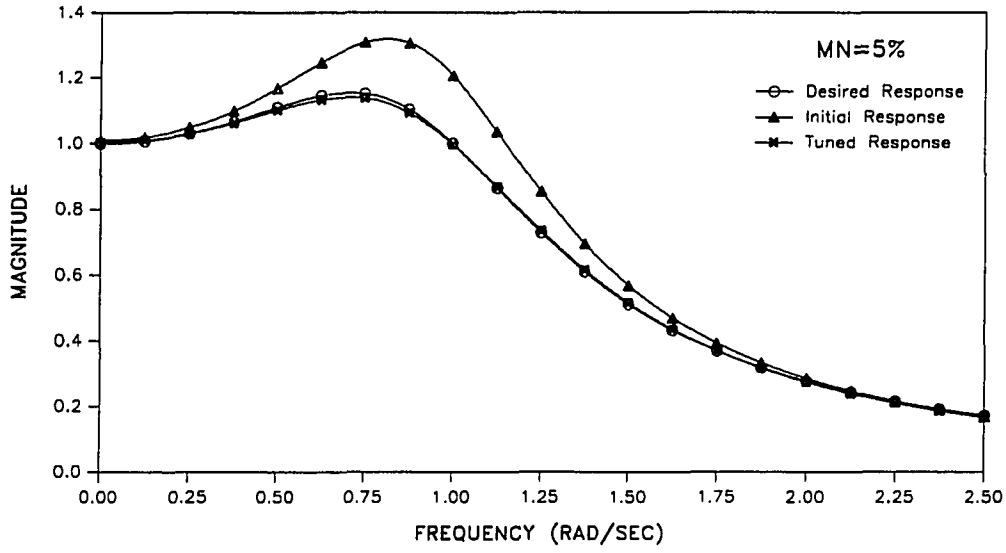


(a)

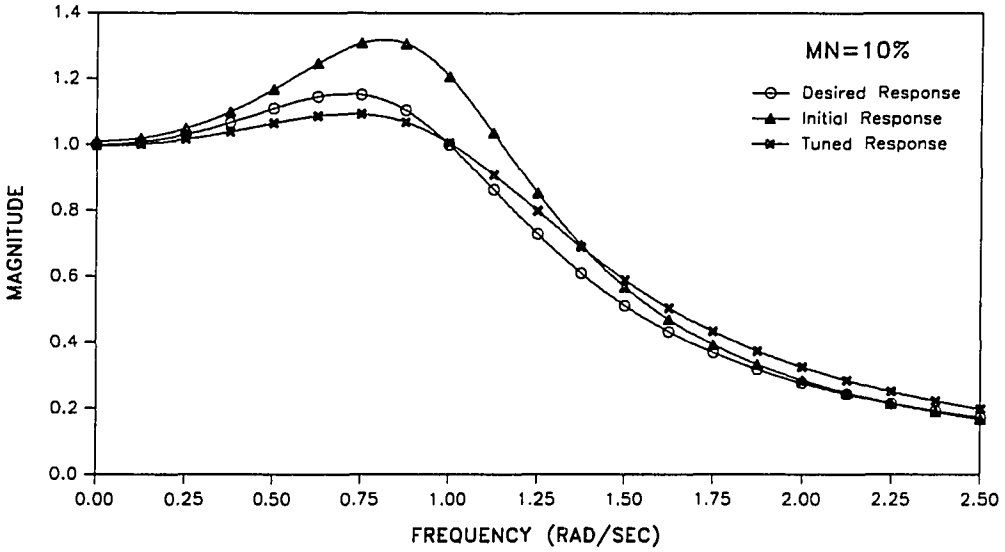


(b)

Figure 2.34: Tuning results of the simple second-order lowpass filter (a) Parameter variations $p = 1\%$ (b) Parameter variations $p = 30\%$ (Over-ordering factor $\omega_o/\omega_p = 0.1$ and Measurement errors $mn = 1\%$)



(a)



(b)

Figure 2.35: Tuning results of the simple second-order lowpass filter (a) Measurement errors $mn = 5\%$ (b) Measurement errors $mn = 10\%$ (Over-ordering factor $\omega_o/\omega_p = 0.1$ and Parameter variations $p = 5\%$)

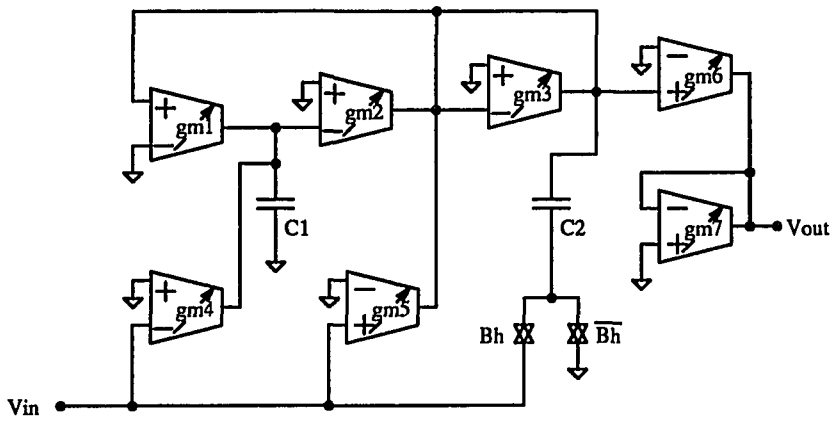


Figure 2.36: One example of the biquads which have the overall gain adjustability

ideal transfer function is given by

$$T(s) = \frac{g_{m6}}{g_{m7}} \frac{B_h s^2 + \frac{g_{m5}}{C_2} s + \frac{g_{m4} g_{m2}}{C_1 C_2}}{s^2 + \frac{g_{m3}}{C_2} s + \frac{g_{m1} g_{m2}}{C_1 C_2}}$$

Small value adjustability can be easily achieved by modifying the OTA structure. For example, the OTA structures presented in [68],[69] have even negative g_m adjustability by using cross-coupled input stage or by connecting two simple OTAs in parallel. Using this structure, a 6th-order elliptic filter is extensively simulated based upon the Monte-Carlo method. In this simulation the LS(ARMA) algorithm presented in Section 2.4.4 is also used to measure the frequency responses of the filter as shown in Fig. 2.37. Yield of tuned filters is also investigated to fully characterize the tuning scheme.

2.5.3.1 Test circuit (6th-order elliptic lowpass filter) To test the whole tuning scheme shown in Fig. 2.37, a 6th-order elliptic lowpass filter has been chosen which has a normalized cutoff frequency of 1.0 (rad/sec), a passband ripple of 1.0 (dB), a stopband starting frequency of 1.5 (rad/sec), and a minimum stopband attenuation of 64.66 (dB). Its magnitude response is shown in Fig. 2.38.

The 6th-order elliptic function

$$T(s) = K \frac{s^4 + \ldots + b_1 s + b_0}{s^6 + \ldots + a_1 s + a_0}$$

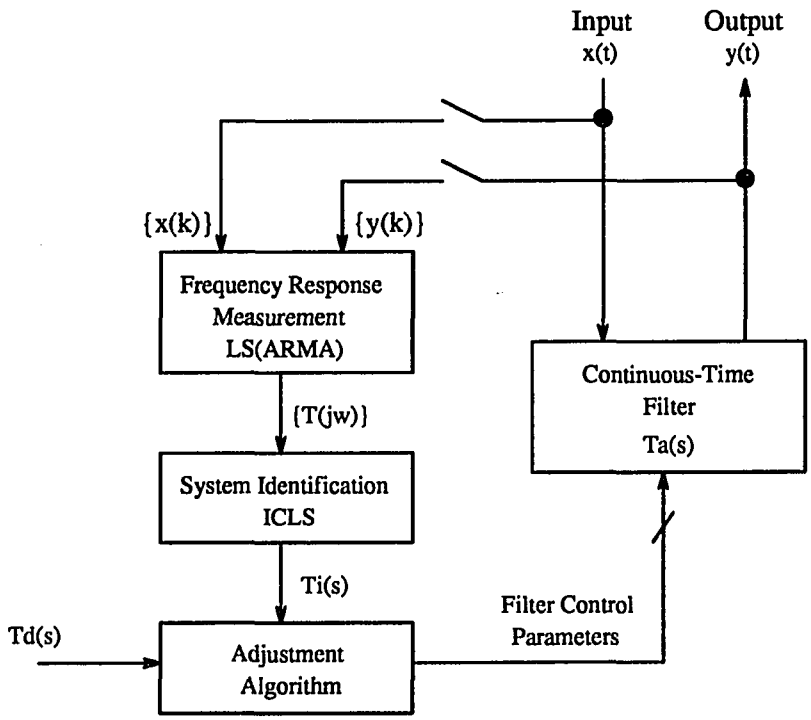
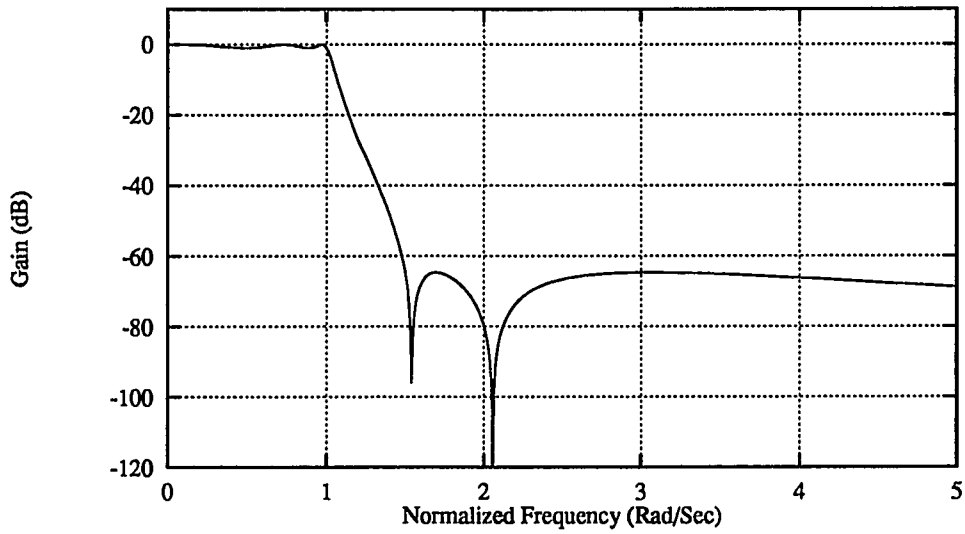
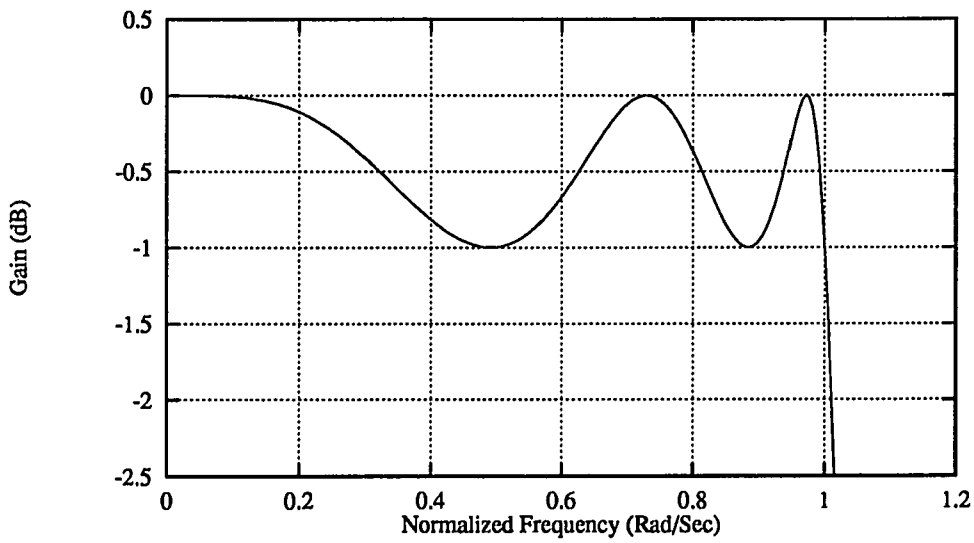


Figure 2.37: Block diagram illustrating the whole tuning scheme



(a)



(b)

Figure 2.38: (a) Magnitude response of the 6th-order elliptic lowpass filter (b) Magnified passband plot

must be decomposed into three 2nd-order biquadratic functions

$$T(s) = \prod_{i=1}^{3} K_i \frac{B_{hi}s^2 + b_{1i}s + b_{0i}}{s^2 + a_{1i}s + a_{0i}}.$$

According to the network decomposition theory [70], there are many different ways to do the decomposition depending upon the following degrees of freedom:

- Pole-zero pairing
- Gain distribution
- Cascading sequence,

and two possible criteria are

- Maximum dynamic range and minimum inband loss
- Minimum overall transmission sensitivity

It is not possible to perform the decomposition such that both criteria are simultaneously satisfied. The first condition is thus selected since it is considered more important from a tuning point of view. To satisfy the first condition, voltage swing at each biquad input is as high as possible while the in-band losses are as low as possible.

There are 6 possible combinations in pole-zero paring of the 6th-order function. Poles and zeros should be paired such that the magnitude response of each biquad is as flat as possible in the frequency range of interest. According to Lueder's method, poles and zeros should be paired such that

$$max\{d_i\}_{i=1,2,3}$$

is minimized where

$$d_i = log\left(\frac{T_{imax}}{T_{imin}}\right)$$

where T_{imax} is maximum gain of biquad i for $\omega \in [0, \infty]$, and T_{imin} is minimum gain of biquad i for $\omega \in [0, 1]$. The pole-zero paring satisfying the above condition can be found using d-table and d-graph as in [70]. In general the rule of thumb is to combine the high-Q poles with the zeros lying closest to them. The pole-zero locations of the 6th-order elliptic filter are depicted in Fig. 2.39. After applying the Leuder's method, we obtained the following pole-zero pairing: $p_1 - z_\infty$, $p_2 - z_2$, and $p_3 - z_1$ which obeys the rule of thumb.

Gain distribution and biquad sequence determination are also done from the viewpoints of minimizing individual biquad overdrive and maximizing the signal to noise ratio. The methods

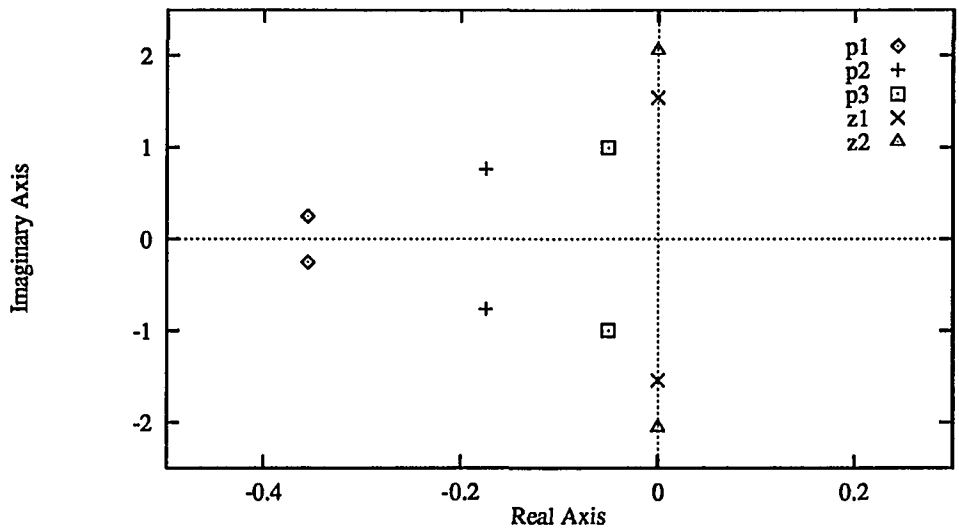


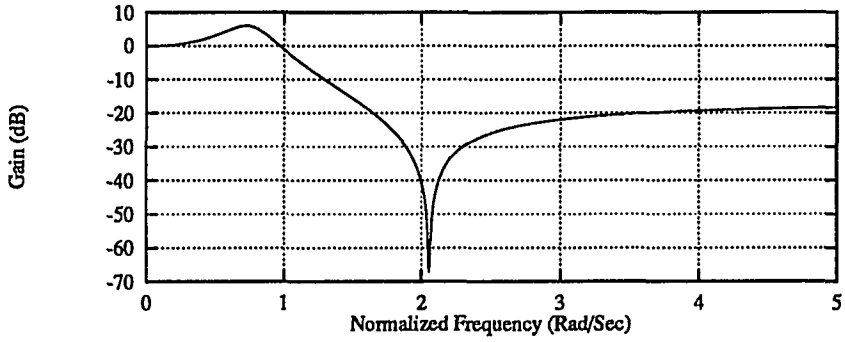
Figure 2.39: Pole-zero locations of the 6th-order elliptic lowpass filter

for these can be also found in [70]. After these procedure, the finally obtained biquad transfer function coefficients are

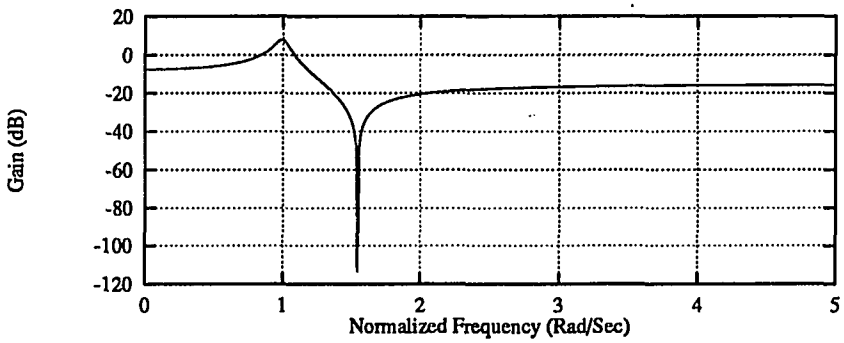
i	K_i	B_{hi}	b_{1i}	b_{0i}	b_{1i}	b_{oi}	ω_{oi}	Q_i
1	0.14218	1	0	4.230449	0.348604	0.611825	0.7822	2.2438
2	0.17511	1	0	2.381154	0.100900	0.994942	0.9975	9.8857
3	0.45618	0	0	1.0	0.710041	0.187943	0.4335	0.6106

where it can be seen that the filter consists of two lowpass notch biquads and one lowpass biquad. Their magnitude responses are shown in Fig. 2.40. This kind of dynamic range optimized filter function can not be implemented with the digitally programmable continuous-time filter under test due to unadjustability of the biquad gains K_i's.

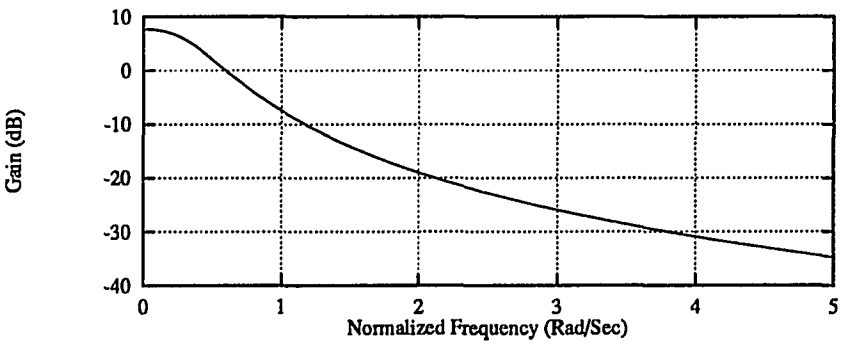
2.5.3.2 Simulation results and tuning yield investigation The 6th-order elliptic lowpass filters have been tuned using the procedure shown in Fig. 2.37. The mean and standard deviation of gain error computed from 100 sample filters before tuning and after tuning are depicted in Fig. 2.41 (a), (b), (c) and (d), respectively, for three different over-ordering factors.



(a)

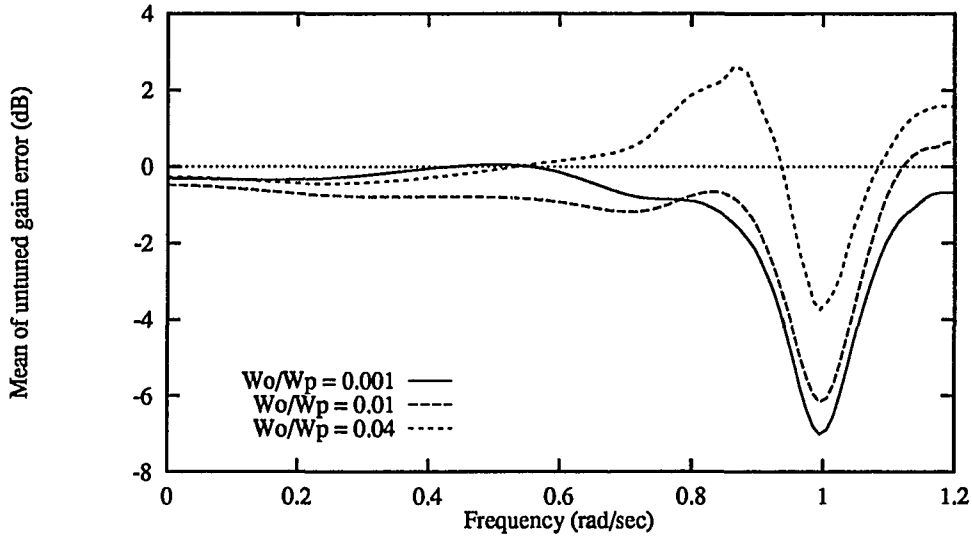


(b)

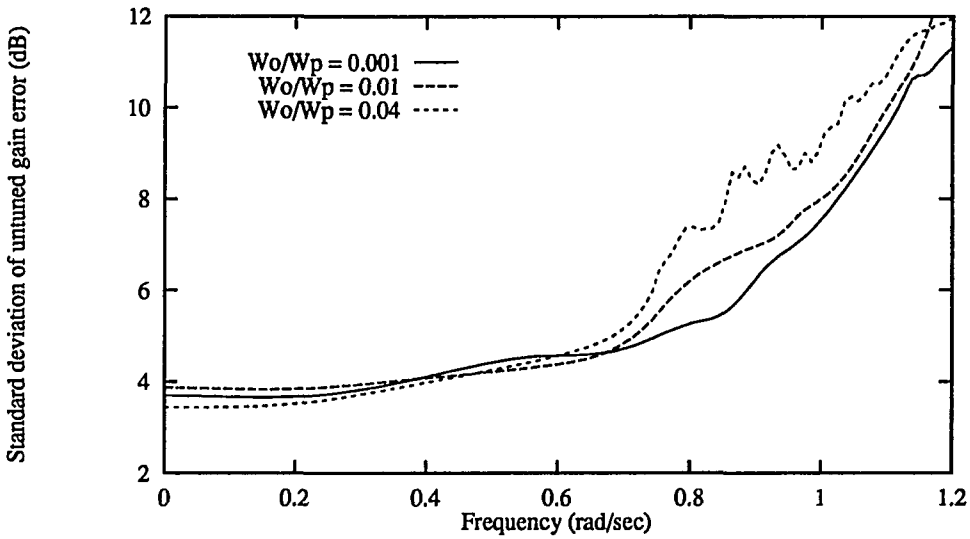


(c)

Figure 2.40: Dynamic range optimized 6th-order elliptic lowpass filter. Magnitude responses (a) Biquad1 (b) Biquad2 (c) Biquad3

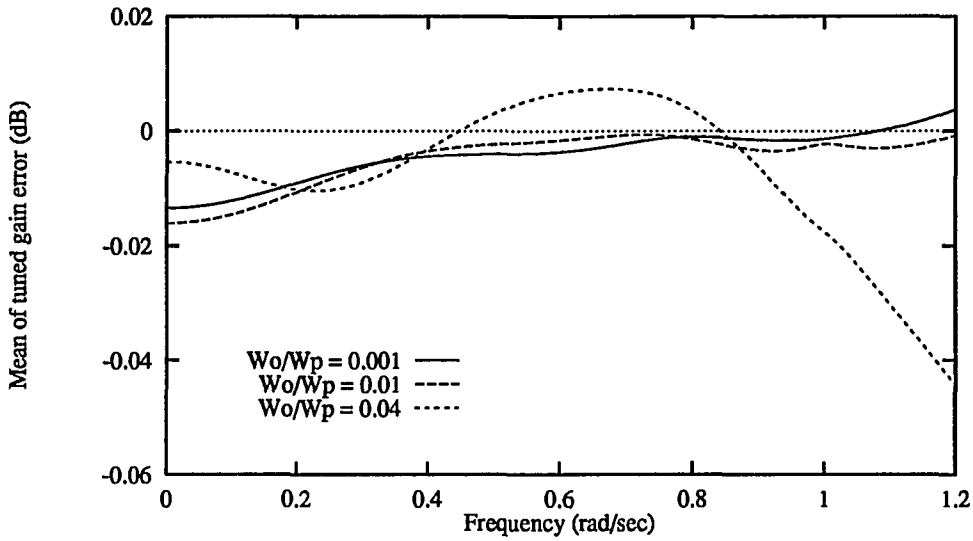


(a)

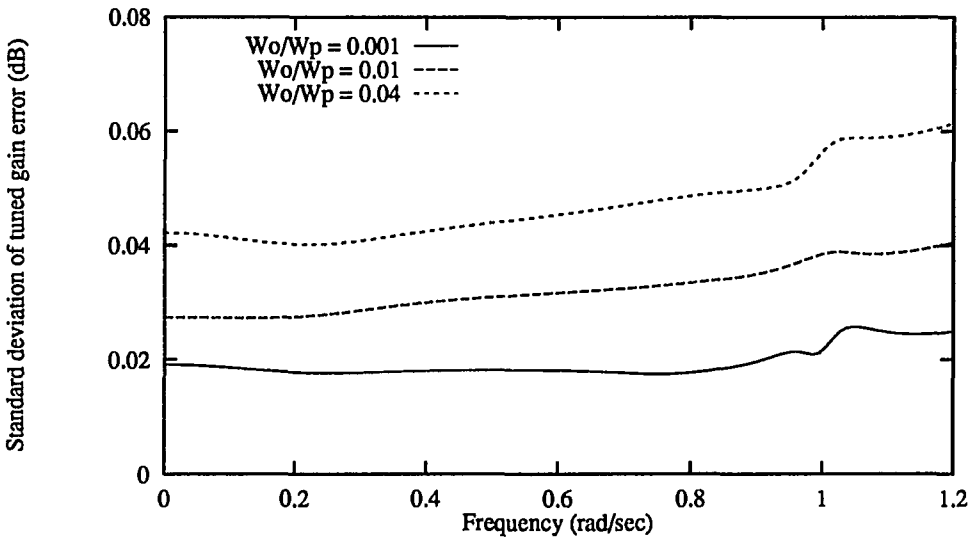


(b)

Figure 2.41: Effects of the over-ordering factor (ω_o/ω_p) on the tuning accuracy. (a) Mean and (b) Standard deviation of untuned gain error computed from 100 independent untuned and tuned filters (SNR=40dB, THD=−40dB, $p = 20\%$)



(c)



(d)

Figure 2.41: (continued) (c) Mean and (d) Standard deviation of tuned gain error

In this simulation, somewhat pessimistic conditions were chosen to simulate worse cases than realistic. The following nonideality factors were used.

- Component variation p=20%
- Signal-to-noise ratio SNR=40dB
- Total harmonic distortion THD=−40dB

The process variations on the capacitances and the OTA transconductance gains were simulated with 20% multiplicative uniform random numbers. The SNR is the ratio of the input excitation signal to additive uniform random noise as defined in Section 2.4.4.4. The input and output signal samples were added to uniformly distributed random numbers. The nonlinearity of input excitation signals and the filter nonlinearity were approximated with the second harmonic distortion. The total harmonic distortion of the input and out signals was set to −40dB. The iteration limit for tuning was set to 20, and the filter is considered to be tuned if all the control voltages are within 10 bit accuracy.

The results indicate that the digital tuning scheme can reduce the standard deviation of gain error by a factor of 100. It should be noted that the standard deviation increases with the over-ordering factor. This is primarily due to the degraded system ID accuracy in the presence of significant parasitics as discussed in earlier sections. The scheme, however, still shows good accuracy with an over-ordering factor (ω_o/ω_p) as high as 0.04. The plots shown in Fig. 2.41 were computed from individual frequency points, and thus, the mean of gain error plot is just the ensemble average of the gain responses of 100 sample filters. With these plots, it is not possible to fully investigate the statistical characteristics of the tuned filters.

To evaluate the performance of the tuning scheme more accurately and to investigate the tuning yield, the window specifications shown in Fig. 2.42 are considered. Fig. 2.42 (a) exhibits the window specification of the ideal 6th-order elliptic lowpass filter. If an error bound is given as shown in Fig. 2.42 (b) such that a sample filter which has a gain response within the error bound is regarded as a tuned or satisfiable filter, then histograms of the tuned filters can be obtained. The error bound e in the passband is defined as a percentage with respect to the passband ripple (1.0dB) as shown in Fig. 2.42. Thus, the filters of which the gain responses in the passband are between $e/100$ (dB) and $-(1+e/100)$ (dB) are considered to have an error bound e in the passband. The error bound is defined differently in the stopband as a percentage with respect to the stopband attenuation (64.66dB) divided by 10. Thus, the filters which have stopband gains less than $-64.66+6.466e/100$ (dB) have an error bound e in the stopband.

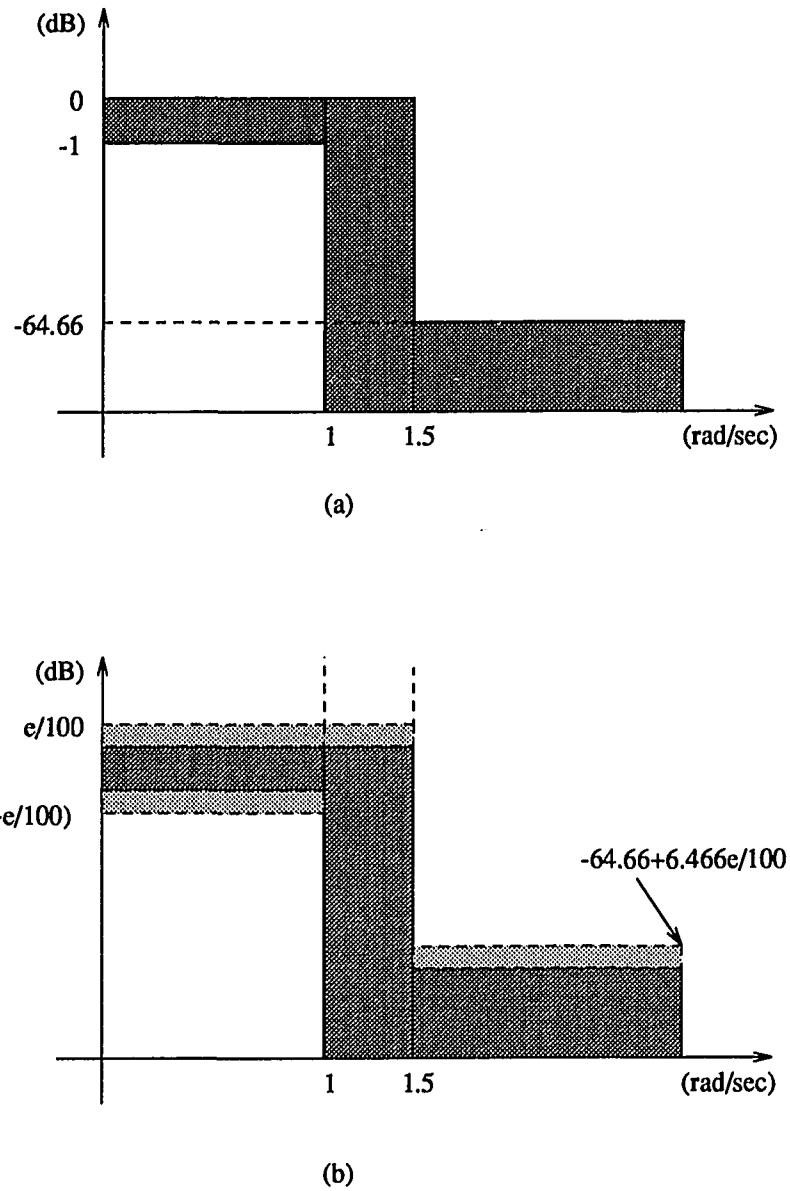


Figure 2.42: (a) Ideal window specification of the 6th-order elliptic lowpass filter (b) Definition of the window error bound

The obtained histograms of the tuned filters are shown in Fig. 2.43, 2.44, and 2.45 for three different over-ordering factors. Each figure contains three histogram plots. The first and second plots were obtained by checking the passband error bound only and the stopband only, respectively. The last one is the histogram taking into account the entire frequency range. From these plots we can more clearly analyze the statistical characteristics of the filters tuned by the digital tuning scheme. If the allowable error bound e is set to 10, i.e., if the allowable passband ripple deviation from the ideal ripple (1.0dB) is 0.1dB and the allowable stopband deviation from the ideal stopband attenuation (64.66dB) is 0.6466dB, then the tuning yields are 100%, 99%, and 92% for $\omega_o/\omega_p = 0.001$, 0.01, and 0.04, respectively. This indicates that if a tighter one than the desired window specification is used for initial filter implementation and tuning, then most tuned filters are expected to satisfy the desired window specification. It can be also seen from the plots that the error bound density looks similar to a Rayleigh or a Gamma distribution.

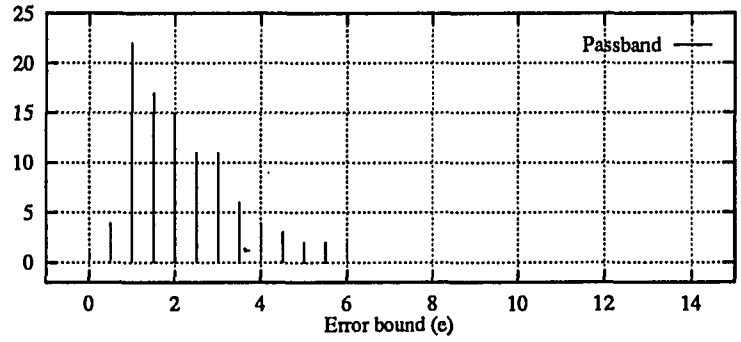
2.5.4 Tuning Experimental Results

The digital tuning scheme has been applied to tune several sample filters. The linear transfer functions were implemented with the digitally programmable monolithic continuous-time filter discussed in Section 2.3 which has only 6 bit resolution for the fine control. The block diagram of the experimental setup is shown in Fig. 2.46.

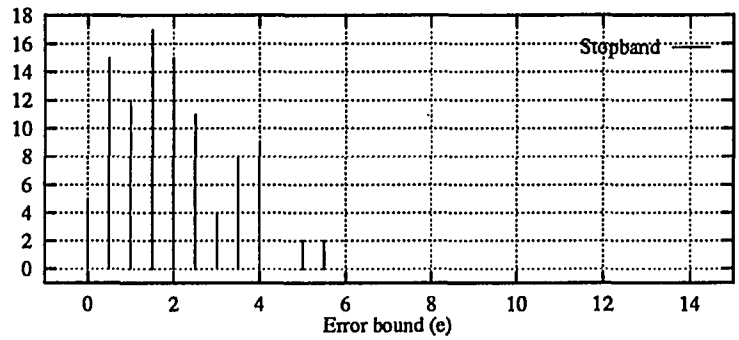
A workstation HP 9000/300 was used as the tuning host, and all instruments were connected on the HP-IB and were controlled by the tuning host. Measurements were made by the HP 54111D digitizing oscilloscope, which has programmable built-in commands for automatic measurements and has 6 bit single-shot accuracy and 8 bit accuracy with averaging. Excitation signals were generated from a HP 3325A programmable function generator.

First, a simple 2nd-order lowpass filter which has a resonant frequency of 500 kHz was implemented and tuned. Its normalized transfer function is given in (2.57). Gain and phase responses were measured at 50 equally spaced frequency points from dc to 600 kHz for each iteration and used for system identification. Fig. 2.47 shows that the tuned filter has a frequency response close to the desired one while the initially implemented filter has an erroneous frequency response. The entire tuning process took 9 iterations.

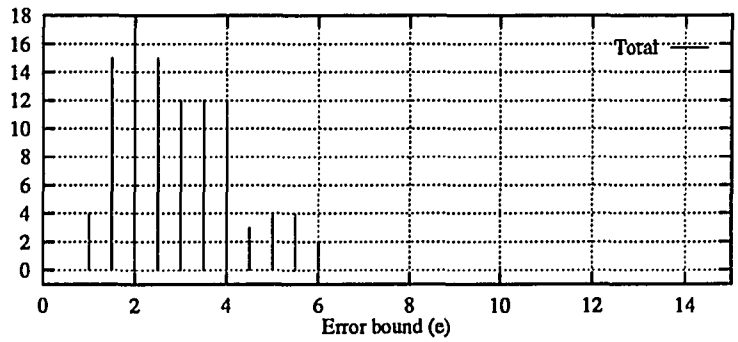
Another tuning experimental result is shown in Fig. 2.48. The filter was tuned to a 2nd-order bandpass filter which has a resonant frequency of 100 KHz, a Q of 10 and a maximum gain of 1. After 7 iterations, the tuned filter had a resonant frequency of 99.7 KHz, a Q of 9.97 and a maximum gain of 0.995. These data were calculated from the identified transfer function



(a)

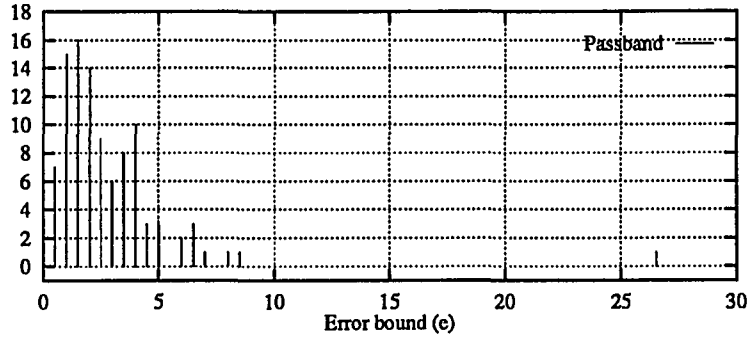


(b)

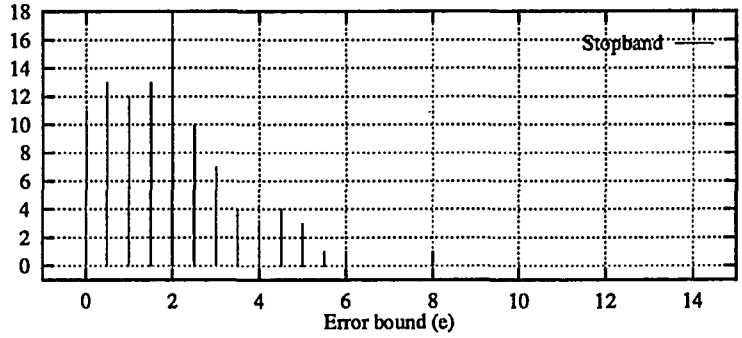


(c)

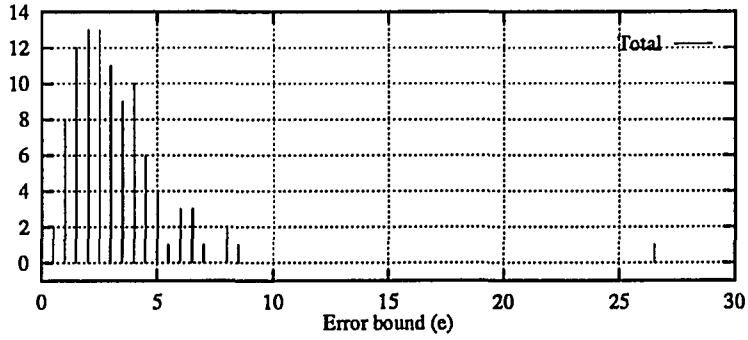
Figure 2.43: Histograms of 100 tuned 6th-order elliptic lowpass filters for $\omega_o/\omega_p = 0.001$ (a) Passband, (b) Stopband, and (c) Entire range



(a)

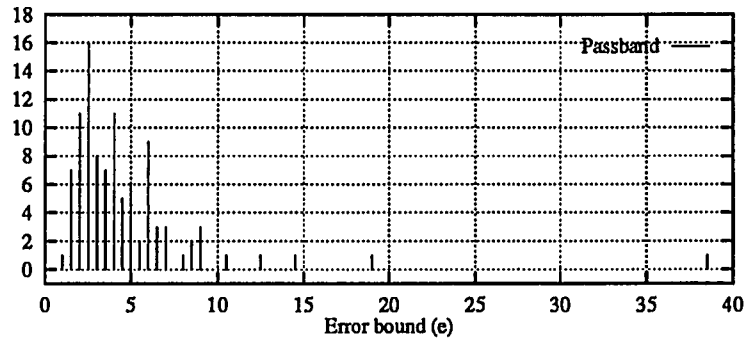


(b)

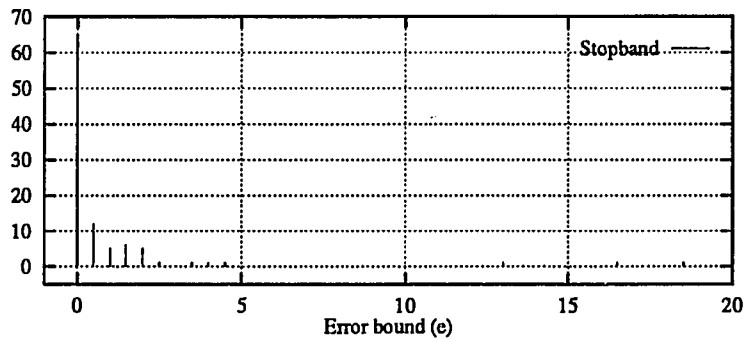


(c)

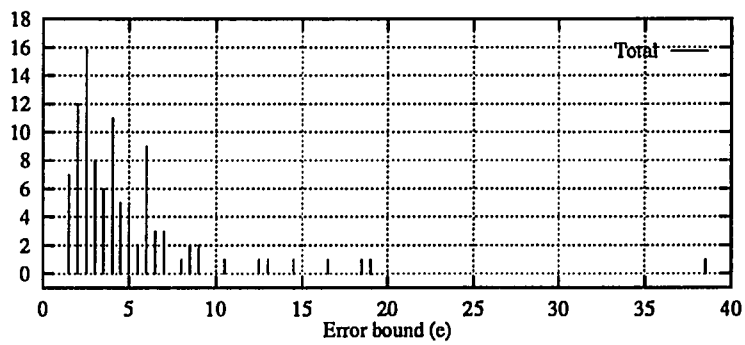
Figure 2.44: Histograms of 100 tuned 6th-order elliptic lowpass filters for $\omega_o/\omega_p = 0.01$ (a) Passband, (b) Stopband, and (c) Entire range



(a)



(b)



(c)

Figure 2.45: Histograms of 100 tuned 6th-order elliptic lowpass filters for $\omega_o/\omega_p = 0.04$ (a) Passband, (b) Stopband, and (c) Entire range

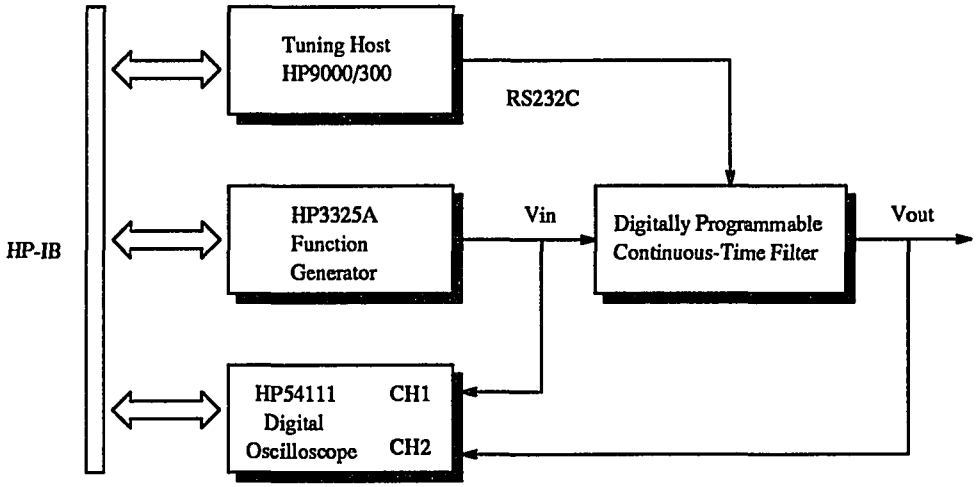


Figure 2.46: Block diagram of the tuning experimental setup

of the tuned filter.

2.6 Conclusions

In this chapter a digital tuning scheme for digitally programmable continuous-time filters has been described. To simplify the tuning problem, the tuning procedure is partitioned into two phases: system identification and adjustment. Various methods for continuous-time filter identification have been discussed. Two indirect methods, time-domain approaches and frequency-domain approaches have been investigated where the system ID problem is simplified by decomposing it into two steps.

In the time-domain approaches, the continuous-time filters are identified by first estimating discrete-time models using z-domain system ID algorithms and then obtaining equivalent continuous-time models using z-to-s transformation methods. It has been demonstrated through extensive simulations that among various LS z-domain methods, the generalized LS algorithm based on an AR noise model shows better performance for various noise characteristics than any others.

Very accurate domain transformation (s-to-z and z-to-s) methods based on the iterative complex LS algorithm have been presented and compared with the well known bilinear trans-

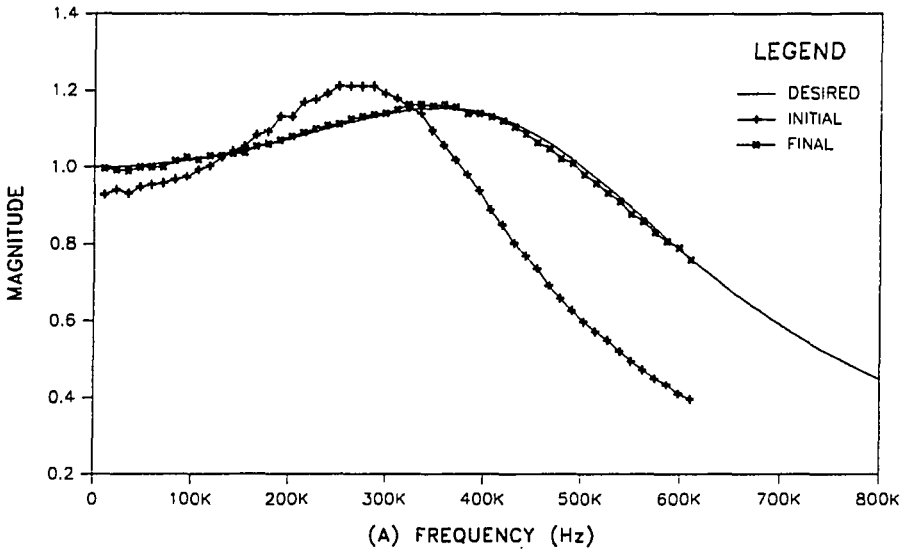


Figure 2.47: Experimental tuning result of a simple second-order lowpass filter: magnitude response measured from the HP 54111D

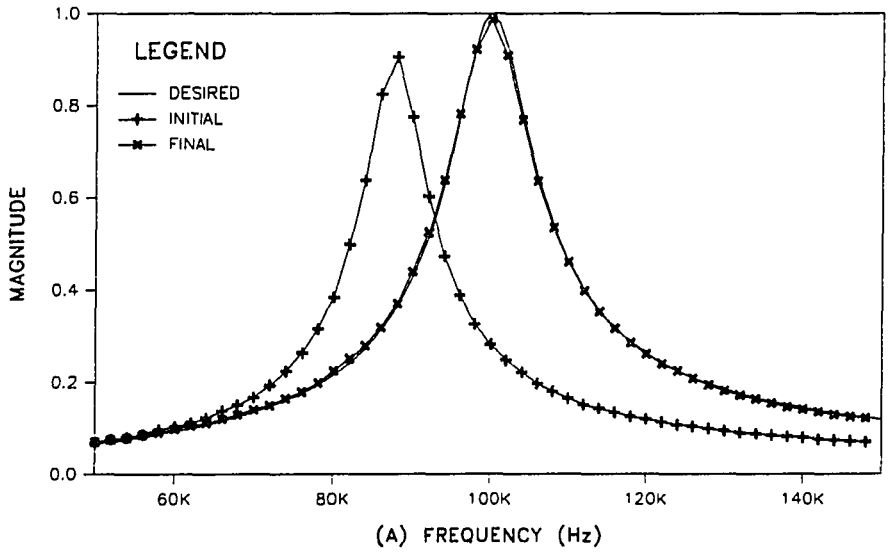


Figure 2.48: Experimental tuning result of a second-order bandpass filter (Q=10): magnitude response measured from the HP 54111D

formation method. It has been shown that both complex LS transformation methods provide improvement in accuracy by a factor of 10 to 100 compared with the bilinear method. The complex LS s-to-z method can thus be well applied with improved accuracy to the digital filter and control system design applications. Although the time-domain approach, i.e., the GLS(AR) algorithm along with the complex z-to-s transformation method, can be well applied to the continuous-time filter ID, it requires high-performance and high-cost data acquisition circuits for high-frequency applications due to the requirement of large number of consecutive samples, which makes the frequency-domain approaches preferable in the digital tuning.

In the frequency-domain approaches, the frequency responses of the filter to be identified are first measured from frequency response measurement algorithms and the measured data are then fed to s-domain system ID algorithms. Frequency response measurement methods based upon the FFT algorithm and the LS algorithms have been comparatively investigated. It has been shown that the LS algorithms based on low-order models such as a 1st-order AR model and a 1st-order ARMA model have similar performance to the FFT method while they can be utilized with lower-cost data acquisition circuits. Several s-domain system ID algorithms have been presented. It has been demonstrated that the iterative complex LS (ICLS) algorithm can reduce the bias existing in the ordinary complex LS algorithm. The combination of the ICLS s-domain system ID algorithm and the LS(ARMA) frequency response measurement algorithm can thus become a good frequency-domain approach for continuous-time system ID. Since actual filters are usually over-ordered due to parasitic effects, system ID should be robust in the presence of parasitics. It has been shown that the requirement can be decently satisfied by the frequency-domain approach.

An adjustment algorithm tailored to the digitally programmable continuous-time filter structure under test has been proposed. Its basic idea is to calculate filter control parameters by estimating process dependent parameters using the system ID results. It is very simple and converges quickly. Extensive simulations demonstrated that the adjustment algorithm along with the system ID method can attain very good accuracy and high convergence rates for low-frequency applications. It has also shown that the digital tuning scheme can be fairly well applicable to high-frequency and high-Q applications. Experimental results have demonstrated that the tuning scheme can be successfully applied to filter tuning with good accuracy.

Since the adjustment algorithm completely relies on the system ID results, the tuning performance is highly affected by the system ID accuracy. Although the proposed system ID method is robust to some extent in the presence of high parasitic effects, its performance will be limited when the parasitic effects are getting more significant. This is primarily due to

that the s-domain system ID algorithm is a model-based one and the performance degradation stemming from model errors are inevitable. The extent of performance degradation due to model error is also strongly related to the biquad structure and the OTA transconductance gain adjustability. If a biquad has lack of adjustability as in the biquad structure discussed in Section 2.3, a reduced degrees-of-freedom model should be used for system ID which leads to increased model error. These kinds of model errors can be considerably reduced by using higher-order models for system ID. This will, however, lead to much complicated procedures for filter adjustment.

The performance of the digital tuning scheme is degraded when the over-ordering effects of the actual filter are very significant. However, simulation results have demonstrated that the tuning scheme can get considerably good accuracy in the presence of high over-ordering effects. Tuning simulation results of a 6th-order lowpass filter have shown that the tuning yields for 0.1dB (about 1% error) passband error bound are 100% and 95% for over-ordering factors 0.001 and 0.04, respectively. These results were obtained with 10 bit resolution of the control voltages (10 bit D/A converters) in somewhat pessimistic environment such as 20% process component variation, 40dB SNR which requires a medium resolution data acquisition system (an 8 bit A/D converter), and -40dB harmonic distorted input signals which can be generated from an inexpensive sinusoidal signal generator circuit. Thus, the tuning scheme can be used to practically build high-performance monolithic continuous-time filters.

The parasitic effects can also be reduced by circuit techniques using excess phase compensation schemes and using very simple OTA structures where the parasitic poles and zeros lies at very high frequencies leading to reduced over-ordering effects. The performance of the digital tuning is limited by the accuracy of the excitation and data acquisition circuits and the resolution of the control circuit. This, however, implies paradoxically that the digital tuning has potential of very high precision at the expense of high cost.

When the digital tuning scheme is applied to high-frequency and high-Q applications, the predistortion technique discussed in Section 2.3 must be used at initial implementation to avoid oscillation. Since the technique using predistortion based on the estimation of the effective parasitic pole does not always result in stable initial implementation and the current digital tuning scheme does not guarantee the stability, one possible future work would be the development of methods such that the stability is guaranteed.

CHAPTER 3. NONIDEALITY CONSIDERATION FOR HIGH PRECISION AMPLIFIERS – ANALYSIS OF RANDOM COMMON-MODE REJECTION RATIO[1]

3.1 Introduction

Numerous nonideal effects impact and generally degrade the performance of practical op-amps. Three factors, finite gain, finite common-mode rejection ratio (CMRR), and nonzero offset, are the major sources which limit the high-precision low-frequency applications of amplifiers. It is well known that precision applications require a high open-loop gain, a large common-mode rejection ratio and a low offset voltage but practical limitations force the designer to make tradeoff between these parameters. Because of the nonlinear relationship between these parameters and the performance parameters of interest, and because of the inherent statistical nature of the offset voltage and CMRR, the relationship between these parameters and the performance of amplifiers is still not fully formulated, causing designers to still commit non-optimal designs to the foundry. For example, an infinite CMRR is often not optimal in the presence of a known finite open-loop gain of the op-amp. This research focuses on a rigorous formulation of the relationship between these parameters and the performance of precision finite-gain amplifiers. Simple mathematically tractable relationships between the finite gain, CMRR and offset voltage are developed and related to the overall performance of high precision finite gain amplifiers.

The CMRR and offset are not totally deterministic but have both deterministic and random components. Unfortunately, the performance and yield of systems using integrated op-amps are often dominated by the random components. These random components which are primarily due to the device mismatch make it difficult to analyze the op-amp errors. The statistical characteristics of these parameters must be well understood to practically obtain high precision

[1]

performance. Several analyses of the random offset [71],[72] and the random CMRR [74]-[76] in differential amplifiers have been made, but these analyses do not focus on the mixed effects of these nonidealities on amplifier performance. The analyses of the random CMRR [74]-[76], made several decades ago, concentrated only on bipolar differential amplifiers. Moreover, they focused on the methods to increase the CMRR, not on the statistical characteristics of this parameter which play a key role in the performance of precision finite gain amplifiers.

The impact of the CMRR may be best appreciated by reviewing the term itself. The term is widely used and has appeared in elementary electronics and instrumentation texts for many years [11],[71]-[73]. For a single sample amplifier with differential input and single-ended output, the term is defined as

$$CMRR = \left|\frac{A_{dm}}{A_{cm}}\right| \tag{3.1}$$

where A_{dm} and A_{cm} are the small signal differential-mode and common-mode gains respectively. Often it is expressed logarithmically rather than linearly. For the single sample amplifier, the CMRR is deterministic and can be readily measured in the laboratory. Of more importance than the CMRR of a single sample amplifier from an operational amplifier yield viewpoint, from a discrete systems designers viewpoint, and from an integrated systems designers viewpoint, is the CMRR of an amplifier architecture in a process. In this case, the common-mode gain which is ideally zero, becomes a key parameter in determining the CMRR. Since the common-mode gain invariably has a random component and a deterministic component, the same comment can be made about the CMRR.

Unfortunately, a rigorous definition of the CMRR has not appeared in the literature. Consequently, designers have been basing designs on inaccurate models and/or expensive "worst case" simulations where it is often difficult to ascertain that the simulations are actually worst case. The impact has often resulted in designs that are overly conservative or designs that have substantially degraded performance. The rigorous definition of the CMRR, though seemingly straightforward, is complicated by the observation that the CMRR is actually a random variable that is ideally infinite and that has a probability density function. The probability density function of the CMRR is nonlinearly related to the probability density functions of several other random variables which characterize the transistors comprising the operational amplifiers.

In this chapter, the CMRR and offset of CMOS op-amps are thoroughly investigated. Op-amp induced errors in precision finite gain amplifiers due to these nonideal effects are compositely analyzed. A model amplifier for these analyses is the two-stage CMOS op-amp shown in Fig. 3.1. The sample op-amp has been designed for high-speed and high-precision applications in a 2μ CMOS process. The device sizes and other performance parameters are

Table 3.1: Transistor size of the op-amp in Fig. 3.1

Transistor	W/L (μm/μm)	Transistor	W/L (μm/μm)
M1	204/2	M2	204/2
M3	75/3	M4	75/3
M5	336/3	M6	100/2
M7	250/2	M8	14/4
V_B	3.3 V	C_C	2.39 pF

Table 3.2: Performance of the op-amp in Fig. 3.1

Specification	Performance
Settling Time (1V Step, 0.1%)	18.3 nS
(2V Step, 5mV)	16.5 nS
Systematic Input Offset Voltage	0.26 mV
Open Loop Voltage Gain	819.4 (58.27 dB)
Unit Gain Frequency (GB)	59 MHz
Phase Margin	75°
Output Voltage Swing	+4.1V, −4.3V
Power Dissipation	16.5 mW
CMRR	62.5 dB

shown in Table 3.1 and 3.2. Although the formulations focus on the two-stage amplifier of Fig. 3.1, the results are readily extendable to other op-amp architectures as well.

3.2 Derivation of the Random and Deterministic CMRR

Since in multistage amplifiers the CMRR of the first stage is usually an important factor in the overall CMRR, the CMRR of the two-stage CMOS op-amp will be dominated by the first stage. The small signal equivalent circuit of the differential stage in Fig. 3.1 is shown in Fig. 3.2, where g_o denotes the internal output conductance of the transistor used as a bias current source. Ideally M1 and M2 are matched as are M3 and M4.

The small-signal output voltage is given by

$$v_o = A_{dm} v_d + A_{cm} v_c \tag{3.2}$$

where

$$v_d = v_{in1} - v_{in2} \tag{3.3}$$

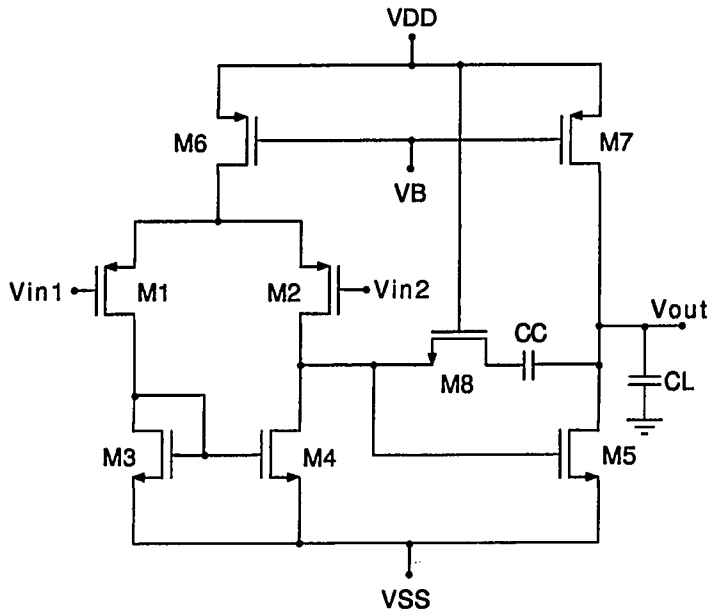


Figure 3.1: Two-stage CMOS operational amplifier

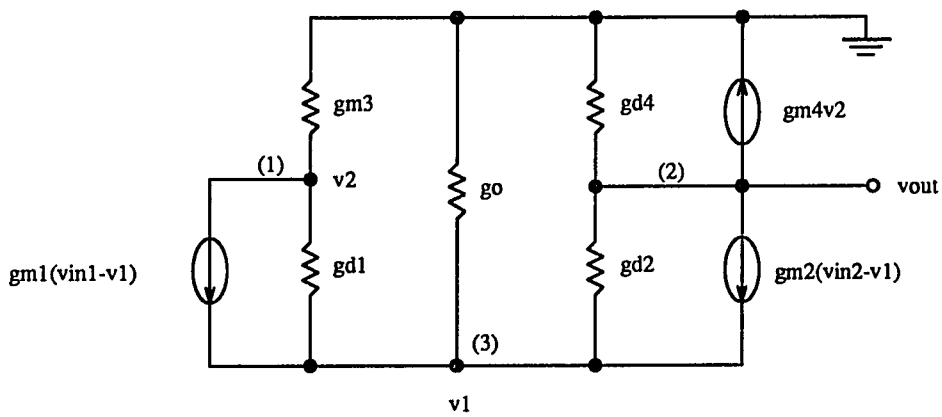


Figure 3.2: Small signal equivalent circuit of the differential stage of Fig. 3.1

$$v_c = \frac{v_{in1} + v_{in2}}{2}. \tag{3.4}$$

The nodal equations at nodes (1), (2), and (3) are

$$\begin{aligned}
(g_{m1} + g_{d1})v_1 - (g_{m3} + g_{d1})v_2 &= g_{m1}v_{in1} \\
(g_{m2} + g_{d2})v_1 - g_{m4}v_2 - (g_{d2} + g_{d4})v_{out} &= g_{m2}v_{in2} \\
(g_{m1} + g_{m2} + g_{d1} + g_{d2} + g_o)v_1 - g_{d1}v_2 - g_{d2}v_{out} &= g_{m1}v_{in1} + g_{m2}v_{in2}.
\end{aligned} \tag{3.5}$$

The model parameters are all random variables and can be expressed as

$$\begin{aligned}
g_{m1} &= g_{m1N} + g_{m1R1} + g_{m1R2} \\
g_{m2} &= g_{m2N} + g_{m2R1} + g_{m2R2} \\
g_{m3} &= g_{m3N} + g_{m3R1} + g_{m3R2} \\
g_{m4} &= g_{m4N} + g_{m4R1} + g_{m4R2} \\
g_{d1} &= g_{d1N} + g_{d1R1} + g_{d1R2} \\
g_{d2} &= g_{d2N} + g_{d2R1} + g_{d2R2} \\
g_{d4} &= g_{d4N} + g_{d4R1} + g_{d4R2},
\end{aligned} \tag{3.6}$$

where the N subscript denotes the nominal value which is deterministic, the R1 subscript denotes a random component that is process dependent but which does not vary from device to device on a wafer and where the R2 subscript denotes a random component that varies randomly from device to device on a wafer. It will be assumed that process dependent random variables (those with an R1 subscript) are totally correlated and identical for matched devices and that the wafer-level random variables (those with an R2 subscript) are identically distributed for ideally matched devices but statistically uncorrelated.

Assuming that $g_{mk} >> g_{dl}$, for all $k, l \in \{1, 2, 3, 4\}$ and that M1 and M2 are nominally matched as are M3 and M4, we can obtain the expressions for the differential-mode gain A_{dm} and the common-mode gain A_{cm}, which are themselves random variables,

$$A_{dm} \simeq \frac{2g_{mi}^2 g_{ml} + g_{mi}^2(2g_{mlR1} + g_{m3R2} + g_{m4R2}) + 2g_{mi}g_{ml}(2g_{miR1} + g_{m1R2} + g_{m2R2})}{2g_{mi}g_{ml}(g_{di} + g_{dl})} \tag{3.7}$$

$$\begin{aligned}
A_{cm} \simeq \frac{1}{2g_{mi}g_{ml}(g_{di} + g_{dl})} &[-g_{di}g_{mi}g_o + (2g_{di}g_{ml} + g_o g_{ml})(g_{m1R2} - g_{m2R2}) \\
&-2g_{mi}g_{ml}(g_{d1R2} - g_{d2R2}) - g_o g_{mi}(g_{m3R2} - g_{m4R2})],
\end{aligned} \tag{3.8}$$

where

$$\begin{aligned} g_{mi} &= g_{m1N} = g_{m2N} \\ g_{miR1} &= g_{m1R1} = g_{m2R1} \\ g_{ml} &= g_{m3N} = g_{m4N} \\ g_{mlR1} &= g_{m3R1} = g_{m4R1} \\ g_{di} &= g_{d1N} = g_{d2N} \\ g_{diR1} &= g_{d1R1} = g_{d2R1} \\ g_{dl} &= g_{d4N}, \end{aligned} \tag{3.9}$$

where the i subscript denotes the input transistors M1 and M2, and the l subscript denotes the load transistors M3 and M4.

Since the random component of the differential gain is very small compared to the deterministic component of the differential gain as can be seen in (3.7), the total differential-mode gain can be approximated by the deterministic gain only. Hence,

$$\begin{aligned} A_{dm} &\simeq \frac{2g_{mi}^2 g_{ml}}{2g_{mi}g_{ml}(g_{di}+g_{dl})} \\ &= \frac{g_{mi}}{g_{di}+g_{dl}}. \end{aligned} \tag{3.10}$$

The random component of the common-mode gain is, however, comparable in magnitude to the deterministic component of the common-mode gain. The deterministic and random common-mode gains, A_{cm}^D and A_{cm}^R, can be defined so that

$$A_{cm} = A_{cm}^D + A_{cm}^R \tag{3.11}$$

From (3.8), natural definitions of A_{cm}^D and A_{cm}^R are

$$\begin{aligned} A_{cm}^D &= -\frac{g_{di}g_{mi}g_o}{2g_{mi}g_{ml}(g_{di}+g_{dl})} \\ &= -\frac{g_{di}g_o}{2g_{ml}(g_{di}+g_{dl})} \end{aligned} \tag{3.12}$$

$$A_{cm}^R = \frac{(2g_{di}g_{ml}+g_o g_{ml})(g_{m1R2}-g_{m2R2}) - 2g_{mi}g_{ml}(g_{d1R2}-g_{d2R2}) - g_o g_{mi}(g_{m3R2}-g_{m4R2})}{2g_{mi}g_{ml}(g_{di}+g_{dl})}$$

$$= \frac{1}{2(g_{di}+g_{dl})}\left[g_o\left(\frac{g_{m1R2}-g_{m2R2}}{g_{mi}} - \frac{g_{m3R2}-g_{m4R2}}{g_{ml}}\right)\right.$$

$$+2g_{di}\left(\frac{g_{m1R2}-g_{m2R2}}{g_{mi}}-\frac{g_{d1R2}-g_{d2R2}}{g_{di}}\right)\Bigg]. \tag{3.13}$$

The ratios of the numerator of (3.13) are readily obtained in terms of the geometric and process device parameters. Details of this calculation appear in Section 3.8. Substituting (3.84), (3.85) and (3.88) into (3.13) gives

$$\begin{aligned} A_{cm}^{R} &= \frac{1}{2(g_{di}+g_{dl})}\Bigg[g_o\left(\frac{W_{1R2}-W_{2R2}}{W_i}+\frac{L_{2R2}-L_{1R2}}{L_i}+\frac{W_{3R2}-W_{4R2}}{W_l}+\frac{L_{4R2}-L_{3R2}}{L_l}\right. \\ &\qquad \left.\frac{V_{T2R2}-V_{T1R2}}{V_{GSi}-V_{Ti}}+\frac{V_{T4R2}-V_{T3R2}}{V_{GSl}-V_{Tl}}\right)+2g_{di}\frac{V_{T1R2}-V_{T2R2}}{V_{GSi}-V_{Ti}}\Bigg] \end{aligned} \tag{3.14}$$

The CMRR, defined in (3.1) where A_{cm} is now a random variable, is itself a random variable. If we define

$$CMRR_D^{-1} = \frac{A_{cm}^{D}}{A_{dm}} \tag{3.15}$$

$$CMRR_R^{-1} = \frac{A_{cm}^{R}}{A_{dm}}, \tag{3.16}$$

then we have

$$\begin{aligned} CMRR &= \left|\frac{A_{dm}}{A_{cm}}\right| \\ &= \left|\frac{A_{dm}}{A_{cm}^{D}+A_{cm}^{R}}\right| \\ &= \left|\frac{1}{CMRR_D^{-1}+CMRR_R^{-1}}\right|. \end{aligned} \tag{3.17}$$

From (3.10), (3.12), and (3.14)-(3.16), the deterministic and random CMRRs are given by

$$CMRR_D^{-1} = -\frac{g_{di}g_o}{2g_{mi}g_{ml}} \tag{3.18}$$

and

$$\begin{aligned} CMRR_R^{-1} &= \frac{1}{2g_{mi}}\Bigg[g_o\left(\frac{W_{1R2}-W_{2R2}}{W_i}+\frac{L_{2R2}-L_{1R2}}{L_i}+\frac{W_{3R2}-W_{4R2}}{W_l}+\frac{L_{4R2}-L_{3R2}}{L_l}\right. \\ &\qquad \left.\frac{V_{T2R2}-V_{T1R2}}{V_{GSi}-V_{Ti}}+\frac{V_{T4R2}-V_{T3R2}}{V_{GSl}-V_{Tl}}\right)+2g_{di}\frac{V_{T1R2}-V_{T2R2}}{V_{GSi}-V_{Ti}}\Bigg]. \end{aligned} \tag{3.19}$$

The deterministic CMRR given by (3.18) is as reported in [72] and [73]. From (3.19) we can see that the random component of the CMRR is caused by the nonzero output conductances of

the bias current source and the input transistors as well as the mismatch of the paired devices. It can be seen that the effect due to g_o on the random CMRR are more dominant than that due to g_{di}.

We are accustomed to characterizing the CMRR by a single number. Unfortunately, it can be seen from (3.17)-(3.19) that the CMRR is actually a random variable and, as such, characterized by a probability density function, not a single number. Nonetheless, it is instructive to develop an appreciation for what the CMRR of sample amplifiers will be and to determine how important the random part of the CMRR actually is. At this stage, we will calculate a pseudo worst case CMRR to compare the magnitude of the random and deterministic components of the CMRR. The probability density function itself will be explored in the next section.

To calculate the pseudo worst case CMRR of the op-amp in Fig. 3.1 whose simulated parameter values are shown in Table 3.3, it is assumed that the wafer-level random component of L and W are normally distributed with zero mean and standard deviation

$$\sigma_L = \sigma_W = 0.014\mu m. \tag{3.20}$$

We chose $\sigma_{\Delta L} = \sigma_{\Delta W} = 0.02\mu m$ which is very reasonable choice as indicated in [79]. From the choice equation (20) was obtained. Since $\Delta L = L_1 - L_2 = L_{1R2} - L_{2R2}$ and $\sigma_{\Delta L} = \sqrt{\sigma_{L1}^2 + \sigma_{L2}^2}$, $\sigma_L = \sigma_{L1} = \sigma_{L2} = \sigma_{\Delta L}/\sqrt{2} = 0.014\mu m$. It is also assumed that the corresponding random component of V_T is normally distributed with zero mean and standard deviation

$$\sigma_{V_T} = \frac{k}{\sqrt{LW}}, \tag{3.21}$$

where k=0.0236 Vμm. The k value was obtained based on the choice of $\sigma_{\Delta V_T} = \frac{5}{3}$mV for LW=20 $\times$ 20μm^2 according to the experimental data in [80].

We define the pseudo worst case CMRR to be the sample CMRR that would result if all random variables comprising the CMRR are in the direction that they add and at the 3σ value that would most degrade the sample CMRR. The corresponding σ values for width, length and threshold voltage variations are summarized in Table 3.4. The deterministic CMRR calculated from (3.18) was 63.7dB which is close to the simulated one shown in Table 3.2. The pseudo worst case random CMRR calculated from (3.19) was 51.6dB which dominates the deterministic CMRR. The worst case total CMRR was thus 49.6dB. Since the random CMRR can have both positive and negative polarity, the total CMRR can be either improved or degraded by the random CMRR.

Table 3.3: Simulated parameter values of the op-amp in Fig. 3.1

g_{mi}	$1030\mu A/V$	g_{ml}	$712\mu A/V$
g_o	$43.7\mu A/V$	g_{di}	$22.0\mu A/V$
$V_{GSi} - V_{Ti}$	$0.393V$	$V_{GSl} - V_{Tl}$	$0.542V$

Table 3.4: Component σ values for the op-amp in Fig. 3.1

σ_L	$0.014\mu m$	σ_W	$0.014\mu m$
$\sigma_{V_{Ti}}$	$1.17mV$	$\sigma_{V_{Tl}}$	$1.57mV$

3.3 Statistical Characteristics of CMRR

In this section the statistical characteristics of the random variable, CMRR as defined by (3.17), will be investigated. For notational convenience we will define

$$\mathbf{c} = CMRR \tag{3.22}$$

$$\mathbf{x} = CMRR_R^{-1} \tag{3.23}$$

$$d = CMRR_D^{-1} \tag{3.24}$$

$$\mathbf{y} = \mathbf{x} + d \tag{3.25}$$

where the bold letters are used to denote random variables. From (3.17), the common-mode rejection ratio can be expressed as

$$\mathbf{c} = \left|\frac{1}{\mathbf{x}+d}\right| = \left|\frac{1}{\mathbf{y}}\right| = \frac{1}{|\mathbf{y}|}. \tag{3.26}$$

Equation (3.19) shows that the random variable $\mathbf{x} = CMRR_R^{-1}$ is a function of 12 random variables. These random variables are assumed to be independent and normally distributed with zero mean.

$$\begin{aligned} W_{1R2}, W_{2R2}, W_{3R2}, W_{4R2} &\sim N(0, \sigma_W^2) \\ L_{1R2}, L_{2R2}, L_{3R2}, L_{4R2} &\sim N(0, \sigma_L^2) \\ V_{T1R2}, V_{T2R2} &\sim N(0, \sigma_{V_{Ti}}^2) \\ V_{T3R2}, V_{T4R2} &\sim N(0, \sigma_{V_{Tl}}^2). \end{aligned} \tag{3.27}$$

Since $\mathbf{x}$ is the sum of 12 uncorrelated zero mean random variables, its mean will also be zero and its variance is equal to the sum of their variances. Thus, $\mathbf{x}$ is distributed as

$$\mathbf{x} \sim N(0, \sigma_x^2) \tag{3.28}$$

where

$$\sigma_x^2 = \frac{1}{2g_{mi}^2}\left[g_o^2\sigma_L^2\left(\frac{1}{L_i^2}+\frac{1}{L_l^2}\right)+g_o^2\sigma_W^2\left(\frac{1}{W_i^2}+\frac{1}{W_l^2}\right)+\frac{\sigma_{V_{Ti}}^2(g_o^2+4g_{di}^2)}{(V_{GSi}-V_{Ti})^2}+\frac{\sigma_{V_{Tl}}^2 g_o^2}{(V_{GSl}-V_{Tl})^2}\right]. \tag{3.29}$$

Since d in (3.24) is deterministic, the random variable $\mathbf{y} = \mathbf{x} + d$ is normally distributed with mean d and variance σ_x^2,

$$\mathbf{y} \sim N(d, \sigma_x^2). \tag{3.30}$$

The mean of $|\mathbf{y}|$ can be expressed as [77].

$$E\{\,|\mathbf{y}|\,\} = \sigma_x\sqrt{\frac{2}{\pi}}\,e^{-d^2/2\sigma_x^2} + 2d\,P\left(\frac{d}{\sigma_x}\right) - d \tag{3.31}$$

where

$$P(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-y^2/2}dy. \tag{3.32}$$

The variance of $|\mathbf{y}|$ is then

$$\begin{aligned}\sigma_{|y|}^2 &= E\{\,|\mathbf{y}|^2\,\} - E^2\{\,|\mathbf{y}|\,\}\\ &= E\{\,\mathbf{y}^2\,\} - E^2\{\,|\mathbf{y}|\,\}\\ &= Var\{\,\mathbf{y}\,\} + E^2\{\,\mathbf{y}\,\} - E^2\{\,|\mathbf{y}|\,\}\\ &= \sigma_x^2 + d^2 - E^2\{\,|\mathbf{y}|\,\}.\end{aligned} \tag{3.33}$$

The probability density function, $f_c(c)$, of the common mode rejection ratio $\mathbf{c}$ can be obtained as follows. We want to determine the density of $\mathbf{c}$ in terms of the density of $\mathbf{y}$. Since $c > 0$, $f_c(c) = 0\ \forall c \le 0$. The equation $c = |\frac{1}{y}|$ has two solutions for $c > 0$,

$$y_1 = \frac{1}{c}, \qquad y_2 = -\frac{1}{c}. \tag{3.34}$$

From the fundamental theorem of determining the density of a function of a random variable [77], the pdf of $\mathbf{c}$ is then

$$\begin{aligned}f_c(c) &= \frac{f_y(y_1)}{|g'(y_1)|} + \frac{f_y(y_2)}{|g'(y_2)|}\\ &= \frac{1}{c^2}\left[f_y\left(\frac{1}{c}\right) + f_y\left(-\frac{1}{c}\right)\right],\end{aligned} \tag{3.35}$$

where $f_y(y)$ is the probability density function of $\mathbf{y}$, and $g(y) = |\frac{1}{y}|$. Since from (3.30) the pdf of $\mathbf{y}$ is

$$f_y(y) = \frac{1}{\sqrt{2\pi}\sigma_x}exp\left[-\frac{(y-d)^2}{2\sigma_x^2}\right], \tag{3.36}$$

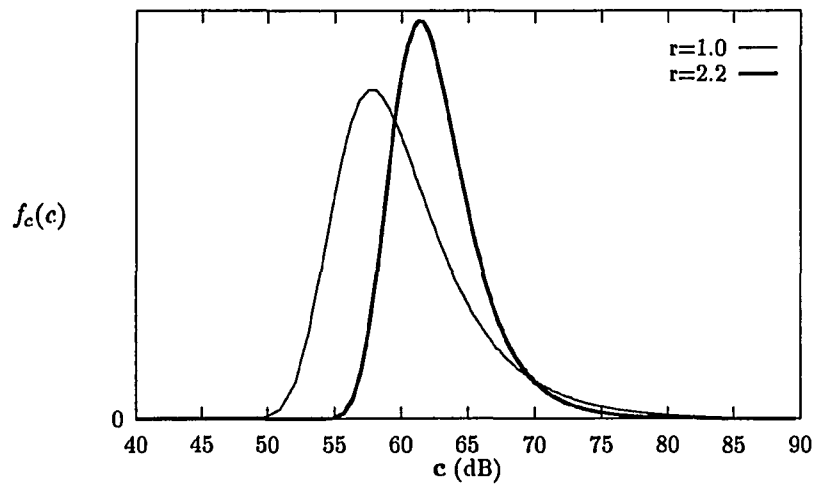


Figure 3.3: Probability density curves of CMRR. $\mathbf{c} = CMRR$ and $r = |d/\sigma_x|$

the pdf of the common-mode rejection ratio **c** becomes

$$f_c(c) = \frac{1}{\sqrt{2\pi}\sigma_x c^2}\left[exp\left\{-\frac{(1-dc)^2}{2\sigma_x^2 c^2}\right\} + exp\left\{-\frac{(1+dc)^2}{2\sigma_x^2 c^2}\right\}\right], \quad c > 0 \tag{3.37}$$

The probability density curves of **c** are shown in Fig. 3.3 where $r = |d/\sigma_x|$ and the $CMRR_D^{-1}$ of the op-amp in Fig. 3.1 was used for d. These curves show that the pdf of **c** is similar to a Gaussian density function, but it is not symmetric, and the left side of the peak point goes to zero faster than the right side, so the mean lies at the right of the peak point. Fig. 3.3 also shows that for the op-amp of Fig. 3.1, the CMRR probability below 55dB is almost zero. Since the pdf of **c** is known, the mean and variance can be found from the expressions,

$$E\{\mathbf{c}\} = \int_0^\infty c f_c(c) dc \tag{3.38}$$

$$\sigma_c^2 = E\{\mathbf{c}^2\} - E^2\{\mathbf{c}\}. \tag{3.39}$$

If $|\mathbf{y}|$ is concentrated near its mean, then $E\{\mathbf{c}\}$ and σ_c^2 can be approximated from the procedure of estimating the mean and variance of the functions of a random variable [77]. Let $\mathbf{c} = f(|\mathbf{y}|) = \frac{1}{|\mathbf{y}|}$ and $m = E\{|\mathbf{y}|\}$. If $f(|y|)$ is approximated by the first three terms of the Taylor series of $f(|y|)$ with center m, then

$$f(|y|) \simeq f(m) + f'(m)(|y| - m) + \frac{f''(m)}{2}(|y| - m)^2. \tag{3.40}$$

Taking the expected values on (3.40), we obtain

$$E\{ f(|\mathbf{y}|) \} \simeq f(m) + \frac{f''(m)}{2} \left(E\{ |\mathbf{y}|^2 \} - m^2 \right). \tag{3.41}$$

The approximated $E\{ \mathbf{c} \}$ is thus

$$E\{ \mathbf{c} \} \simeq \frac{1}{E\{ |\mathbf{y}| \}} \left[1 + \left(\frac{\sigma_{|y|}}{E\{ |\mathbf{y}| \}} \right)^2 \right]. \tag{3.42}$$

The first-order estimate of σ_c^2 is given by

$$\begin{aligned} \sigma_c^2 &\simeq |f'(m)|^2 \sigma_{|y|}^2 \\ &= \left(\frac{\sigma_{|y|}}{E^2\{ |\mathbf{y}| \}} \right)^2. \end{aligned} \tag{3.43}$$

The mean and variance of $|\mathbf{y}|$ are given in (3.31) and (3.33). From (3.37)-(3.39), (3.42), and (3.43) it is clear that the statistical characteristics of the common-mode rejection ratio, i.e., its mean, variance, and pdf, can be readily obtained if the variance of the process parameters are known.

The statistical parameters of the CMRR of the sample op-amp in Fig. 3.1 were calculated using the derived equations and the data in Table 3.3 and 3.4. The approximated equations (3.42) and (3.43) were used to calculate $E\{ \mathbf{c} \}$ and σ_c. The calculated results are listed in Column A of Table 3.5. In order to investigate the correctness of these derived equations, 200 Gaussian random numbers with zero mean and variance σ_x^2 were generated and used to calculate the corresponding parameters. From these sample data of the random variable $\mathbf{x}$, the sample data of $|\mathbf{y}|$ and $\mathbf{c}$ can be obtained using (3.25) and (3.26). Their calculated mean and variance are shown in Column B of Table 3.5. The $E\{ |\mathbf{y}| \}$ and $\sigma_{|y|}$ from the derived equations are very close to those from the generated sample data, but the $E\{ \mathbf{c} \}$ and σ_c of Column A somewhat differ from those of Column B because the $E\{ \mathbf{c} \}$ and σ_c were calculated from the approximated equations (3.42) and (3.43). The histogram of the generated random data of $\mathbf{x}$ and the CMRR histogram are shown in Fig. 3.4 and Fig. 3.5. Since the $r(= |d/\sigma_x|)$ of the sample op-amp in Fig. 3.1 is 2.2, Fig. 3.5 corresponds to the curve ($r = 2.2$) of Fig. 3.3. These two plots are very similar and support the model of equation (37) for the pdf of $\mathbf{c}$.

3.4 Definition of the CMRR for Processes

The random offset of a CMOS amplifier has been defined for processes as three times its standard deviation. The reason is that the offset voltage has a Gaussian distribution, so 99.7%

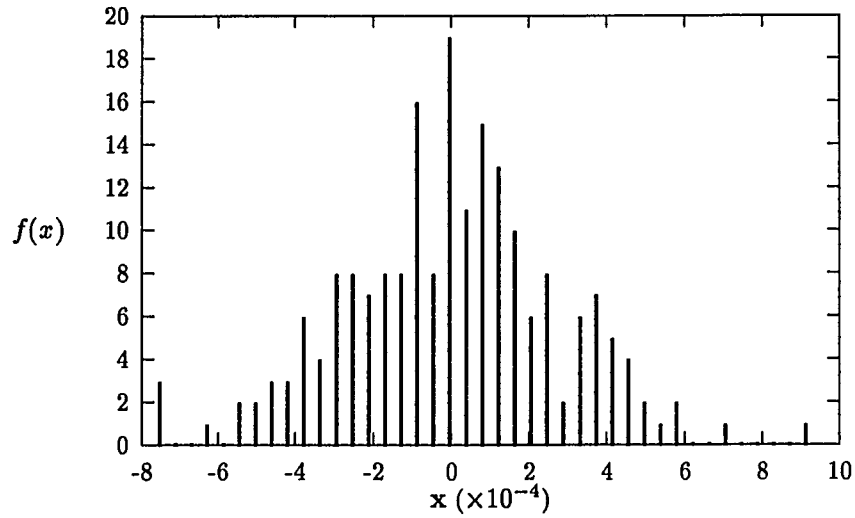


Figure 3.4: Histogram of the 200 samples generated for the random variable **x**. $\mathbf{x} = CMRR_R^{-1}$

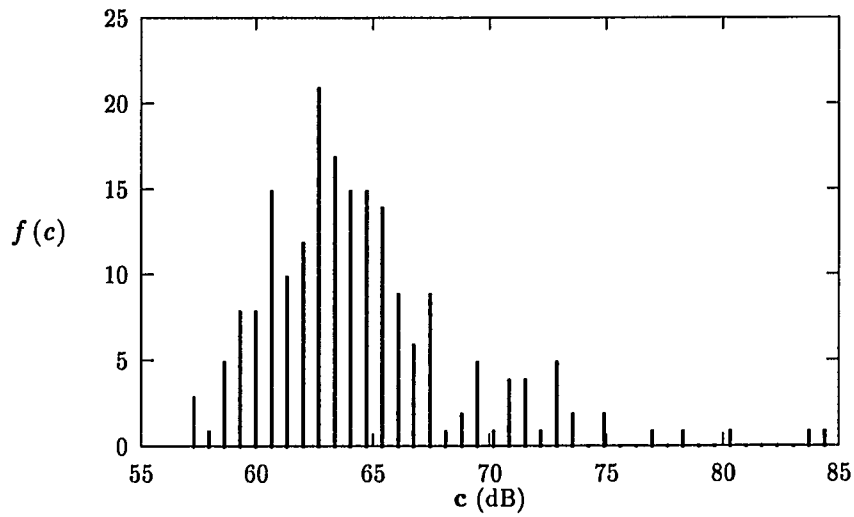


Figure 3.5: Histogram of the 200 samples calculated from the data in Fig. 3.4 for the random variable **c**. $\mathbf{c} = CMRR$

Table 3.5: The CMRR statistical characteristics of the op-amp in Fig. 3.1 calculated (A) from the derived equations (B) from the 200 generated random numbers

	A	B
d	-6.55×10^{-4}	
σ_x	2.976×10^{-4}	2.763×10^{-4}
$E\{\|y\|\}$	6.579×10^{-4}	6.612×10^{-4}
$\sigma_{\|y\|}$	2.797×10^{-4}	2.691×10^{-4}
$E\{c\}$	1.795×10^{3} (65 dB)	2.017×10^{3} (66 dB)
σ_c	6.462×10^{2}	1.847×10^{3}

of a sample satisfies the specification. Attention, however, has not been paid to the random CMRR of CMOS amplifiers, and no definition of the CMRR including random components has been made. Thus, the CMRR of CMOS op-amps for processes will be defined in this section.

In the previous section we found the probability density function $f_c(c)$ of the CMRR. We will define the CMRR to the value of $\hat{c}$ such that 99.86% of a sample set has a CMRR greater than $\hat{c}$. The choice of the 99.86% which is close to the 99.7% used in the definition of offset voltages discussed above will be discussed later. Integration of the pdf, $f_c(c)$, from $\hat{c}$ to infinity gives the following results:

$$\int_{\hat{c}}^{\infty} f_c(c)dc = P(a) + P(b) - 1 \tag{3.44}$$

where

$$a = \frac{1/\hat{c} - d}{\sigma_x} \tag{3.45}$$

$$b = \frac{1/\hat{c} + d}{\sigma_x}. \tag{3.46}$$

Since d is negative for the sample op-amp, we can rewrite a and b as

$$a = \frac{1}{\sigma_x \hat{c}} + \left|\frac{d}{\sigma_x}\right|, \qquad b = \frac{1}{\sigma_x \hat{c}} - \left|\frac{d}{\sigma_x}\right|. \tag{3.47}$$

From equation (3.47) we can see that a is always greater than b by $2|d/\sigma_x|$. Thus, $P(a)$ is also always greater than $P(b)$ because $P(x)$ defined in (3.32) increases from 0.5 to 1.0 as x increases from 0 to ∞.

Since we want to make

$$\int_{\hat{c}}^{\infty} f_c(c)dc = 0.9986, \tag{3.48}$$

$P(b)$ should be very close to 1.0. This means that $P(a)$ is almost 1.0 because $P(a)$ is greater than $P(b)$, and the maximum value of the function $P(x)$ is 1.0. In most cases, $|d/\sigma_x| > 0.5$, so

$a > b + 1$. Therefore, under the condition of (3.48), the approximation

$$\begin{aligned} \int_{\hat{c}}^{\infty} f_c(c)dc &\simeq P(b) \\ &= P\left(\frac{1/\hat{c}+d}{\sigma_x}\right) \end{aligned} \tag{3.49}$$

can be used. From the equation (3.48) and (3.49) we obtain

$$P\left(\frac{1/\hat{c}+d}{\sigma_x}\right) = 0.9986. \tag{3.50}$$

It now follows from tables for $P(x)$ [78] that (3.50) will be satisfied provided

$$\frac{1/\hat{c}+d}{\sigma_x} = 3 \tag{3.51}$$

which can be expressed as

$$\hat{c} = (3\sigma_x - d)^{-1}. \tag{3.52}$$

The reason why we chose a figure of 0.9986 in (3.48) was to obtain the integer 3 in (3.51). If we use $(3\sigma_x - d)^{-1}$ as the CMRR specification in designing CMOS amplifiers, then 99.86% of a large sample will satisfy the specification. If d is positive, then $P(b)$ is greater than $P(a)$ and finally we have

$$\hat{c} = (3\sigma_x + d)^{-1}. \tag{3.53}$$

Therefore, we can define the CMRR for processes as

$$CMRR = (3\sigma_x + |d|)^{-1} \tag{3.54}$$

where d and σ_x are $CMRR_D^{-1}$ and the standard deviation of $CMRR_R^{-1}$. The $CMRR_D^{-1}$ and $CMRR_R^{-1}$ were defined in (3.15) and (3.16). The calculated CMRR for the sample op-amp in Fig. 3.1 was $56.2dB$. Comparing with the density curve (r=2.2) in Fig. 3.3, we can see that the value $56.2dB$ is very reasonable.

The CMRR definition for processes of (3.54) and the CMRR pdf of (3.37) are general for the op-amps whose deterministic and random components comparably contribute to the total CMRR. This case usually corresponds to the op-amps whose first stage has a single-ended output. If op-amps have a first stage with differential output, then their deterministic common-mode gains are significantly reduced by the next stages [75]. In these cases the deterministic CMRR can be ignored, i.e., $d \simeq 0$ and the above CMRR definition and the pdf should be changed. If d is nearly zero, then the pdf of the total CMRR is

$$f_c(c) = \frac{2}{\sqrt{2\pi}\sigma_x c^2} exp\left[-\frac{1}{2\sigma_x^2 c^2}\right], \qquad c > 0. \tag{3.55}$$

The integration of the pdf from $\hat{c}$ to ∞ becomes

$$\int_{\hat{c}}^{\infty} f_c(c)dc = 2P\left(\frac{1}{\sigma_x \hat{c}}\right) - 1. \tag{3.56}$$

The CMRR definition for processes is thus

$$CMRR = (3\sigma_x)^{-1} \tag{3.57}$$

where 99.73% of a sample set will be greater than $(3\sigma_x)^{-1}$. The approximated mean and variance of the CMRR have the same equations (3.42) and (3.43), but the $E\{|\mathbf{y}|\}$ and $\sigma_{|\mathbf{y}|}$ should be modified as follows:

$$E\{|\mathbf{y}|\} = \sigma_x\sqrt{\frac{2}{\pi}} \tag{3.58}$$

$$\sigma_{|\mathbf{y}|} = \sigma_x^2(1 - \frac{2}{\pi}) \tag{3.59}$$

3.5 Offset Analysis

The offset voltage of an op-amp consists of two components: a deterministic offset and a random offset. The former results from improper dimensions and/or bias conditions, so it can be reduced to a very small value by careful design. The latter is due to the random errors in the fabrication process, i.e., mismatches in identically designed pairs of devices. For two-stage op-amps the first-stage will have a dominant effect on the offset. Therefore, the total input referred offset voltage of the two-stage op-amp will be highly affected by the first-stage random offset voltage. The input offset voltage, V_{OS}, is defined as the differential input voltage that is required to make the differential output voltage exactly zero. If both input terminals are grounded, then the input referred offset voltage of the first stage can be expressed as

$$\begin{aligned} V_{OS} &= \frac{V_o}{A} \\ &= \frac{\Delta I_D}{g_m} \\ &= \frac{\Delta I_D}{2I_D/(V_{GSi} - V_{Ti})} \\ &= \frac{V_{GSi} - V_{Ti}}{2}\frac{\Delta I_D}{I_D}, \end{aligned} \tag{3.60}$$

where V_o is the first-stage output voltage, and A is the first-stage small-signal voltage gain.

Since ΔI_D is mainly affected by the mismatch in the threshold voltage and the device width and length, and other factors can be ignored [79], we will consider only offsets in the V_T

and W/L of the input differential pair (M1 and M2) and the current mirror pair (M3 and M4) in Fig. 3.1. The $\Delta I_{Di} = I_{D1} - I_{D2}$ due to the mismatch of the input differential pair and the $\Delta I_{Dl} = I_{D3} - I_{D4}$ due to the mismatch of the current mirror pair are given from Section 3.8. Substituting (3.89) and (3.90) into (3.60), we have the input offset voltage due to the mismatch of the input differential pair,

$$V_{OSi} = V_{T2R2} - V_{T1R2} + \frac{V_{GSi} - V_{Ti}}{2}\left(\frac{W_{1R2} - W_{2R2}}{W_i} + \frac{L_{2R2} - L_{1R2}}{L_i}\right), \tag{3.61}$$

and the input offset voltage due to the current mirror pair,

$$V_{OSl} = \frac{V_{GSi} - V_{Ti}}{2}\left(\frac{W_{3R2} - W_{4R2}}{W_l} + \frac{L_{4R2} - L_{3R2}}{L_l}\right) + \frac{V_{GSi} - V_{Ti}}{V_{GSl} - V_{Tl}}(V_{T4R2} - V_{T3R2}). \tag{3.62}$$

The total input referred offset will be the sum of these terms (3.61) and (3.62),

$$\begin{aligned} V_{OS} &= V_{OSi} + V_{OSl} \\ &= \frac{V_{GSi} - V_{Ti}}{2}\left[\frac{W_{1R2} - W_{2R2}}{W_i} + \frac{L_{2R2} - L_{1R2}}{L_i} + \frac{W_{3R2} - W_{4R2}}{W_l}\right. \\ &\qquad \left. + \frac{L_{4R2} - L_{3R2}}{L_l} + \frac{2(V_{T2R2} - V_{T1R2})}{V_{GSi} - V_{Ti}} + \frac{2(V_{T4R2} - V_{T3R2})}{V_{GSl} - V_{Tl}}\right]. \end{aligned} \tag{3.63}$$

Since the offset voltage is the sum of 12 uncorrelated zero mean Gaussian random variables, it is also normally distributed with zero mean and standard deviation

$$\begin{aligned} \sigma_{V_{OS}} &= \frac{(V_{GSi} - V_{Ti})}{\sqrt{2}}\left[\sigma_L^2\left(\frac{1}{L_i^2} + \frac{1}{L_l^2}\right) + \sigma_W^2\left(\frac{1}{W_i^2} + \frac{1}{W_l^2}\right) + \right. \\ &\qquad \left. \frac{4\sigma_{V_{Ti}}^2}{(V_{GSi} - V_{Ti})^2} + \frac{4\sigma_{V_{Tl}}^2}{(V_{GSl} - V_{Tl})^2}\right]^{\frac{1}{2}}. \end{aligned} \tag{3.64}$$

Therefore, the offset has a Gaussian density function with zero mean and variance $\sigma_{V_{OS}}$.

Assuming again the pseudo worst case as in Section 3.2, and using the data of Table 3.3 and 3.4, the calculated pseudo worst case random offset of the sample op-amp in Fig. 3.1 is 27.9 mV. The offset due to the (W/L) mismatch is 14.1 mV while the offset due to the V_T mismatch is 13.8 mV. It shows that the two factors give almost equal contribution to the random offset for the sample op-amp.

3.6 Analysis of Op-amp Errors

The gain of a unity-gain configured op-amp will be exactly one if the op-amp is ideal. Practical op-amps, however, don't offer the exact gain because of finite differential gains, finite

Table 3.6: Simulated gains of the op-amp in Fig. 3.6

A	386.8	A'	386.6
A_d	386.5	A'_d	386.2
A_c	0.4811	A'_c	0.4803
$CMRR$	805.8	$CMRR'$	805.8
V_{OS}	$-20.4\mu V$		

common-mode rejection ratios, and nonzero offset voltages. In this section, the op-amp errors associated with these nonideal effects are analyzed. First, we define the different open-loop gains as shown in Fig. 3.7. We denote the finite open-loop gains of the op-amps which have different characteristics as follows:

A : Finite CMRR and nonzero offset.

A_d : Infinite CMRR and nonzero offset.

A' : Finite CMRR and zero offset.

A'_d : Infinite CMRR and zero offset.

Simulated results of these gains for the op-amp in Fig. 3.6 obtained by neglecting statistical variations are shown in Table 3.6, where A_c, $CMRR$, A'_c, and $CMRR'$ are the common mode gains and the common mode rejection ratios of a nonzero offset op-amp and a zero offset op-amp, respectively. The V_{OS} is the input referred offset voltage. The op-amp in Fig. 3.6 differs from that in Fig. 3.1. It has a programmable current mirror instead of a simple one as a load of the differential input pair. The programmable current mirror can be used to compensate the offset voltage of the op-amp by adjusting the bias voltages VT1 and/or VT2 as described in [6] and [81]. Basic concepts concerning the influence of each nonideal factor are briefly reviewed in the following three subsections. This is followed by discussions about the combined effects of the nonideal factors.

3.6.1 Finite Open-loop Gain Effect

Assuming that an op-amp has an infinite CMRR and a zero offset, the output voltage of the unity-gain configured op-amp will be

$$V_o = \frac{A'_d}{1 + A'_d} V_i. \tag{3.65}$$

If the pure differential gain A'_d is infinity, then the input V_i will be equal to the output V_o, but the output of a practical op-amp will be less than the input due to the finite open-loop

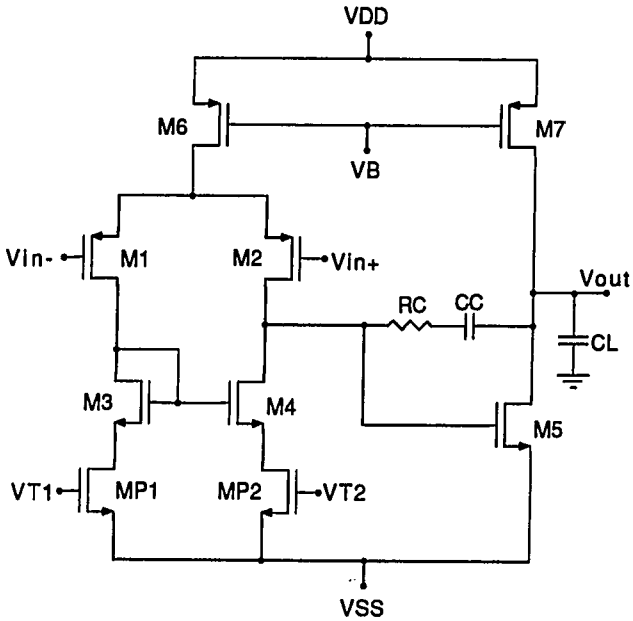


Figure 3.6: Two-stage CMOS operational amplifier with a programmable current mirror

gain. Hence, the gain of a unity-gain configured op-amp will be always less than one under the assumption of infinite CMRR and zero offset.

3.6.2 Finite CMRR Effect

Considering a finite-CMRR and zero-offset op-amp which is equivalent to the op-amp in Fig. 3.7 (c) if the voltage source V_{OS} is removed, the output of the op-amp will consist of two terms.

$$\begin{aligned} V_o &= V_c A_c' + V_d A_d' \\ &= \frac{V_1 + V_2}{2} A_c' + (V_2 - V_1) A_d' \\ &= \frac{V_1 + V_2}{2CMRR'} A_d' + (V_2 - V_1) A_d'. \end{aligned} \tag{3.66}$$

From these equations the op-amp can be modeled as in Fig. 3.7 (d) if the voltage source V_{OS} in Fig. 3.7 (d) is removed, where

$$V_{CMRR} = \frac{V_1 + V_2}{2CMRR'}. \tag{3.67}$$

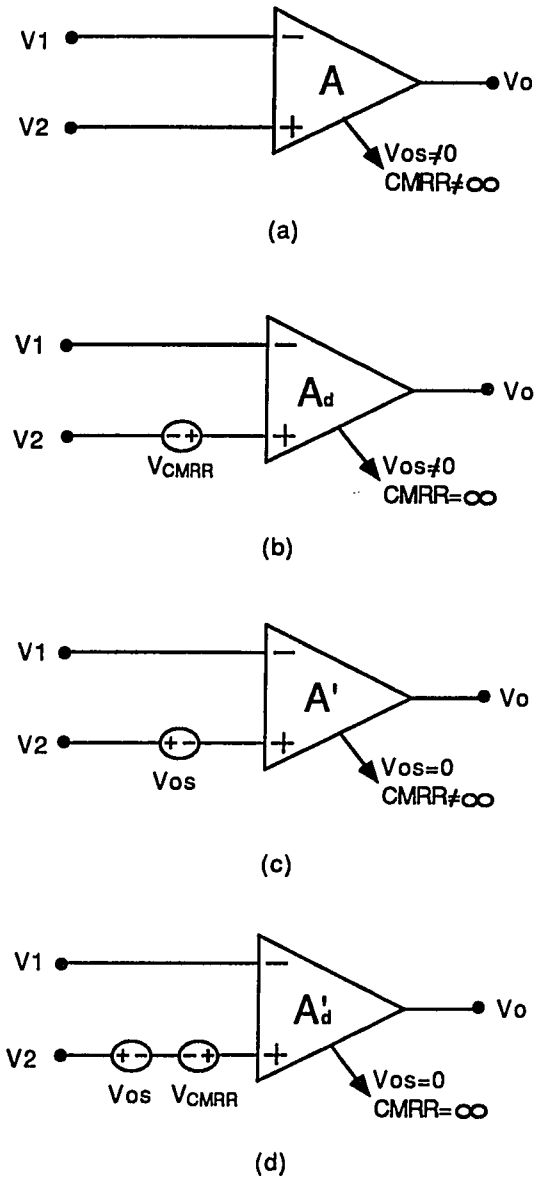


Figure 3.7: Equivalent models for a nonideal op-amp interpreting CMRR and offset and showing differently defined open-loop gains

If this op-amp is used for a unity-gain configuration, i.e., $V_1 = V_o$ and $V_2 = V_i$, then the output voltage will be

$$\begin{aligned} V_o &= A_d'(V_i + V_{CMRR'} - V_o) \\ &= A_d'(V_i + \frac{V_i + V_o}{2CMRR'} - V_o). \end{aligned} \tag{3.68}$$

Hence,

$$V_o = \frac{A_d'(1 + \frac{1}{2CMRR'})}{A_d'(1 - \frac{1}{2CMRR'}) + 1} V_i. \tag{3.69}$$

It can be seen that an infinite CMRR reduces the equation (3.69) to (3.65). The equation (3.69) shows that the finite CMRR can compensate or overcompensate the gain decreasing effect due to the finite open-loop gain.

3.6.3 Nonzero Offset Effect

To investigate the effect of nonzero offset, we consider an nonzero-offset and infinite-CMRR which is equivalent to the op-amp in Fig. 3.7 (b) if the voltage source V_{CMRR} is removed. The input referred offset voltage can be defined as the voltage applied at the positive input so that the voltage existing at the output becomes zero. Thus, the nonzero-offset and infinite-CMRR op-amp can be modeled as a voltage source V_{OS} which is equivalent to the input offset voltage and a pure differential op-amp. This model is equivalent to Fig. 3.7 (d) if the voltage source V_{CMRR} is removed. If this op-amp is used for a unity-gain configuration, then the output voltage will be

$$V_o = A_d'(V_i - V_{OS} - V_o). \tag{3.70}$$

Hence,

$$V_o = \frac{A_d'}{1 + A_d'}(V_i - V_{OS}), \tag{3.71}$$

where it is well known that the offset voltage can be either positive or negative.

3.6.4 Total Op-amp Error

Now, the three effects are combined to derive the total op-amp error. The nonideal op-amp shown in Fig. 3.7 (a) can be modeled as two voltage sources, V_{OS} and V_{CMRR}, applied at the positive input and a pure differential op-amp which has an infinite CMRR and a zero offset voltage as shown in Fig. 3.7 (d). The output is then

$$V_o = A_d'(V_2 - V_{OS} + V_{CMRR} - V_1)$$

$$= A_d'(V_2 - V_{OS} + \frac{V_1 + V_2 - V_{OS}}{2CMRR'} - V_1). \tag{3.72}$$

If this op-amp is used for a unity-gain configuration as shown in Fig. 3.8 (a), then the output will be

$$V_o = A_d'(V_i - V_{OS} + \frac{V_o + V_i - V_{OS}}{2CMRR'} - V_o). \tag{3.73}$$

The total output affected by a finite gain, a finite CMRR, and a nonzero offset is thus given by

$$V_o = \frac{A_d'(1 + \frac{1}{2CMRR'})}{A_d'(1 - \frac{1}{2CMRR'}) + 1}(V_i - V_{OS}) \tag{3.74}$$

where

$$CMRR' = \frac{A_d'}{A_c'} \approx \frac{A_d}{A_c} = CMRR. \tag{3.75}$$

If the op-amp is used for a high-gain configuration as shown in Fig. 3.8 (b), then the output becomes

$$V_o = \frac{A_d'(1 + \frac{1}{2CMRR'})}{A_d'\beta(1 - \frac{1}{2CMRR'}) + 1}(V_i - V_{OS}) \tag{3.76}$$

where

$$\beta = \frac{R_1}{R_1 + R_2}. \tag{3.77}$$

From the equation (3.74), it can be easily seen that the equations (3.71), (3.69), and (3.65) can be obtained by setting $V_{OS} = 0$, $CMRR' = \infty$, and both of them, respectively.

From the equation (3.74) and the data given in Table 3.6, the calculated unity-gain configured output voltage of the op-amp in Fig. 3.6 is 0.9987V when $V_i = 1.0$V while the simulated settling point of the output voltage is 0.9988V. This result shows that the equation (3.74) gives a very consistent result with the simulated one. In this example the random CMRR and the random offset have not been considered, but the correctness of the equation (3.74) has been demonstrated. In practical op-amps that kind of accuracy could not be obtained because of the random components described in the previous sections. With the assumption that $V_{OS} = 0$, the output errors of the op-amp in Fig. 3.6 as a function of CMRR were calculated at different closed-loop gains, and the results are shown in Fig. 3.9. Even though the offset is zero and the CMRR is very high, the output error of the unity-gain configured op-amp ($\beta = 1$) is about 0.3% due to the finite open-loop gain. If the CMRR is 52dB, then the output error is nearly zero. This shows that the finite CMRR can reduce the error attributable to the finite gain as mentioned in Section 3.6.2. From the figure it can be also seen that high-gain configured op-amps show more errors than low-gain op-amps.

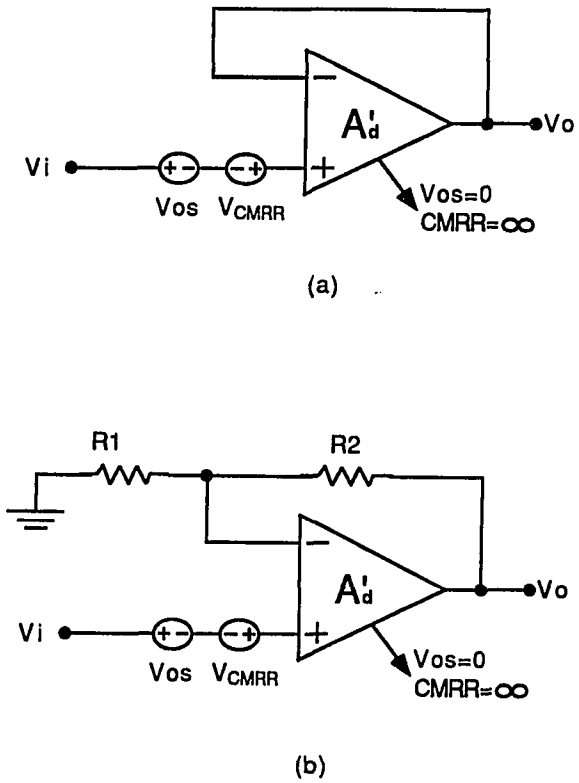


Figure 3.8: (a) Model of a unity-gain configured op-amp, (b) Model of a high-gain configured op-amp

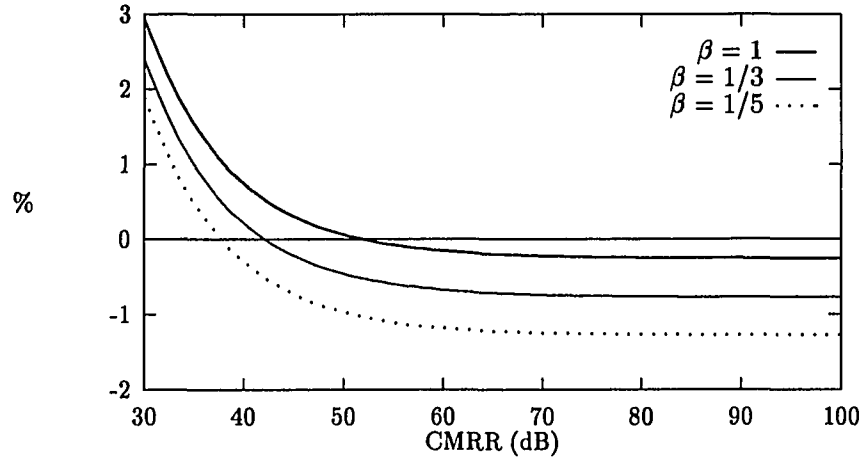


Figure 3.9: Output error of the op-amp in Fig. 3.6 versus CMRR with β as a parameter. The offset voltage $V_{OS} = 0$, and the open-loop gain $A = 52$dB

If the offset of a given op-amp is compensated, and the compensated offset range is known, then the output error of the given op-amp can be analyzed from (3.74) and (3.76) because the CMRR range of the op-amp can be easily found from the pdf of the CMRR derived in Section 3.3 or the CMRR definition in Section 3.4. Assuming that the offset is adjusted to less than 1mV in magnitude, the output errors of the sample op-amp in Fig. 3.1 were analyzed. It was shown in Section IV that the sample op-amp in Fig. 3.1 had CMRR for the process of about 56dB. Thus, the CMRR of most individual amplifiers will be greater than 56dB. Fig. 3.10 shows the output errors relative to 2V of the unity-gain configured sample op-amp as a function of the input V_i. From the 56dB CMRR curves in Fig. 3.10(a) and the 100dB CMRR curves in Fig. 3.10(b), it can be seen that the output errors are less than 0.2% through the input range of -2V to $+2$V if the magnitude of the input offset is less than 1mV. As expected, the 56dB CMRR curves show reduced errors compared to those of the 100dB CMRR curves.

3.7 Conclusions

The CMRR and offset of two-stage CMOS op-amps have been analyzed. Equations representing their statistical characteristics have been derived. Using these equations, we can readily find the distribution, mean, and variance of the CMRR and offset if the process parameter variations are given. The derived equations have shown that the CMRR pdf is similar to that of

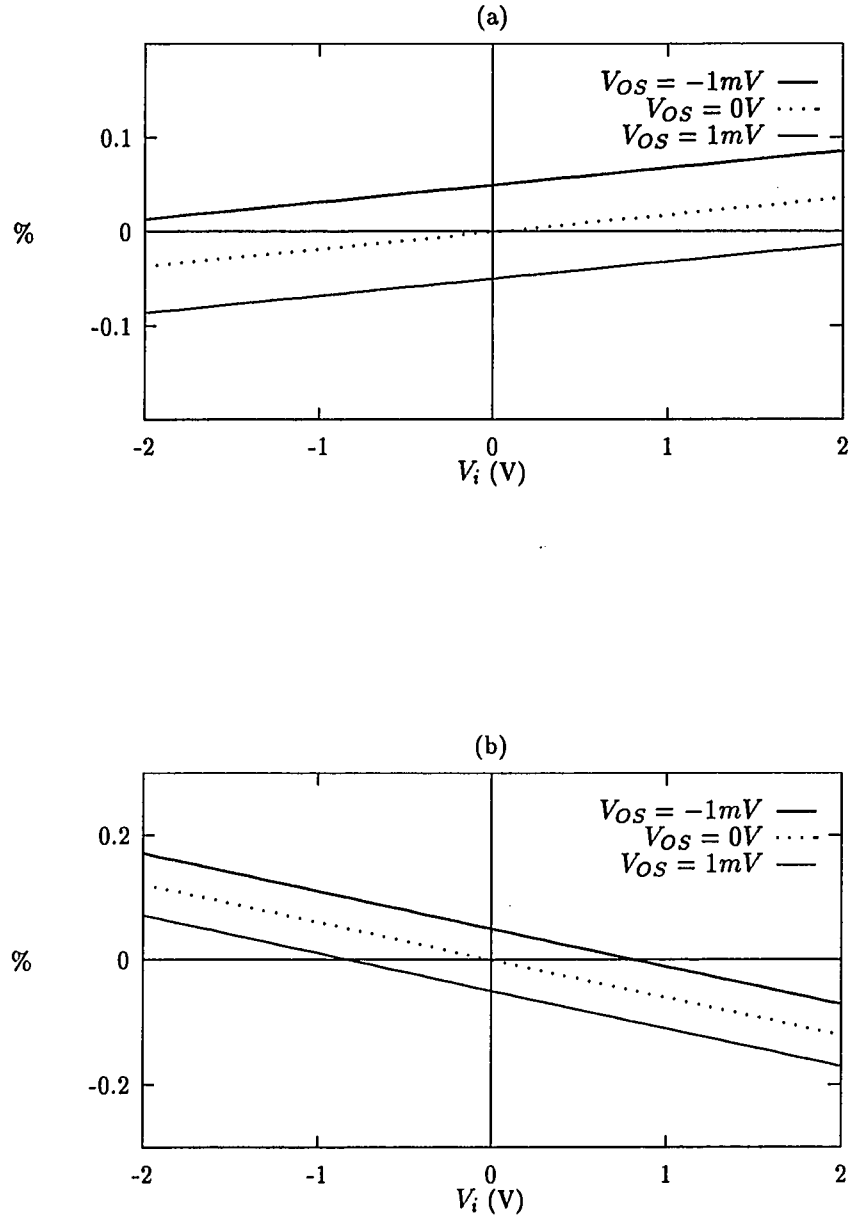


Figure 3.10: Output error of the op-amp in Fig. 3.1 versus V_i with V_{OS} as a parameter. The open-loop gain $A = 58$dB, and CMRRs are: (a) 56dB, (b) 100dB

a Gaussian random variable, but the mean is not zero and the symmetry is somewhat skewed, whereas the offset has a Gaussian distribution with zero mean. The CMRR for the processes has been defined. The CMRR is defined by $(3\sigma_x+|d|)^{-1}$ for the op-amps which have both dominant deterministic and random CMRR so that 99.86% of a large sample can be greater than the defined value. For the op-amps whose deterministic CMRRs are nearly zero, $(3\sigma_x)^{-1}$ can be used for the definition of the CMRR, where 99.73% of a large sample satisfies the specification. The variable d is the ratio of the deterministic common-mode gain to the differential-mode gain and σ_x is the standard deviation of the ratio of the random common-mode gain to the differential-mode gain.

The op-amp errors due to finite open-loop gains, finite CMRRs, and nonzero offsets have been analyzed. A finite differential open-loop gain always makes the gain of a unity-gain configured op-amp less than one, and a finite CMRR can compensate for the error attributable to the finite open-loop gain unless it is too small. If the compensated offset range is known, then the op-amp error range can be found.

3.8 Derivation of mismatch components

If the channel-length modulation effect is ignored, the small-signal transconductance gains of the paired transistors M1 and M2 which act in the saturation region are given by

$$g_{m1} = 2K'\left(\frac{W}{L}\right)_1 (V_{GSi} - V_{T1}) \tag{3.78}$$

$$g_{m2} = 2K'\left(\frac{W}{L}\right)_2 (V_{GSi} - V_{T2}). \tag{3.79}$$

where $K' = \mu C_{OX}/2$. Only mismatches in the V_T and W/L are considered. The similar expressions as in (3.6) for the random variables, L, W, and V_T, can be used as follows:

$$\begin{aligned} L_1 &= L_i + L_{iR1} + L_{1R2}, \qquad L_2 = L_i + L_{iR1} + L_{2R2} \\ W_1 &= W_i + W_{iR1} + W_{1R2}, \qquad W_2 = W_i + W_{iR1} + W_{2R2} \\ V_{T1} &= V_{Ti} + V_{TiR1} + V_{T1R2}, \qquad V_{T2} = V_{Ti} + V_{TiR1} + V_{T2R2}, \end{aligned} \tag{3.80}$$

where L_i, W_i, and V_{Ti} are the nominal values, and the subscript $R1$ and $R2$ are the same as before.

Using these definitions, g_{m1} can be approximated by ignoring higher order terms,

$$g_{m1} = 2K'\left(\frac{W_i + W_{iR1} + W_{1R2}}{L_i + L_{iR1} + L_{1R2}}\right)(V_{GSi} - V_{Ti} - V_{TiR1} - V_{T1R2})$$

$$
\begin{aligned}
&= 2K'\left(\frac{W_i}{L_i}\right)(V_{GSi}-V_{Ti})\left(\frac{1+(W_{iR1}+W_{1R2})/W_i}{1+(L_{iR1}+L_{1R2})/L_i}\right)\left(1-\frac{V_{TiR1}+V_{T1R2}}{V_{GSi}-V_{Ti}}\right)\\
&\simeq g_{mi}\left(1+\frac{W_{iR1}+W_{1R2}}{W_i}\right)\left(1-\frac{L_{iR1}+L_{1R2}}{L_i}\right)\left(1-\frac{V_{TiR1}+V_{T1R2}}{V_{GSi}-V_{Ti}}\right)\\
&\simeq g_{mi}\left(1+\frac{W_{iR1}+W_{1R2}}{W_i}-\frac{L_{iR1}+L_{1R2}}{L_i}-\frac{V_{TiR1}+V_{T1R2}}{V_{GSi}-V_{Ti}}\right). \qquad (3.81)
\end{aligned}
$$

By the same way,

$$
g_{m2} \simeq g_{mi}\left(1+\frac{W_{iR1}+W_{2R2}}{W_i}-\frac{L_{iR1}+L_{2R2}}{L_i}-\frac{V_{TiR1}+V_{T2R2}}{V_{GSi}-V_{Ti}}\right). \qquad (3.82)
$$

Hence,

$$
g_{m1}-g_{m2} = g_{mi}\left(\frac{W_{1R2}-W_{2R2}}{W_i}+\frac{L_{2R2}-L_{1R2}}{L_i}+\frac{V_{T2R2}-V_{T1R2}}{V_{GSi}-V_{Ti}}\right). \qquad (3.83)
$$

Since $g_{m1}-g_{m2}=g_{m1R2}-g_{m2R2}$ from (3.6) and (3.9),

$$
\frac{g_{m1R2}-g_{m2R2}}{g_{mi}} = \frac{W_{1R2}-W_{2R2}}{W_i}+\frac{L_{2R2}-L_{1R2}}{L_i}+\frac{V_{T2R2}-V_{T1R2}}{V_{GSi}-V_{Ti}}. \qquad (3.84)
$$

By the same procedure,

$$
\frac{g_{m3R2}-g_{m4R2}}{g_{ml}} = \frac{W_{3R2}-W_{4R2}}{W_l}+\frac{L_{4R2}-L_{3R2}}{L_l}+\frac{V_{T4R2}-V_{T3R2}}{V_{GSl}-V_{Tl}}. \qquad (3.85)
$$

Since the drain current I_D and the output conductance g_d can be expressed as

$$
I_D = K'\left(\frac{W}{L}\right)(V_{GS}-V_T)^2 \qquad (3.86)
$$

$$
g_d = \lambda I_D, \qquad (3.87)
$$

by the same method as above we can obtain

$$
\frac{g_{d1R2}-g_{d2R2}}{g_{di}} = \frac{W_{1R2}-W_{2R2}}{W_i}+\frac{L_{2R2}-L_{1R2}}{L_i}+\frac{2(V_{T2R2}-V_{T1R2})}{V_{GSi}-V_{Ti}}, \qquad (3.88)
$$

and

$$
\frac{I_{D1}-I_{D2}}{I_{Di}} = \frac{W_{1R2}-W_{2R2}}{W_i}+\frac{L_{2R2}-L_{1R2}}{L_i}+\frac{2(V_{T2R2}-V_{T1R2})}{V_{GSi}-V_{Ti}} \qquad (3.89)
$$

$$
\frac{I_{D3}-I_{D4}}{I_{Dl}} = \frac{W_{3R2}-W_{4R2}}{W_l}+\frac{L_{4R2}-L_{3R2}}{L_l}+\frac{2(V_{T4R2}-V_{T3R2})}{V_{GSl}-V_{Tl}} \qquad (3.90)
$$

where $I_D = I_{Di} = I_{Dl}$.

CHAPTER 4. AN AUTOMATIC OFFSET COMPENSATION SCHEME WITH PING-PONG CONTROL FOR CMOS OPERATIONAL AMPLIFIERS

4.1 Introduction

In many op-amp applications, offset cancellation or reduction is critical because an amplifier input offset voltage limits the capability of the system. An offset voltage of $10mV$ to $30mV$ is typical for CMOS amplifiers. This can not be tolerated in many applications. For continuous-time integrated applications, a number of offset cancellation schemes have been reported [6],[81],[82]-[89]. Classical approaches to build low-offset MOS op-amps through device optimization are inefficient and have performance limitations. To obtain low offset, special circuit techniques are additionally required. Commonly used auto-zero techniques use analog switches and capacitors to implement low-offset amplifiers. The offset cancellation of these techniques is degraded by the charge injection due to the autozero switches. The schemes have 50% duty cycles making them unsuitable for continuous-time applications.

In this chapter a digital correction technique is presented to keep the noise of the offset compensation circuit small. The objective is to compensate for inherent matching-induced offsets to achieve an op-amp with an offset voltage of less than 500μV. The proposed architecture is available to achieve even much lower offsets. A ping-pong architecture is employed to obtain a 100% duty cycle. With the ping-pong control the op-amp is capable of continuous-time operation, yet the offset is periodically adjusted making the offset compensation scheme insensitive to time and temperature drift. The scheme also requires no off-chip components and no adjustments during manufacturing. This compensation is obtained at the expense of modest extra chip area for the digital correction circuit. The scheme is most practical, from an area viewpoint, for large chips where many low offset op-amps are required. In these applications the digital correction circuit can be used in common, thus the area required for the digital correction circuit comprises a small fraction of the total die area.

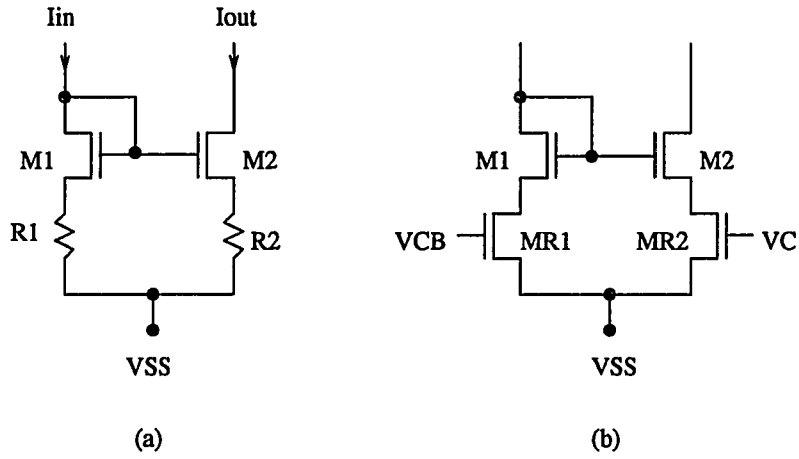


Figure 4.1: Programmable current mirror biased with (a) resistors (b) transistors

4.2 Offset Tuning Strategy

To adjust offset the programmable current mirror shown in Fig. 4.1 is used as the load of the differential input stage. The current mirror gain of the programmable current mirror in Fig. 4.1(a) is given by

$$A_i = \frac{I_{out}}{I_{in}} \simeq \frac{g_{m2}}{g_{m1}} \left(\frac{1 + g_{m1} R_1}{1 + g_{m2} R_2} \right) \tag{4.1}$$

The current mirror gain can thus be adjusted or programmed by changing the resistor values. A variable resistor can be implemented with a MOS transistor which is biased to operate in a linear region as shown in Fig. 4.1(b). For the case of $V_{DS} << V_{GS} - V_T$, MR1 and MR2 behave as linear resistors of value

$$1/R \simeq \mu C_{OX} \left(\frac{W}{L} \right) (V_{GS} - V_T) \tag{4.2}$$

By changing the bias voltages VCB and VC, the resistor values and thus the current mirror gain can be adjusted.

The adjustable range of the current mirror gain varies with the device sizes of MR1 and MR2 as shown Fig. 4.2. A small W/L increases the resistor value and thus increases the adjustable range of the current mirror gain which is directly related to the offset adjustable range of the op-amp. Therefore, the sizes of MR1 and MR2 should be selected according to the expected offset voltage range of the op-amp to be compensated.

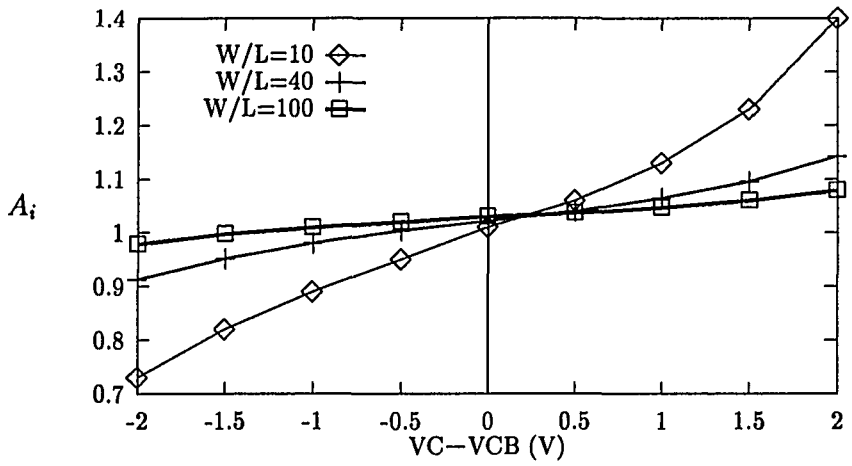


Figure 4.2: Simulated current mirror gain adjustable range of the programmable current mirror

These kinds of adjustable current mirrors have been used for auto-zero offset compensation [81],[83], where the op-amp output is fed back to a control port (e.g. VC) of a programmable current mirror during auto-zero periods, and the compensation voltage is stored on a capacitor to be utilized during signal processing periods. This analog scheme is very simple and requires small area, but its performance is limited by two factors. First, the compensation voltage is actually the op-amp output offset which is small but can not become zero. Secondly, the compensation voltage is contaminated by the charge injection of analog switches. Thus, there exists a lower limit of compensation at given supply voltages although it can be optimized by carefully selecting the op-amp gain and/or the gain from the control voltage to the op-amp output. The lower limit can be further reduced by using a compensation voltage generated from a digital correction circuit instead of one fed directly from the op-amp output. The digital scheme does not suffer from the charge injection problem. The performance of digital compensation is limited mainly by the resolution of the control voltages, and thus much smaller offset voltages are obtainable at the expense of more chip area.

Fig. 4.3 shows an offset adjustable two-stage CMOS op-amp with a programmable current mirror as the load of the input stage. The op-amp has been designed for high-speed and high-precision applications in a 1.0-μm CMOS technology. Since the first stage has the dominant

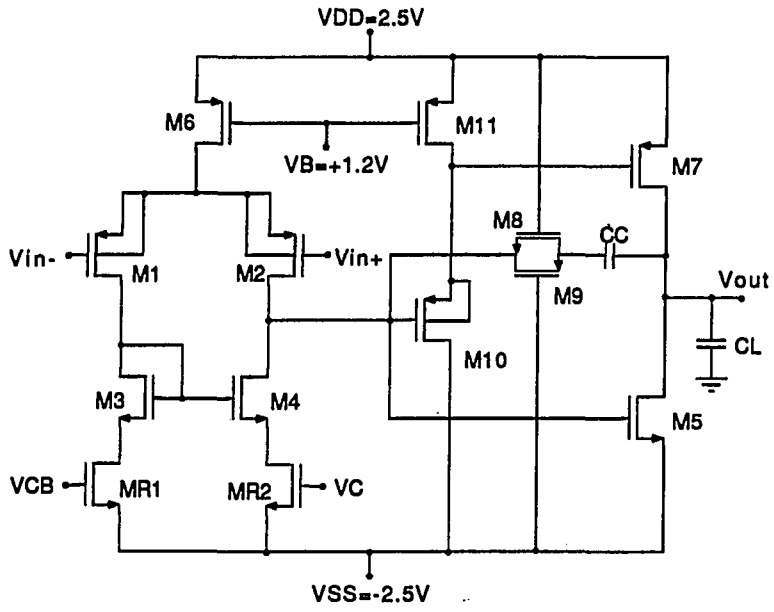


Figure 4.3: Offset adjustable two-stage CMOS op-amp with a programmable current mirror

effect on the offset, the input referred offset voltage of the op-amp can be expressed as

$$V_{OS,in} = V_{OSi} + V_{OSl} + V_{OSr} + V_{OS,sys} \tag{4.3}$$

where V_{OSi}, V_{OSl}, and V_{OSr} are the input referred random offset voltages due to the mismatches of the pairs (M1, M2), (M3, M4), and (MR1, MR2), respectively, and $V_{OS,sys}$ is the systematic offset voltage. If VC=VCB, the systematic offset is usually small and can be reduced to a very small value by careful design. Clearly, $V_{OS,sys}$ is a function of the bias voltage VC if VCB is fixed. For appropriate device sizes, there exists a certain value VC such that the total offset voltage $V_{OS,in}$ is zero. The random offset voltage can thus be compensated by intentionally introducing an offsetting systematic offset voltage that is dependent upon VC.

Assuming VC=VCB, the random offset voltages V_{OSi}, V_{OSl}, and V_{OSr} can be obtained as in the previous chapter or in [5]. The standard deviation of the sum of V_{OSi}, V_{OSl}, and V_{OSr} is given by

$$\sigma_{V_{OS,in}} = \frac{|V_{GSi} - V_{Ti}|}{\sqrt{2}} \left[\sigma_L^2 \left(\frac{1}{L_i^2} + \frac{1}{L_l^2} + \frac{1}{L_r^2} \right) + \sigma_W^2 \left(\frac{1}{W_i^2} + \frac{1}{W_l^2} + \frac{1}{W_r^2} \right) + \frac{4\sigma_{V_{Ti}}^2}{(V_{GSi} - V_{Ti})^2} + \frac{4\sigma_{V_{Tl}}^2}{(V_{GSl} - V_{Tl})^2} + \frac{\sigma_{V_{Tr}}^2}{(V_{GSr} - V_{Tr})^2} \right]^{\frac{1}{2}}, \tag{4.4}$$

where the subscript i denotes the input transistors M1 and M2, the subscript l denotes the load transistors M3 and M4, and the subscript r denotes the transistors MR1 and MR2 used as resistors. The standard deviation of the input offset voltage of the op-amp in Fig. 4.3 can be calculated based on (4.4). In this calculation

$$\sigma_L = \sigma_W = 0.014\mu m \tag{4.5}$$

$$\sigma_{V_T} = \frac{0.0236}{\sqrt{LW}} \tag{4.6}$$

were used as in the previous chapter. The designed transistor sizes, $L_i = L_l = L_r = 1\mu m$, $W_i = 240\mu m$, $W_l = 86\mu m$, and $W_r = 16\mu m$, and the simulated excess voltages, $V_{GSi} - V_{Ti} = -0.626V$, $V_{GSl} - V_{Tl} = 0.427V$, and $V_{GSr} - V_{Tr} = 3.753V$ were used for the calculation. The calculated $\sigma_{V_{OS}}$ is 12.2mV. This is somewhat high due to the short channel lengths of the input-stage transistors. The minimum sizes of the channel lengths were selected to obtain the fast settling characteristics of the op-amp.

To obtain a 99.7% offset yield, the sizes of MR1 and MR2 will be determined such that the offset voltage of $\pm 3\sigma_{V_{OS}}$ can be covered by adjusting VC. For the present design the ratio of (16/1) was selected to be more conservative for the achievable offset resolution. With this selection the offset adjustable range of the op-amp was simulated using a unity gain configuration. The results are shown in Fig. 4.4. An offset adjustable range from -16.2mV to $+19.6$mV can be obtained by changing the bias voltage VC from 1.5V to 2.5V with VCB fixed to 2V. This offset adjustable range which is equivalent to $-1.33\sigma_{V_{OS}}$ to $+1.61\sigma_{V_{OS}}$ leads to a 85.5% offset yield, provided the mean of the offset voltages is zero. The lower and upper limits of the bias voltage are determined by the reference voltages of the D/A converter (DAC) which will be discussed later. A wider adjustable range can be obtained by changing the DAC reference voltages, but the linearity will be degraded due to the nonlinearity of the NMOS resistors. The resolution will also be degraded by increasing the range of the DAC reference voltages at a fixed number of bits.

A simple way to find VC such that the total input offset voltage $V_{OS,in}$ becomes zero is as follows:

1. The output voltage of the op-amp is compared to zero when both the input terminals are grounded.

2. The bias voltage VC of the transistor MR2 is adjusted in the direction of reducing the output offset voltage.

3. The procedure is repeated until the op-amp output voltage crosses zero.

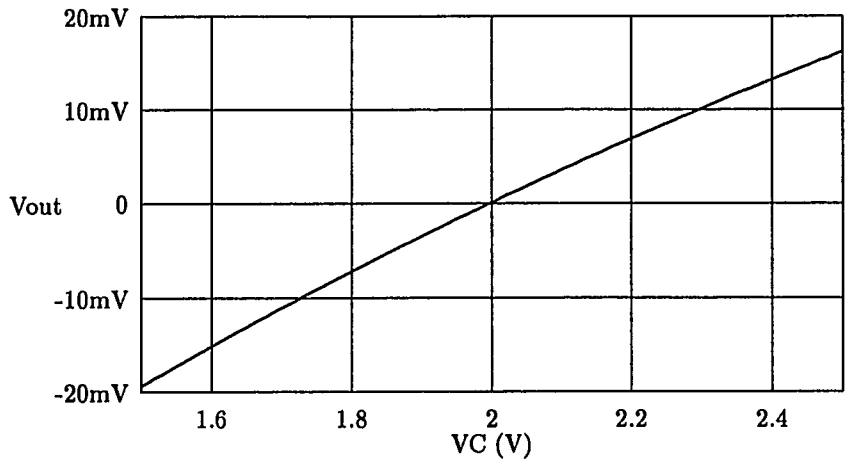


Figure 4.4: Simulated offset adjustable range of the op-amp in Fig. 4.3. VCB is fixed to 2V

This offset tuning strategy is depicted in Fig. 4.5. The up/down counter is initially set to half of full scale such that the DAC output VC is equal to the fixed bias voltage VCB. The performance of the offset reduction depends upon the resolution of the bias voltage VC and the offset of the comparator.

Fig. 4.6 shows a simplified block diagram of the entire offset compensated op-amp. It consists of three blocks: an op-amp block, a timing signal generator, and an offset tuning block. The op-amp block consists of two identical op-amps and several analog switches for ping-pong operation. The timing signal generator produces signals, P1-P4, to control the ping-pong structure. The offset tuning block will reduce the offset voltages of the op-amps by adjusting the bias voltages VC1 and VC2. Circuit details and functions of the blocks are presented in the following sections.

4.3 Op-amp Block with a Ping-Pong Structure

The op-amp block diagram is shown in Fig. 4.7. The block consists of two identically designed op-amps of Fig. 4.3 and several switches which are used for implementing a ping-pong structure. VC1 and VC2 are the bias voltages of the transistor MR2 of OPA1 and OPA2, respectively. One of the op-amps will be in a normal mode at any one time while the other is in an offset tuning mode. A 100% duty cycle can be obtained by interchanging their roles. The

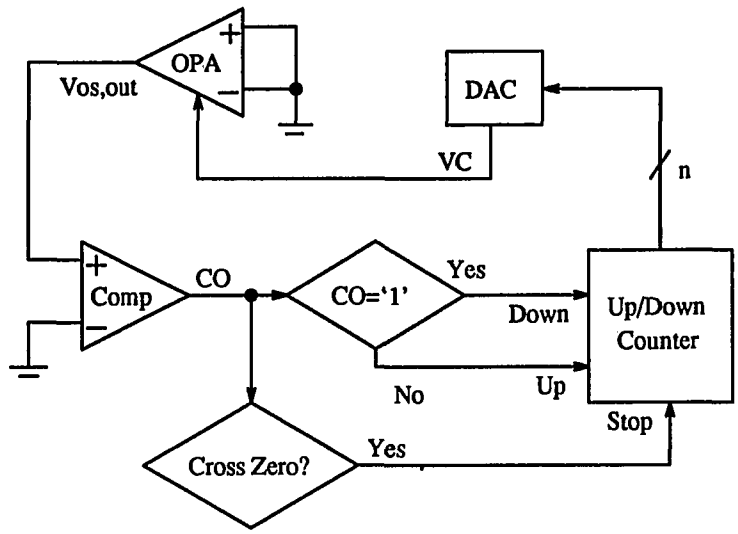


Figure 4.5: Concept of the offset tuning scheme

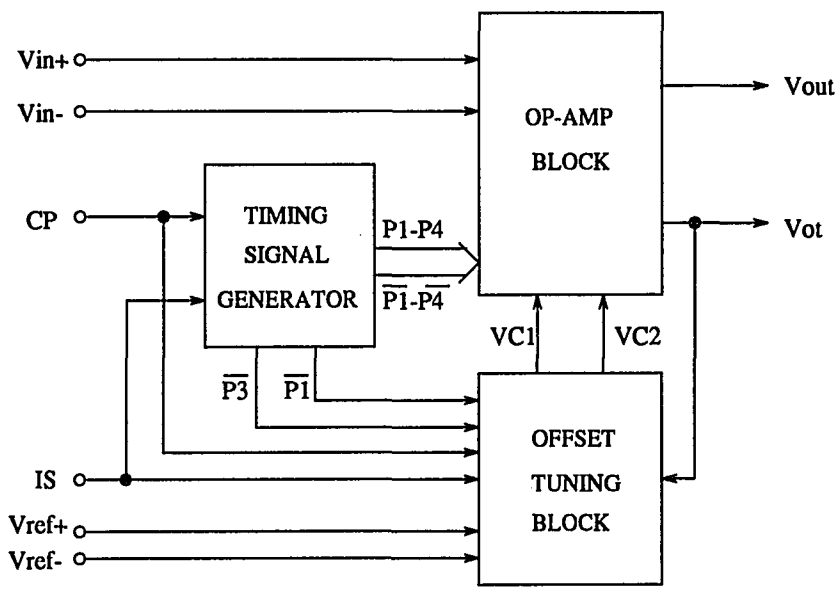


Figure 4.6: Block diagram of the digitally offset compensated op-amp

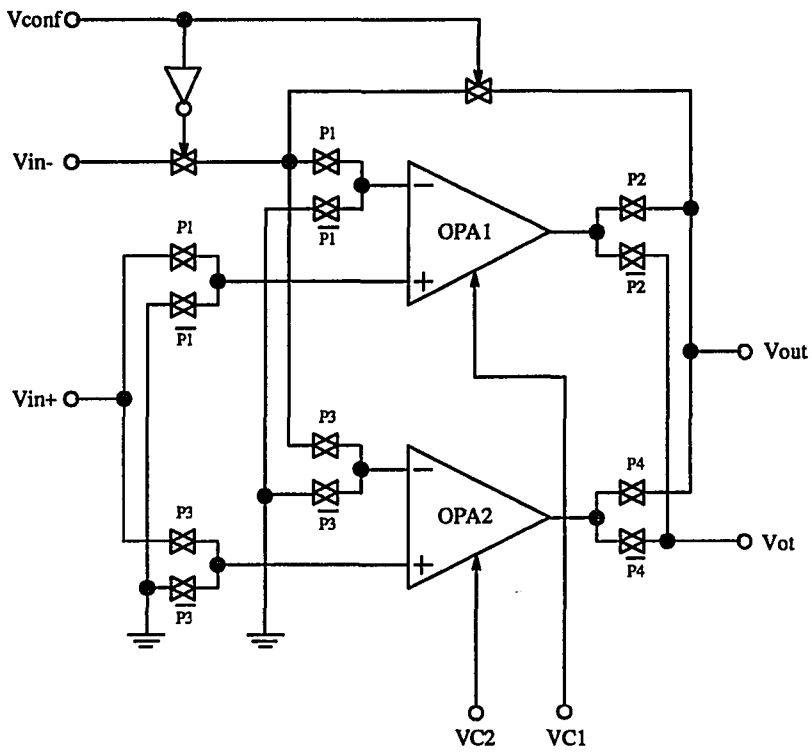


Figure 4.7: Op-amp block diagram (Ping-Pong structure)

signal V_{conf} is used to configure the op-amps either in an open-loop or in a closed-loop with unity gain. The other switches are controlled by the timing signals, P1-P4, generated by the timing signal generator to achieve the ping-pong operation. A similar operation was introduced in [90],[91], where two identical resistors which are implemented with NMOS transistors biased in the ohmic region and capacitors are alternatively tuned to obtain accurate RC products for continuous-time filters. Simulated performance of the uncompensated op-amp block is shown in Table 4.1. The SPICE simulation was performed with the circuit extracted from its layout.

Fig. 4.8(a) shows the timing diagrams of P1-P4. During the first phase, OPA1 processes the input signal while OPA2 is in the offset tuning mode. A phase is defined here as the time duration from the moment that two op-amps interchange their roles to the next interchanging moment. Each phase consists of 2^{n-1} (128 for n=8) clock (CP) periods. The offset tuning block uses V_{ot} which is the output offset voltage of OPA2 at this time to generate an offset

Table 4.1: Simulated performance of the op-amp block of Fig. 4.7

Specification	Performance
Settling Time (−1V to 1V Step Input)	
0.2% Error Limit, C_L=1PF	21.6 nsec
Input Systematic Offset Voltage	160 μV
Open Loop Voltage Gain	700 (56.9 dB)
Unity Gain Frequency	113 MHz
Phase Margin	61°
Input Common Mode Range	±2.0 V
Output Voltage Swing	±1.9 V
Power Dissipation	45.9 mW
CMRR	56 dB

control signal VC2. When the op-amps interchange their roles, P3 first goes to high at the 128th CP falling edge, such that the signal input is also connected to OPA2. After that, other three timing signals, P1, P2, and P4 change their states after one CP period. At this time the transient in OPA1 to be used for tuning does not affect the tuning process because updating the bias voltage VC1 is made after one clock period, and the one clock period is made long enough for the op-amp to finish its transient.

An expected offset voltage waveform is shown in Fig. 4.8(b). During the reset (when the signal 'IS' is '0') the states of the four timing signals are (P1,P2,P3,P4)=(1,0,1,1), and thus, the output of the op-amp block V_{out} will be the initial uncompensated output offset voltage of OPA2 if the inputs are grounded. Of course, the magnitude of the offset voltage depends on the configuration of the op-amp. During the first phase, i.e., (P1,P2,P3,P4)=(1,1,0,0), V_{out} will be the uncompensated output offset voltage of OPA1, and the offset of OPA2 is compensated by the offset tuning circuit. Therefore, during the second phase (P1,P2,P3,P4)=(0,0,1,1), the compensated offset voltage of OPA2 will appear at the output while the offset of OPA1 is being adjusted. The compensated offset voltage of OPA1 can be thus found during the third phase. In the following phases only the changes of the offset due to the temperature and time drift will be compensated.

The compensated offset voltages of the two op-amps can be different from each other, so the equivalent offset voltage of the op-amp block after calibration can be defined by

$$V_{OSeq} = max\{|V_{OS1}|, |V_{OS2}|\} \tag{4.7}$$

The ping-pong operation makes it possible that the op-amp can operate in a continuous-time mode while the offset voltages are kept small after the calibration time that corresponds to

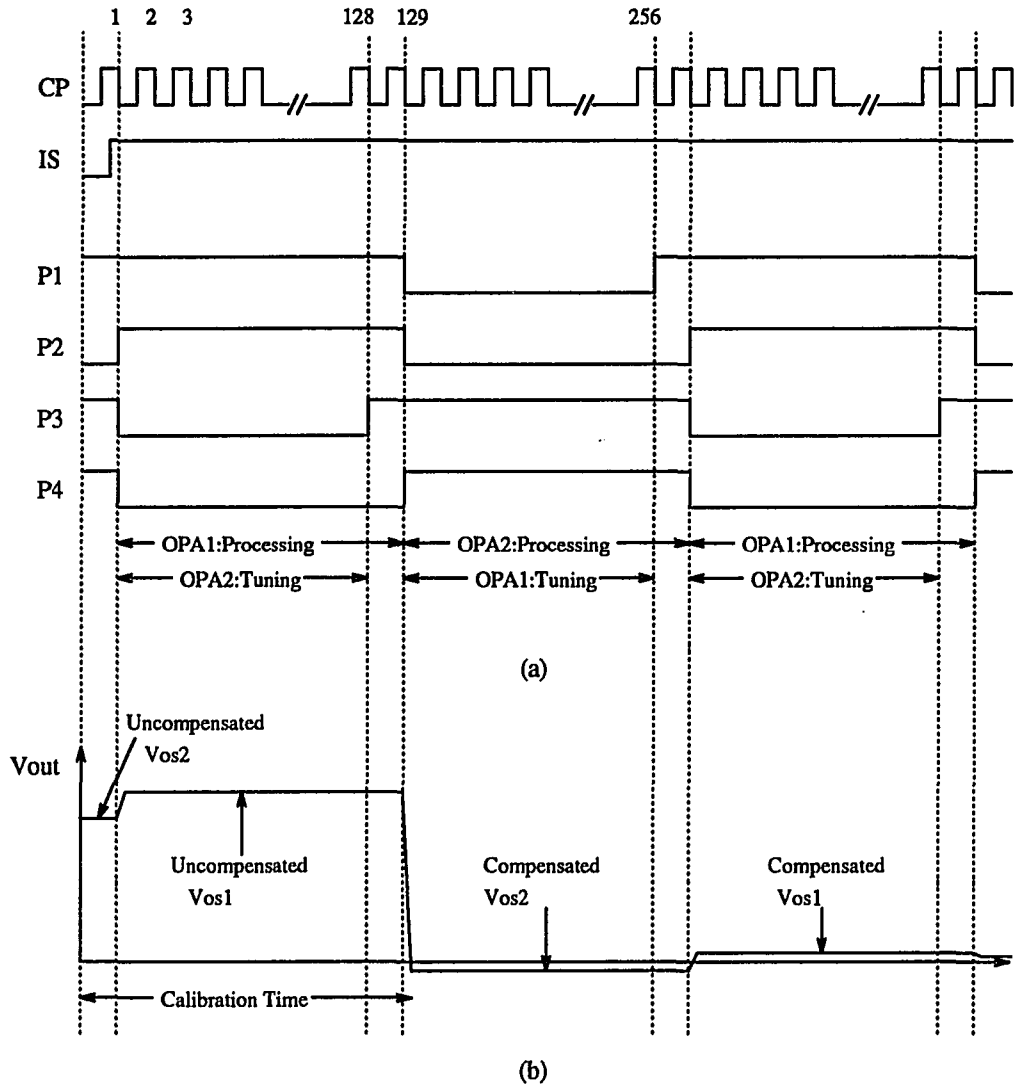


Figure 4.8: (a) Timing diagram of the ping-pong structure (b) An expected offset voltage waveform of the op-amp block

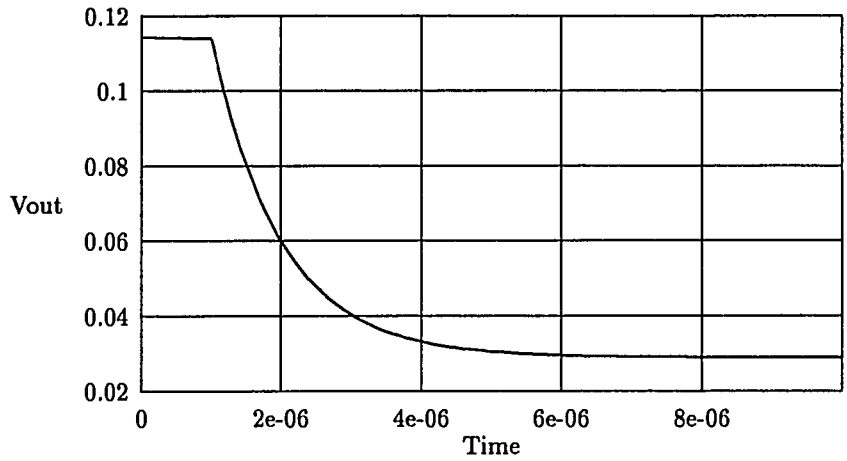


Figure 4.9: Output delay of the op-amp block with respect to VC change

128 CP periods. The temperature drift of the offset voltages can be also compensated by the ping-pong operation.

Fig. 4.9 shows the simulated output delay of the op-amp block when VC is changed. The initial output voltage of the open-loop op-amp block was assumed to be 0.114V which is due to the systematic input offset voltage. The random offset was not considered. To reduce the offset, VC was decreased from 2V to 1.996V at t=1μsec. The step, about 4mV, corresponds to the resolution of the DAC, i.e., 1LSB of 8-bit. The VC change causes the reduction of the output voltage, and thus, the input offset voltage. The delay time of the op-amp block with respect to the VC change is about 5μsec. Thus, one CP period must be longer than the delay time to correctly update the bias voltage VC.

4.4 Offset Tuning Block

The block diagram of the offset tuning block is shown in Fig. 4.10. It consists of a comparator, a zero crossing detector (ZCD), two 8-bit up/down counters (UDC block), and two small 8-bit D/A converters (DAC). This block detects the output voltage V_{ot} of the op-amp to be tuned and then provides an updated bias voltage VC1 for OPA1 and VC2 for OPA2 such that the offset voltages are reduced. The timing signals, $\overline{P1}$ and $\overline{P3}$, generated from the timing signal generator determine which op-amp will be tuned, so only one of the two up/down counters is enabled to count. The counters are initially set through the signal IS to half of full

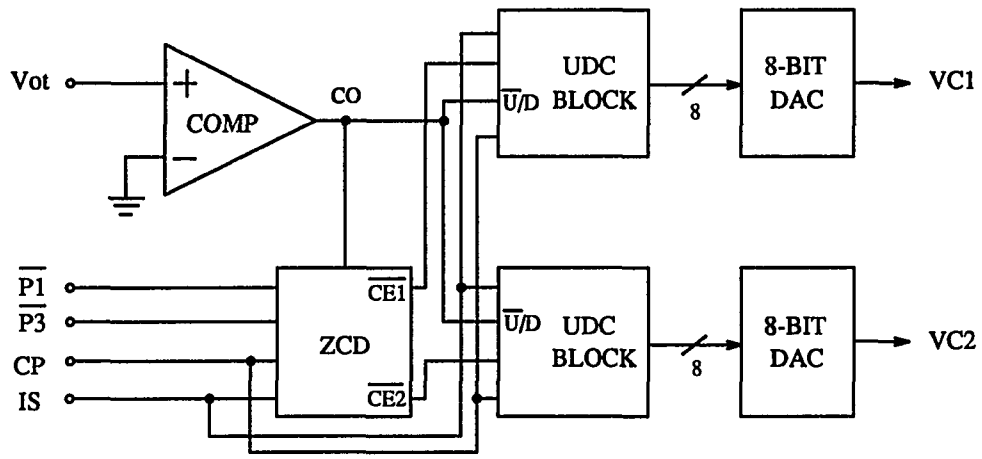


Figure 4.10: Block diagram of the offset tuning block

scale. This is done to accommodate for the inherent bipolarity of the offset voltages.

If the comparator output is high, indicating that the offset is greater than zero, then the down signal of the up/down counter is set to '1', so that the counter counts down to decrease the bias voltage VC. The current mirror gain of the programmable current mirror is then decreased, and the op-amp output voltage is also decreased, i.e., the offset voltage is reduced. As long as the comparator output does not change, this procedure is repeated until $\overline{P1}$ or $\overline{P3}$ is changed. If the comparator output goes to '0' before the phase is changed, then the zero crossing detector detects this change and sets the count enable signal CE to '0' to prevent the op-amp output from oscillating. A change of the comparator output means that the op-amp output crosses zero, and the minimum offset is achieved. Thus, no further update of the bias voltage is required.

4.4.1 Digital-to-Analog Converter

A simple R and 2R resistor ladder network shown in Fig. 4.11 is used for an 8-bit DAC. The resistor ladder is implemented by PMOS transistors with the W/L ratios of 2/5 for R and 2/10 for 2R. The ratio of the PMOS transistors used for the decoding switches is 10/1. The binary signals D0-D7 are the outputs of the up/down counter. One advantage of this simple structure is that the area is very small compared to other structures and increases only linearly with the number of bits. For the proposed offset compensation scheme the DAC does not require

excellent linearity because some degree of nonlinearity can be tolerated unless the resolution is significantly degraded. Even nonmonotonicity in the DAC can be tolerated. Nonlinearity and nonmonotonicity result in minor degradation of the resolution without affecting the correct operation of the offset compensation. The simple structure has been chosen to keep the area small.

The simulated output voltages of the designed DAC at different digital settings are shown in Fig. 4.12, where V_{ref+} =2.5V and V_{ref-} =1.5V. The simulated result shows that the DAC has the expected nominal nonlinear characteristics. The DAC output voltage at digital setting 128 is 2.04V instead of 2.0V due to the nonlinearity. Fig. 4.4 shows that the op-amp output also exhibits a modest nonlinear relationship between the bias voltage of the programmable current mirror and the offset voltage. Due to the nonlinearity of both the DAC and the programmable current mirror the simulated worst-case resolution is 0.22mV when the offset adjustable range is −16.2mV to +19.6mV, which is degraded from the theoretical resolution of 0.14mV(=35.8mV/256) but satisfies the targeted resolution of 500μV.

The resolution can be readily improved by increasing the number of bits of the DAC and the up/down counters at the cost of a small increase in die area. If a 10-bit DAC is used, then a theoretical resolution of 35μV(=35.8mV/1024) can be obtained at the cost of two more flip-flops for a counter and 8 more PMOS transistors for a DAC. Another way to improve the resolution is to reduce the range of the DAC reference voltages. This will, however, reduce the offset adjustable range.

4.4.2 Comparator

A simple two-stage comparator with the output buffered is employed to compare the op-amp output offset voltage to zero. The designed device sizes are $W_i/L_i = 28/2$ and $W_l/L_l = 9/4$. The standard deviation of the random offset voltage of the designed comparator was calculated based on (4.4), resulting in $\sigma_{V_{OS}}$=6.5mV. This comparator offset voltage can be tolerated because the op-amp in an offset tuning phase is in an open-loop configuration (see Fig. 4.7), and thus, the output offset voltage preserved at the input of the comparator is 700 (the open-loop gain) times greater than the input offset voltage of the op-amp. Since 8-bit DACs are used for the present design, and the simulated resolution of the offset adjustment is 0.22mV, the minimum output voltage of the op-amp which must be resolved by the comparator is 154mV (0.22mV×700) which is much greater than the comparator offset voltage. Thus, only a small fraction of the comparator input offset which is the comparator input offset voltage divided by the open-loop gain of the op-amp will contribute to degradation in the achievable offset

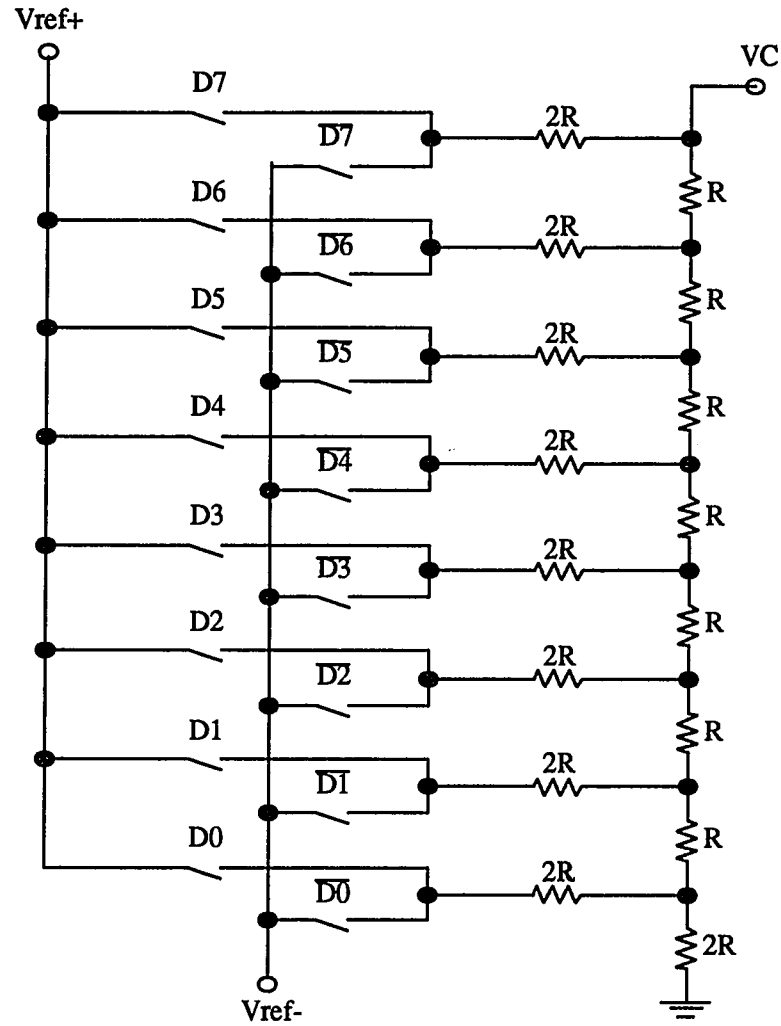


Figure 4.11: R and 2R resistor ladder 8-bit D/A converter

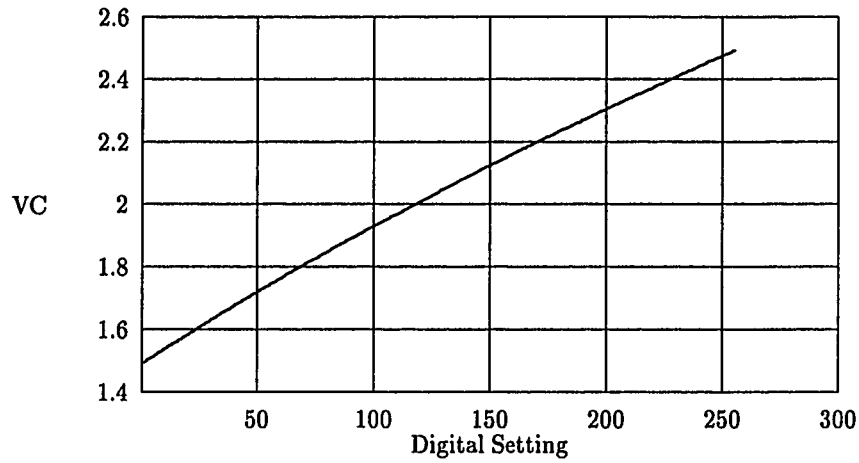


Figure 4.12: Simulated DAC output voltages at 256 digital settings ($V_{ref+} = 2.5$V and $V_{ref-} = 1.5$V)

resolution.

4.5 Experimental Results

The test circuit was fabricated in a 1.0-μm n-well CMOS process. The chip photomicrograph is shown in Fig. 4.13. The total circuit area excluding pads is $0.99mm^2$. The op-amp block occupies 14.7% of the total area, and the comparator and the two DACs occupy 0.84% and 2.12%, respectively. The remaining 82.3% is for the digital control circuits and connections. The extra large area of the digital section can be compensated somewhat by using it in common for several op-amps. One simple example is to use time sharing operation. For the case of using two op-amp blocks, only one up/down counter can be used along with 4 DACs by including latches before DACs, and other blocks can remain unchanged. With additional multiplexing circuits and connections for time sharing operation, one digital tuning circuit can serve for two op-amp blocks such that two among four op-amps are always available for signal processing while one of the remaining two op-amps is in an offset tuning mode. This scheme can be extendable for circuits including more op-amp blocks.

The CP frequency is set to 46.9 kHz ($T_{CP} = 21.3\mu s$). The period of one phase is thus

$$T_p = 128T_{CP} = 2.73ms$$

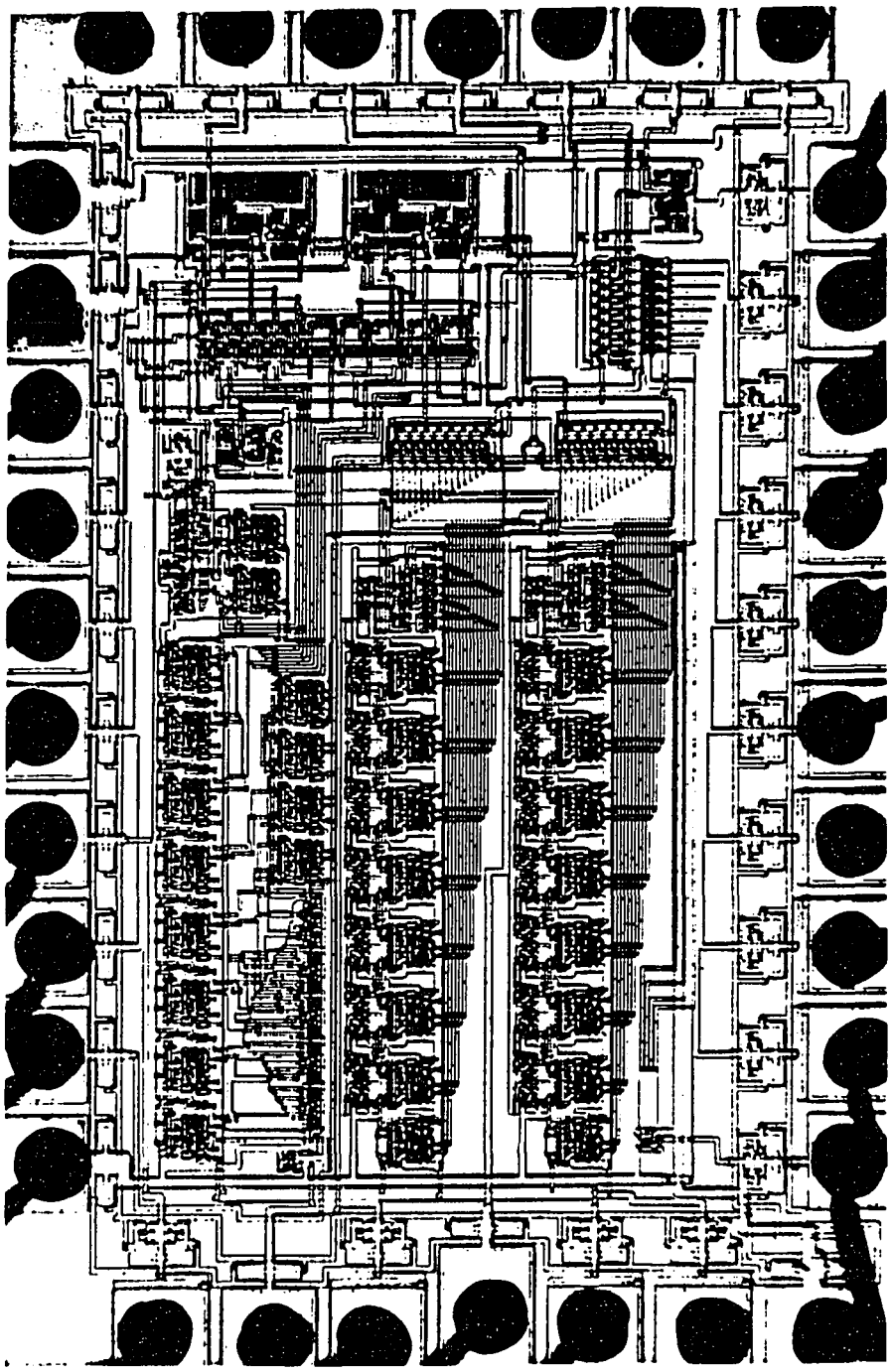


Figure 4.13: Chip photomicrograph

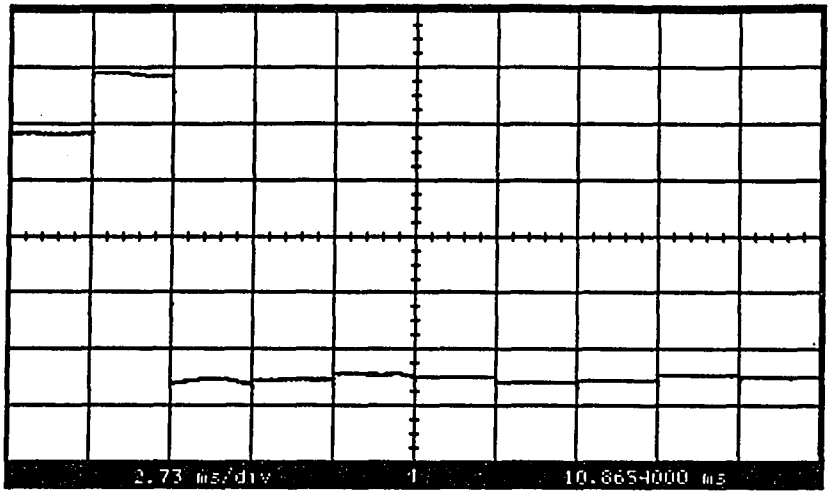
A closed-loop configuration with a gain of 100 is used to characterize the offset voltage waveform of the op-amp block. The closed loop configuration is obtained by connecting two resistors ($R1 = 1K\Omega$ and $R2 = 100K\Omega$) around the inverting input terminal, when the signal 'V_{conf}' in Fig. 4.7 is set to '0'. With the inputs grounded 100 times the input offset voltage will appear at the output terminal V_{out}. In this case a periodic square waveform that is generated from a function generator is used for the reset ('IS') signal, and this serves as a triggering signal to help an oscilloscope catch the nonperiodic offset voltage waveforms more easily.

4.5.1 Measured Offset Waveforms

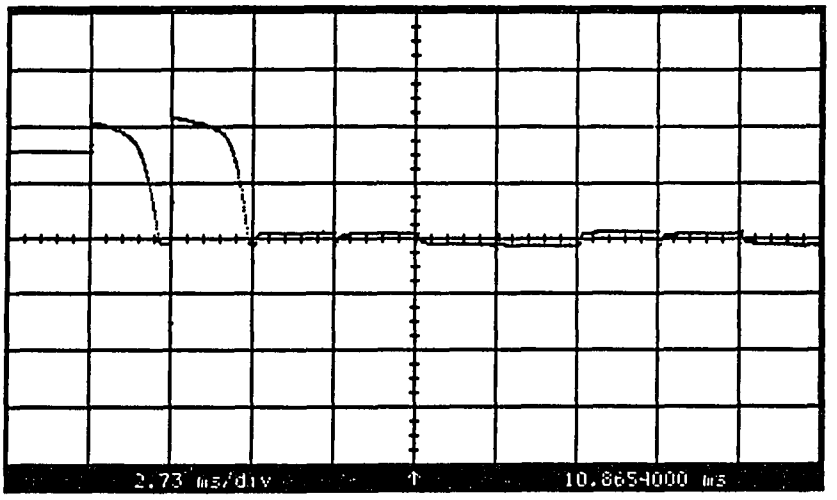
A typical output offset waveform measured from one of the test chips is shown in Fig. 4.14(a). The waveform was obtained from the closed-loop feedback amplifier in the gain of 100 configuration. The reset signal is changed from '0' to '1' at $t = 2.73ms$ when $t = 0$ is referenced to the left edge of the trace. The horizontal scale is 2.73 ms/div which is the period of one phase T_p. The vertical scale is 200 mV/div, and the vertical axis offset is 500 mV. It can be seen that the measured offset waveform is very similar to the expected waveform shown in Fig. 4.8(b). The first two high states are due to the uncompensated offset voltages of OPA2 and OPA1, respectively. The following low states are the compensated offset voltages. The initially uncompensated input offset voltages and the compensated offset voltages can be obtained from the plots by dividing the output voltages by the closed-loop gain of 100.

By connecting V_{ot} (see Fig. 4.7) instead of V_{out} to the oscilloscope input, the offset compensation process can be observed more clearly because V_{ot} is the output offset voltage of the open-loop op-amp being tuned, and this signal is compared with zero by the comparator and finally reduced by the offset tuning circuit. The measured V_{ot} is shown in Fig. 4.14(b). From the plot it can be seen that the initial uncompensated large offset voltages of OPA2 and OPA1 are reduced significantly during the first and second phases, respectively. It can be also seen that because of the large uncompensated input offset voltages and the large open-loop gains of the op-amps, the outputs of the open-loop op-amps initially saturate.

During the first phase the initial offset voltage of OPA2 is continuously decreased, and the offset compensation is stopped when the offset crosses the zero line. Thus the offset sits at a small negative level. In the next OPA2 tuning phase, i.e., the third phase, the tuning circuit detects the negative polarity of the offset and starts to adjust the offset toward the positive direction. After the initial tuning, a single one-step adjustment (1LSB change of the DAC) is usually enough to make the offset cross the zero line. The zero crossing detector then makes the adjustment stop. This process is repeated in the following tuning phases. Therefore, each



(a)



(b)

Figure 4.14: (a) Output offset waveform measured at V_{out} with the op-amp having a closed-loop gain of 100 (VS: 200mV/div, Offset: 500mV) (b) Output offset waveform measured at V_{ot} which is the output offset voltage of the open-loop op-amp being adjusted (VS: 1V/div, Offset: 0V)

Table 4.2: Measured initial and compensated offset voltages

#		Initial V_{OS}	Final $\|V_{OS}\|$	#		Initial V_{OS}	Final $\|V_{OS}\|$
1	OPA1	9.9 mV	157 μV	2	OPA1	15.0 mV	150 μV
	OPA2	8.9 mV	191 μV		OPA2	11.2 mV	309 μV
3	OPA1	−6.3 mV	219 μV	4	OPA1	9.5 mV	395 μV
	OPA2	2.4 mV	97 μV		OPA2	11.6 mV	318 μV
5	OPA1	8.0 mV	72 μV	6	OPA1	5.0 mV	245 μV
	OPA2	6.3 mV	158 μV		OPA2	2.3 mV	57 μV
7	OPA1	−370 μV	172 μV	8	OPA1	8.4 mV	102 μV
	OPA2	8.5 mV	261 μV		OPA2	7.4 mV	185 μV
9	OPA1	13.8 mV	313 μV	10	OPA1	590 μV	299 μV
	OPA2	12.6 mV	163 μV		OPA2	1.5 mV	269 μV

op-amp will have two compensated offset levels, i.e., one is positive and the other is negative as shown in Fig. 4.14(b). Thus, the compensated op-amp will have bipolar offset voltages. Unipolar offset compensation and correspondingly an overall decrease in offset voltages can be readily obtained by slightly modifying the control logic (the zero crossing detector block) such that compensated offset voltages approach the zero crossing from the same direction during each phase. Doing this, the compensated offset variation will be substantially reduced, and the corresponding offset voltages are reduced by a factor of 2.

4.5.2 Offset Compensation Results

The compensated offset voltages are measured from V_{ot} by expanding the vertical scale and dividing the values by the measured open-loop gain. This will be more accurate for small offset voltages than measurements from a feedback amplifier at V_{out} because of the large open-loop gain. The measured open-loop gains are between 400 and 700. The initial large offset voltages were, however, measured from V_{out} using closed-loop configurations with proper closed-loop gains which were selected low enough to guarantee that the op-amp outputs do not saturate.

Of 13 chips tested, three showed initial offset voltages outside the compensatable range. This can be expected because of the low designed offset yield as mentioned in Section 4.2. The measured initial offset voltages and the compensated final offset voltages are shown in Table 4.2. Since each op-amp has two compensated offset levels as mentioned above, the greater one in magnitude is reported in the table. The table shows that most op-amp initial offset voltages are biased in the positive direction. This suggests a small wafer-level and/or die-level systematic offset.

All compensated offset voltages are less than 400μV in magnitude which met our design specification of 500μV. These offsets are, however, somewhat degraded from the simulated resolution of 220μV. The degradation is attributable, in part, to the systematic offset which causes the final steps of the offset control voltage VC to be placed around the lower left corner of the curve in Fig. 4.4 where the resolution is degraded. The nonlinearity of the fabricated circuit may be more severe than the simulated one. The comparator offset also partially contributes to the degradation in offset voltages.

Although the resolution can be improved by a more careful layout, the systematic offset can be reduced, and the variations of the random offset voltages can be reduced by using a more linear portion of the plot in Fig. 4.4, an easier way to improve the resolution without reducing the offset adjustable range is to increase the number of bits of the DAC. In this case the comparator offset will ultimately become the dominant factor limiting resolution. Further improvement with higher-bit DACs can be achieved by compensating the comparator offset.

4.5.3 Transient Characteristics

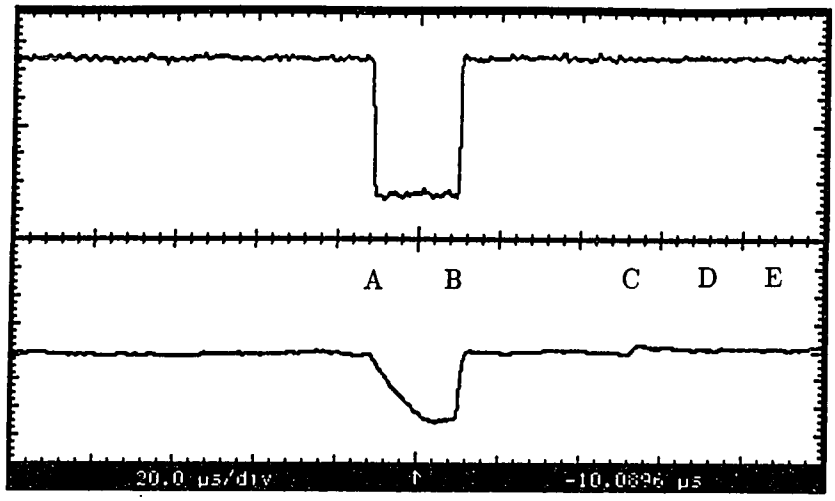
In many applications there are brief periods of time where the amplifier need not be operational, and in such applications the transient responses associated with switching the op-amp from the compensation state into the application state are not of concern since this switching can occur during these brief periods. Furthermore, the re-compensation rate can be very small ranging from minutes to days on even weeks in many environments. The effects of this switching transient even in true continuous operation are, however, very small.

The measured transient characteristics due to the ping-pong operation are shown in Fig. 4.15. The upper waveform of Fig. 4.15 (a) and (b) was used as the triggering signal which will be denoted as S_T, where

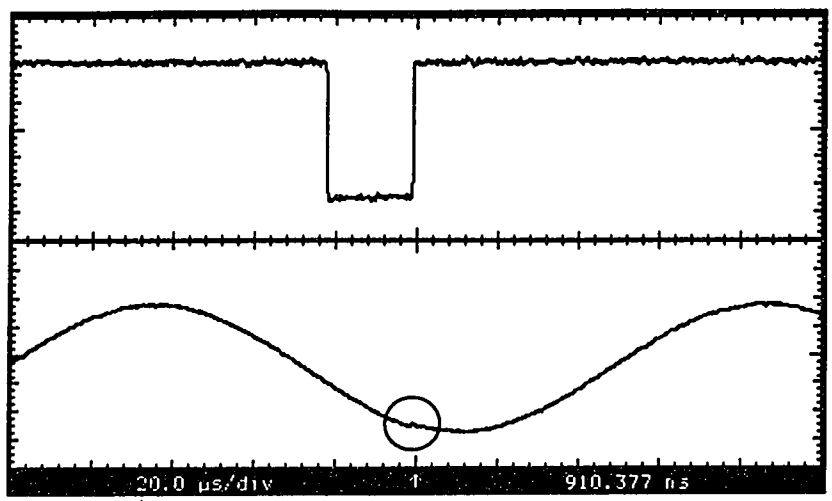
$$S_T = \overline{P1 \cdot P3}.$$

Therefore, as can be seen in Fig. 4.8(a), the period of S_T is $128T_{CP}$, and the duration of S_T='0' is T_{CP}. At time A the inputs of the op-amp which have been connected to ground for offset tuning are switched to the input signal. After one CP period the two op-amps interchange their roles completely at time B.

The lower waveform of Fig. 4.15(a) is V_{ot} which is the output of the op-amp in an offset tuning mode. The waveform V_{ot} exhibits the compensated output offset voltage of one op-amp until time A and shows the output offset of the other op-amp after time B. It can be seen that since the compensation of the op-amp which is in a new tuning phase is started after one CP

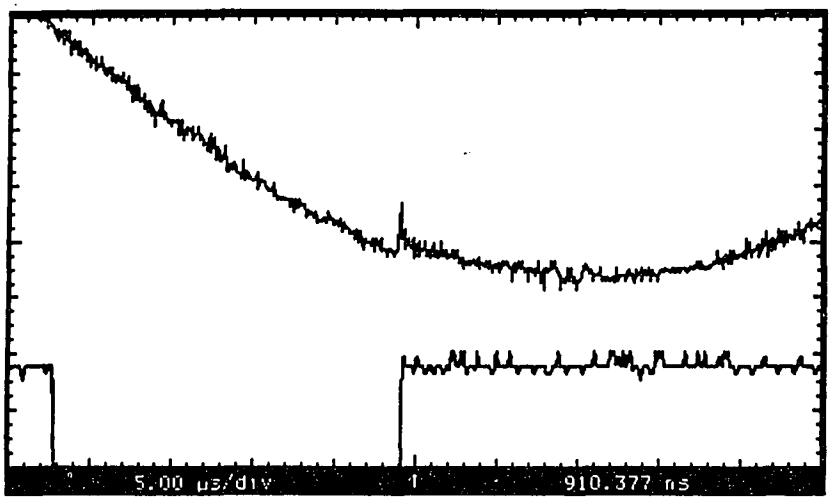


(a)



(b)

Figure 4.15: Measured transient characteristics (a) Output offset voltage measured at V_{ot} (Lower trace, VS: 500mV/div) (b) Output signal measured at V_{out} when the op-amp is in a closed-loop configuration with a gain of 33 (Lower trace, VS: 500mV/div)



(c)

Figure 4.15: (continued) (c) Magnified plot of (b) (Upper trace, VS: 200mV/div, Offset: -1.24V). The triggering signal S_T is the upper trace of (a) and (b) and the lower trace of (c) (VS: 1V/div)

period as mentioned in Section 4.3, the offset adjustment can be observed at time C which is two CP periods after time B. It can be also seen that no further offset adjustment is observed at the following CP falling edges such as time D and E since the offset crossed the zero line at time C, and after calibration a single one-step adjustment is usually enough to make the offset cross the zero line.

The lower waveform of Fig. 4.15(b) was measured at V_{out} when the op-amp block was in a closed-loop configuration with a gain of 33 and a sinusoidal input was applied. The high-gain configuration was used to examine the transient characteristics more clearly. With low-gain configurations the transients were hardly observed. The output transients can be observed inside the circle on Fig. 4.15(b). Fig. 4.15(c) is a magnified plot of the circle on Fig. 4.15(b). The displacement of the output signal is due to the difference between the offset voltages of two op-amps. This displacement can be greatly reduced by using the unipolar offset compensation scheme as mentioned before.

From the experimental results it can be seen that the transient energy in the output due to the ping-pong operation is not significant. The transient energy can be further reduced if the following schemes are incorporated with the current structure:

1. Modifying the switching process such that the output of the op-amp to be used for signal processing follows the output of the op-amp which is currently being used. This can be done by connecting the output V_{out} with the input of the op-amp in a unity-gain configuration as shown in Fig. 4.16. This tracking process can be done during S_T='0'.

2. Reducing the CP frequency greatly after the first two tuning phases. This can be readily obtained by modifying the control logic.

3. Disabling the ping-pong operation after the initial power-up calibration and enabling only on demand. This will also reduce the digital noise associated with the clock pulse CP.

Using these schemes, the transients will be almost negligible, and continuous-time operation can be achieved without any significant dynamic range loss.

4.6 Conclusion

An automatic digital offset correction method for continuous operation CMOS op-amps has been presented. A programmable current mirror is used to adjust the offset voltage. A ping-pong structure is employed to obtain a 100% duty cycle while the offset voltage is constantly kept small. The proposed offset compensation is not sensitive to time and temperature drift.

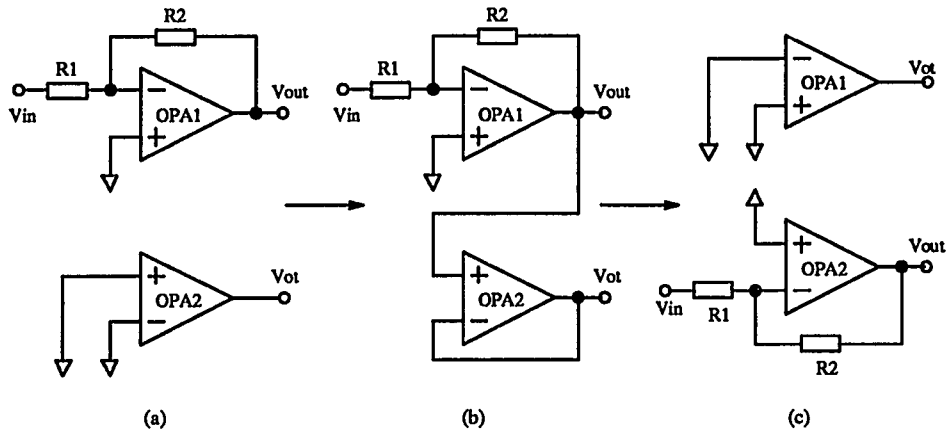


Figure 4.16: An example switching process to reduce the transient effects due to the ping-pong operation (a) OPA1: signal processing, OPA2: offset tuning (b) OPA2 tracks the output signal of OPA1 during S_T='0' (c) OPA1: offset tuning, OPA2: signal processing

Experimental results show that the designed op-amps can be digitally adjusted to have input offset voltages of less than 400μV in magnitude. The resolution can be substantially improved by increasing the number of bits of the DAC, doing layout more carefully, using a unipolar offset compensation scheme, and employing an offset compensated comparator. Experimentally measured transients due to the ping-pong operation are not significant. Several schemes have been proposed to further reduce the transient effects.

CHAPTER 5. VERY LOW VOLTAGE OPERATIONAL AMPLIFIERS USING FLOATING GATE MOS TRANSISTORS[2]

5.1 Introduction

With emergence of increasing number of battery-operated applications, great interest has been aroused in low voltage circuit techniques. The research efforts for low supply voltage based operation have been focused mainly on digital circuits [92], especially on the high density memory circuits such as DRAMs and SRAMs [93],[94].

The current technology trends for low voltage operation are paralleling the scaling of device feature sizes and threshold voltages. Very low voltage operation can be possible through device scaling if the threshold voltage can be scaled down in proportion to the supply voltage scaling [95]. A 0.1μm CMOS device called a low-impurity-channel transistor has been reported in [96], where the threshold voltage can be scaled down below 0.1V. The scaling of the threshold voltage, however, requires much more complicated technologies called "substrate engineering". One problem with the scaling is the increased threshold voltage variation. There also exists a lower limit in scaling down the threshold voltage because the scaled down devices experience problems such as increased leakage currents, short channel effects and parasitic effects which are much more severe than in large feature size devices.

On the other hand, a floating gate MOS transistor is capable of having a very low threshold voltage without device scaling and without any substrate engineering. The floating gate transistor (FGT) has been used primarily as a data storage device in EPROM and EEPROM circuits [97],[98]. Recently, however, the device has started to attract considerable interests as a nonvolatile analog storage device and as a precision analog trim element because it has the threshold voltage programmability with nearly infinite resolution as well as the long term charge retention. Experimental results have demonstrated that the threshold voltage of a test

[2] *Proceedings of IEEE International Symposium on Circuits and Systems,* Chicago, vol. 2, pp. 1152-1155, May 1993.

FGT can be adjusted in sub-milivolt range increments with a charge loss less than 2% in 10 years at room temperature [99].

Motivated by the unique and promising characteristics of the floating gate MOS transistors, a threshold voltage tunable op-amp structure for very low voltage (e.g. 0.5V) operation is presented in this chapter. To utilize the FGTs as op-amp circuit elements, their threshold voltages must be programmed and tuned. A two-step tuning method is used. One is coarse tuning, and the other is fine tuning. Two fine tuning methods are presented.

5.2 Threshold Voltage Tunable Op-amp Structure

To obtain an op-amp which can operate with a very low power supply, the threshold voltages of the floating gate MOS transistors composing the circuit must be reduced. The FGTs should also have very similar characteristics with those of conventional MOS transistors. This has been validated in the literature [102].

The structure and the programming operation principles of a floating gate MOS transistor are well known and will not be described in detail here, except to note that when a large enough field is present across the gate oxide, in most existing FGTs Fowler-Nordheim electron tunneling allows charge to be transferred to or from the floating gate, depending on the polarity of the field. The charge amount to be transferred depends on the magnitude and duration of the programming pulse that is needed to produce a large enough electric field in the tunnel oxide. Since charge transfer to or from the floating gate affects the threshold voltage of the FGT, three variables, the magnitude, polarity, and duration of the programming pulse, can be used to control the threshold voltage.

One method of tuning the threshold voltage entails placing the FGTs in an array as shown in Fig. 5.1. Each cell terminology is shown in Fig. 5.2. Each cell consists of 6 transistors: a FGT, three select transistors (SD, SG, and SS) and two switch transistors (S1 and S2). The select MOS transistors are required to tune the threshold voltage of the selected FGT only, and thus, the other FGTs that are not selected will not be affected by the tuning process. The switch transistors are used to connect or disconnect the FGT with other FGTs. A switch transistor at the source terminal of the FGT is not required since high voltages are not applied at the source terminal during the threshold voltage tuning. Although each cell has two switch transistors as shown in Fig. 5.1, two switch transistors are not always required for all cells. The number of switch transistors can be reduced and depends upon the circuit topology.

The threshold voltage (V_{th}) tunable op-amp circuit has two operating modes: a V_{th} tuning

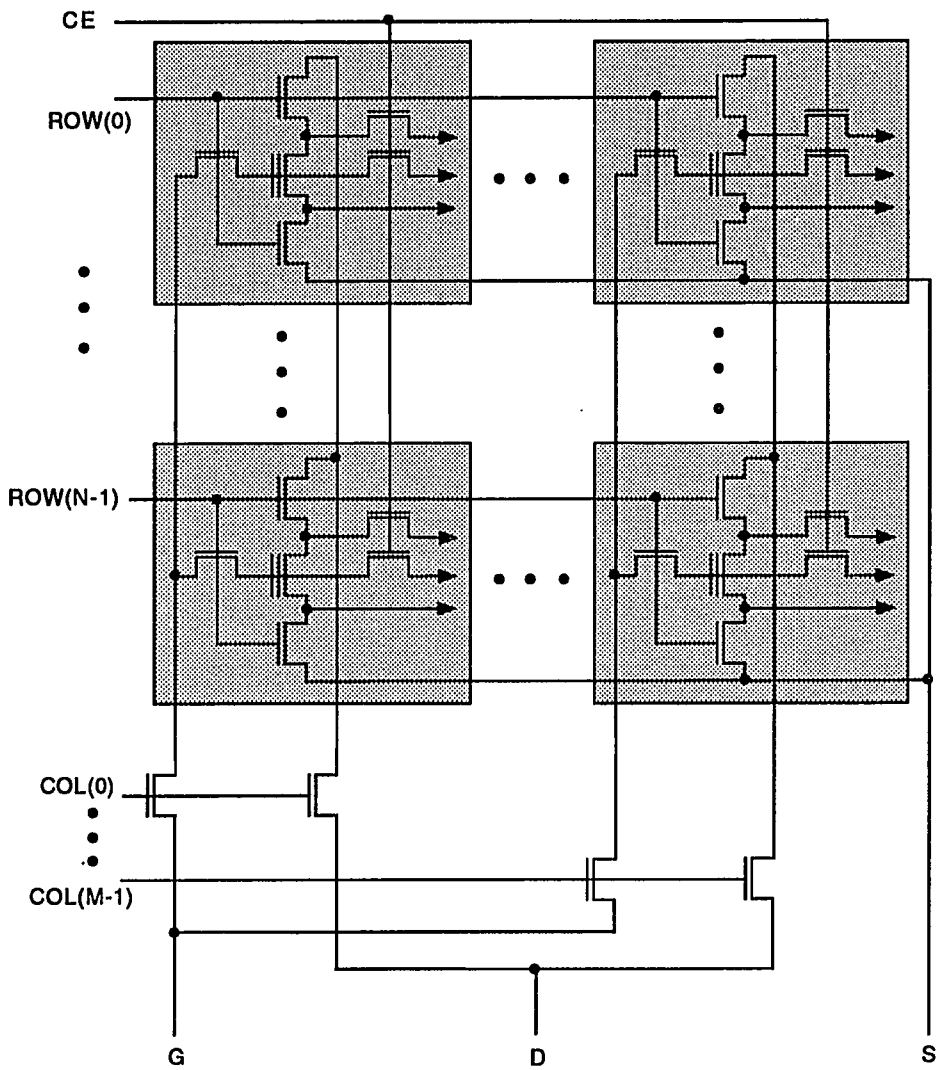


Figure 5.1: Floating gate MOS transistor array

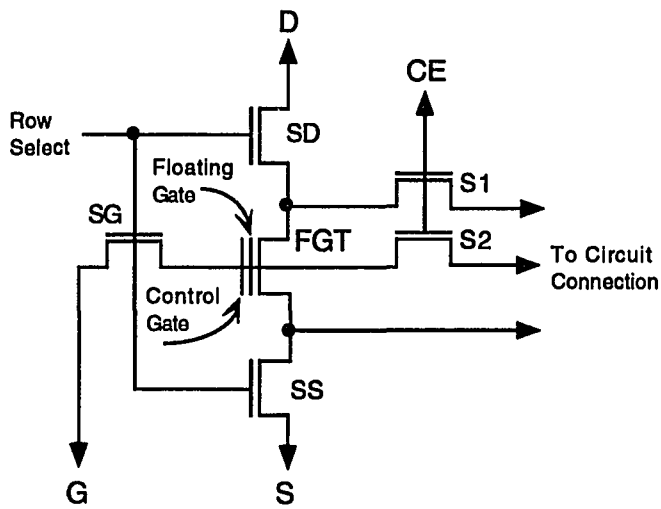


Figure 5.2: Floating gate MOS transistor cell terminology

mode and a normal mode. In a V_{th} tuning mode the cells are disconnected from the main circuit and sequentially selected through the row and column selection lines so that the threshold voltages of the FGTs can be tuned. In a normal mode the cells are connected to each other according to the circuit topology by turning on the switch transistors, S1, and S2 in Fig. 5.2. The signal CE is used to connect or disconnect the cells from the circuit.

5.3 V_{th} Tuning Strategy

One V_{th} tuning strategy is presented here. The simplified entire block diagram of the V_{th} tunable low voltage op-amp circuit is shown in Fig. 5.3. A counter, a row decoder, and a column decoder can be used to sequentially select the cells of the FGT array. The V_{th} tuning is performed in two steps: a coarse tuning and a fine tuning. The coarse tuning is a preliminary step to provide an environment where the on-chip charge pump and the main circuit are capable of operating with a low voltage. The fine tuning is for providing a good matching properties and a desired operating point.

In the coarse tuning all the FGTs that are the elements of the FGT array are approximately programmed in a one-tuning cycle to have a very low threshold voltage (e.g. 100mV) using an external high voltage (e.g. 20V). The coarse tuning can be performed using either a closed-loop mode or an open-loop mode. This action is performed only once just after the circuit is fabricated. After the coarse tuning the entire circuit operates with a very small power supply

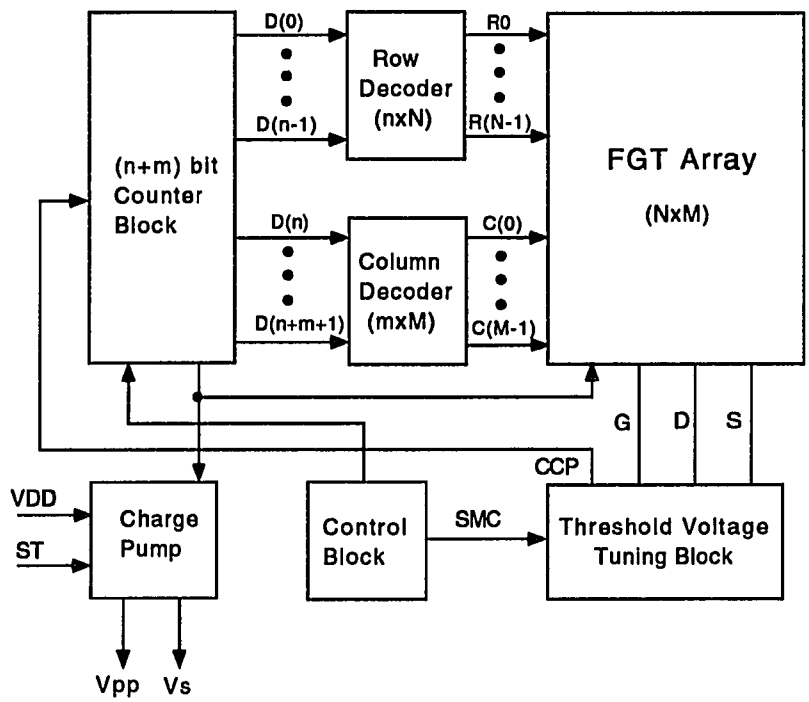


Figure 5.3: Simplified block diagram of the V_{th} tunable low voltage circuit

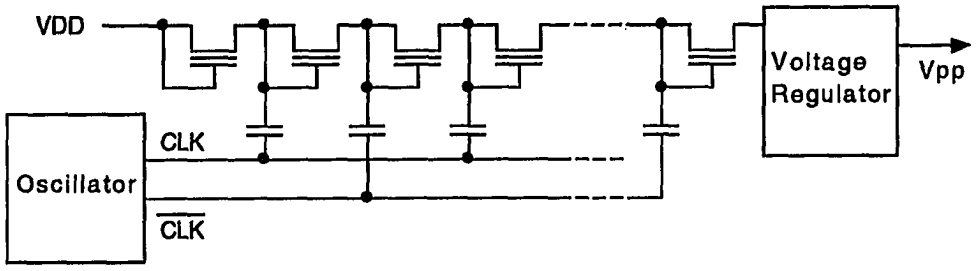


Figure 5.4: The charge pump circuit constituted of floating gate transistors

(e.g. VDD=0.5V). The fine tuning is performed under the external control signal ST. Whenever the signal ST is detected, the fine tuning is performed, and the circuit automatically returns to a normal mode when the fine tuning is finished. The fine tuning need not be a frequent event because of the long-term charge retention characteristics of the FGTs. This can afford the possibility of near continuous-time operation.

In a fine tuning mode a high voltage is also required to adjust the threshold voltages of the FGTs. The high voltage V_{pp} is developed from VDD with an on-chip charge pump, and thus, no external high voltages are required. The charge pump consists of an oscillator, diode-connected floating gate MOS transistors, capacitors, and a voltage regulator as shown in Fig. 5.4. The detailed operation principles of the charge pump can be seen in [100]. To make the charge pump operate with a low power supply, the oscillator and the voltage regulator circuits are also constituted of FGTs. The threshold voltages of the FGTs in the charge pump circuit are also adjusted to a very low value (e.g. 100mV) during the coarse tuning step. After that the charge pump can generate a high voltage V_{pp} from VDD, and the internally generated V_{pp} is used for the fine tuning.

When the tuning of all cells in the FGT array is completed, the circuit automatically returns to a normal mode, and the oscillator of the charge pump circuit is also disabled. Thus, the charge pump does not generate the high voltage V_{pp} any more in the normal mode and is left in a state where it awaits another ST signal. In the normal mode the switch transistors must be turned on for the circuit to function correctly. Since the switch transistors are conventional MOS transistors, they will not be turned on by the very low supply voltage. Hence, another charge pump circuit is required which can generate a voltage that is high enough to turn the

switch transistors on in the normal mode. The power dissipation by the generated voltage will be very small since it is applied at only gate terminals of the MOS transistors and possibly one or two small drain or source diffusions. The frequency of the charge pump oscillator for the switch transistors does not have to be high because the load resistance is very high.

5.4 Fine Tuning Method

There can be many fine tuning methods to obtain a good matching property and a desired operation of the coarsely tuned op-amp circuit. Two fine tuning methods are presented here. The first method is to adjust the threshold voltages of all FGTs to a predetermined value. The second method is to adjust the intermediate node voltages of the circuit to pre-assigned values by adjusting the threshold voltages. The first method will be simpler and more generally applicable to all kinds of circuit structures than the second method. However, the second method will give better matching results because it can provide the node voltage matching while the first method can provide only the threshold voltage matching of the presumably matched transistors.

5.4.1 Threshold Voltage Fine Tuning Method

The block diagram of the fine tuning block for threshold voltage tuning is shown in Fig. 5.5. During a V_{th} tuning mode this block alternately measures the threshold voltage of a selected FGT and adjusts it to a desired threshold voltage. It is controlled by the control signal SMC generated from the control block. The VT control block shown in Fig. 5.6 compares the measured V_{th} with the desired V_{th} and then, determines the adjustment direction and also determines through a zero crossing detector whether to keep adjusting or to finish it. If further adjusting is required, the unit pulse generator generates a unit programming pulse for V_{th} adjustment. The resolution of the V_{th} tuning depends on the magnitude and width of the unit programming pulse.

The accuracy of the tuning results also depends on the performance of the VT measurement block. A simple scheme can be used to measure the threshold voltage. For example, $V_{th} \approx V_{GS}$ may be assumed at low I_{DS} [101]. This method is very simple, but the measurement error is somewhat large. Using this kind of simple method, good absolute accuracy can not be obtained, but a good threshold voltage matching can be obtained for equally sized devices. To obtain more accurate tuning, a more complicated measurement circuit is required at the expense of much larger area.

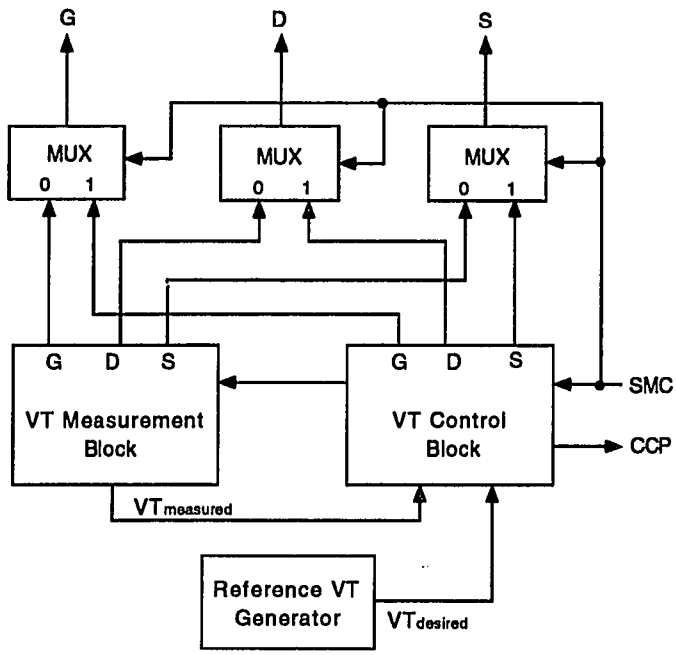


Figure 5.5: The block diagram of the fine tuning block

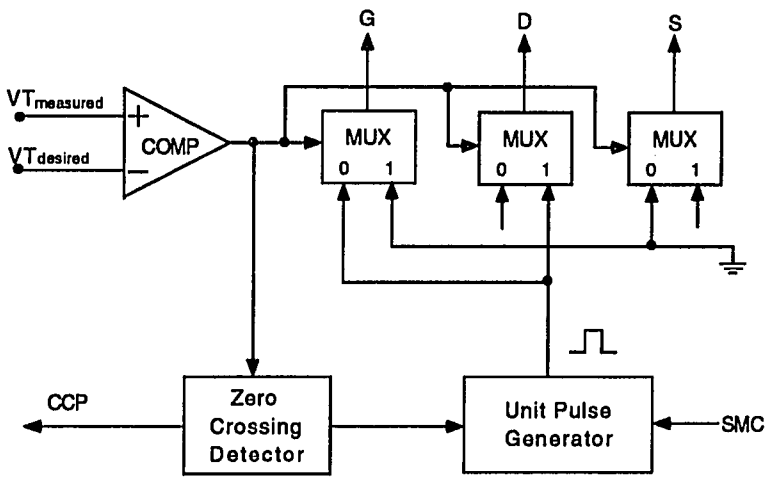


Figure 5.6: The VT control block

5.4.2 Node Voltage Fine Tuning Method

To obtain a better transistor pair matching of the op-amp, the internal node voltages can be adjusted to pre-assigned values by adjusting the threshold voltages. For node voltage fine tuning, a sequential tuning method for a circuit structure is required because the adjustment of one node voltage may affect the other node voltages. The order of the FGTs to be tuned must be carefully determined, and the FGTs must be placed in the FGT array according to the order. To use the node voltage fine tuning method, the circuit should be changed a little. Switch transistors for the nodes to be tuned are additionally required in the FGT array for comparison with pre-assigned values. No measurement circuit is, however, required because the node voltages are directly compared with the pre-assigned values, as defined by a reference voltage generator which can generate the same number of reference voltages as the number of nodes to be tuned. The reference voltages can be selected by the same counter and decoding circuits that are used for the FGT array.

To demonstrate the node voltage fine tuning method, an example circuit of a NMOS, instead of an n-type FGT, differential input stage which is shown in Fig. 5.7 was simulated because there does not exist a good simulator for FGTs. In this simulation it was assumed that an n-type floating gate transistor and the NMOS transistor have very similar characteristics except that the former can be adjusted to have various threshold voltages. This assumption has been validated in the literature [102]. The node voltage changes were simulated by changing the threshold voltages of the NMOS transistors. The simulation results have shown that the change of V_{T5} changes all the node voltages, V1, V2, and V3, and the change of V_{T1} also changes all the node voltages significantly. The change of V_{T3}, however, causes very small changes in V2 and V3. A change of 10mV in V_{T3} results in changes of less than 0.3mV in V2 and V3. The node voltages can thus be approximately adjusted to pre-assigned values by selecting the order of transistors to be adjusted. A possible node voltage fine tuning scheme for the differential input stage is as follows:

Step 1 The threshold voltage of M5 is tuned to adjust the node voltage V3 to its desired value.

Step 2 The transistor M3 is tuned to adjust V1 to its desired value. This step little affects the node voltage V3.

Step 3 The transistor M4 is tuned to adjust V2 to its desired value. V3 will also be little affected by this step, but V1 will be a little bit changed.

Step 4 The threshold voltage of M1 is adjusted to obtain V1$\approx$V2 by comparing V1 with V2

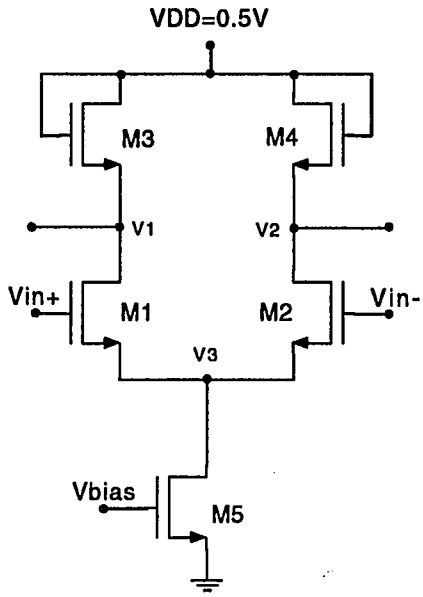


Figure 5.7: NMOS differential input stage

and adjusting M1 in the direction of reducing the difference between V1 and V2. This step will provide a good matching although the exactly pre-assigned node voltages can not be achieved. However, their difference will be very small and the convergence speed will be very fast.

The differential input stage has been tuned according to the above procedure. The tuning results are shown in Table 5.1. The unit step of the threshold voltage which can be adjusted was assumed to be 0.2mV. The pre-assigned node voltages were assumed to be 0.23V, 0.23V, and 0.12V for V1, V2, and V3, respectively. The initial node voltages are shown in Table 5.1. Step 1 took 44 unit steps to result in 0.11996V for V3. Step 2 took 14 unit steps to obtain 0.23001V for V1. This step changed the node voltage V3 by 0.11mV. Step 3 took 13 unit steps to obtain 0.23001V for V2. This step also changed V3 by 0.11mV and V1 by 0.18mV. After Step 3 all the node voltages are very close to the pre-assigned values. To obtain a good matching between V1 and V2, Step 4 was performed. Step 4 took only 1 unit step, and the consequence is that V1 differs from V2 by 0.13mV.

Further reduction of the difference can be obtained by using a narrower unit step, but the improvement will be restricted by the performance of a comparator that will be used for comparing between the node voltages and the pre-assigned values. It should also be noted that

Table 5.1: Tuning results of the node voltages of a NMOS differential input stage

	V1(V)	V2(V)	V3(V)	# of unit steps
Pre-assigned	0.23	0.23	0.12	
Initial	0.24175	0.24175	0.12842	
Step 1	0.22748	0.22748	0.11996	44
Step 2	0.23001	0.22766	0.12007	14
Step 3	0.23019	0.23001	0.12018	13
Step 4	0.23003	0.23016	0.12027	1

adjusting the node voltages to pre-assigned values may push one of the transistors into a linear region, so careful design and selection of the desired node voltages are initially required. The node voltage fine tuning method can provide better matching property and does not require a measurement circuit. The tuning scheme should, however, be changed for other types of op-amp structures.

5.5 Conclusions

In this chapter a method to obtain very low voltage op-amps has been presented. The op-amp are constituted of floating gate MOS transistors. Adjusting the threshold voltages of the FGTs makes the op-amp have a capability of operating with very low power supply voltages. In the simulation here, operation at a supply voltage of 0.5V was obtained. Operation at substantially lower supply voltage levels can also be readily achieved.

A two-step tuning scheme has been presented. The coarse tuning is used to adjust the threshold voltages of all FGTs to a small value (e.g. 100mV) in a one-shot cycle so that the charge pump can operate with the low voltage and thus, the FGTs of the op-amp can be tuned with an internally generated high voltage from the charge pump. The coarse tuning is performed only once. The fine tuning which is performed under an external control signal is used to provide a good matching property and a desired operation of the op-amp. Two fine tuning methods have been presented and compared, which are a threshold voltage fine tuning and a node voltage fine tuning. The basic low-voltage methodology can be extended to achieve low voltage operation in other analog as well as digital applications.

CHAPTER 6. AN ACCURATE AND MATCHING-FREE V_T EXTRACTOR USING A RATIO-INDEPENDENT SC SUBTRACTING AMPLIFIER AND A DYNAMIC CURRENT MIRROR

6.1 Introduction

Numerous numerical techniques exist for accurately extracting device model parameters from measured data [82],[11]. One example of such a technique is the MOS transistor threshold voltage (V_T) extraction using a linear regression on measurements of I_{DS} at many V_{GS} values. Such techniques are not well suited for real-time on-chip threshold voltage extraction.

Recently, several real-time V_T extraction methods based on circuit implementations have been proposed for overcoming the above disadvantages [101],[103]-[106]. These methods are very fast although the accuracy is degraded compared to that attainable by the numerical methods. Most methods [103]-[105] require matched devices to extract V_T for one test device of a fixed geometry. The resultant accuracy thus depends on the matching between two or more devices. These methods are inefficient when extraction of V_T is required for many transistors with various geometries and particularly unsuitable for V_T extraction of small devices since the matching of small-size transistors is poor. These methods also require other component matching in their extraction circuits such as current mirror transistors and resistors. Mismatches of these components further degrade the accuracy of the extracted V_T values. Moreover, the methods [103]-[105] are not applicable for transistors with different bias conditions, i.e., nonzero substrate-to-source voltages ($V_{BS} \neq 0$) since they need a cascode configuration of matched test-transistors with the same V_{BS} of all the test-transistors which is not possible due to their cascode configuration. In contrast to the methods discussed in [103]-[105], the method discussed in [101] uses only one test device and thus does not require device matching. Although the latter method is very simple, it produces relatively large errors (about 100mV) due to the uncertainty of choosing the proper threshold current which is used to measure V_T.

In this chapter a matching-free V_T extraction scheme is presented which does not require

Table 6.1: Feature comparison of V_T extraction schemes

Schemes	Required number of matched test-transistors	Required components that should be matched	Applicability at different geometries	Applicability at different substrate bias conditions	Comments
Numerical [1]	None	None	Efficient	Yes	Accurate but Not suitable for real time
Wang [4] & Johnson [3]	9	Current mirror trs	Inefficient	No	Using a transistor array
Tsividis [5]	3	Current mirror trs	Inefficient	No	Using a transistor string
Alini [6]	2	Resistor & Current mirror trs	Inefficient	Yes	BICMOS implementation
Lee [7]	None	None	Efficient	Yes	Simple but Poor accuracy
Proposed	None	None	Efficient	Yes	Dynamic implementation

any replica of the device under test and which is applicable to transistors with different geometries and different substrate bias conditions. The proposed extraction circuit is implemented in a matching-free way by using a ratio-independent switched-capacitor amplifier and a dynamic current mirror. Thus, the accuracy of the proposed scheme does not depend on the test-transistor matching and other component matching in the extraction circuit. The features of the proposed scheme is comparatively summarized in Table 6.1 with other extraction schemes mentioned above.

6.2 Principle of the Matching-Free V_T Extractor

6.2.1 Basic Scheme

A conceptual schematic of the proposed V_T extraction scheme is depicted in Fig. 6.1. Applying the outputs of a current mirror, I_{D1} (with S1 closed and S2 open) and I_{D2} (with S1 open and S2 closed), to a test transistor which operates in the saturation region and assuming that the transistor has square-law characteristics, we obtain respectively

$$K(V_{GS1} - V_T)^2 = I_{D1} \tag{6.1}$$

$$K(V_{GS2} - V_T)^2 = I_{D2}, \tag{6.2}$$

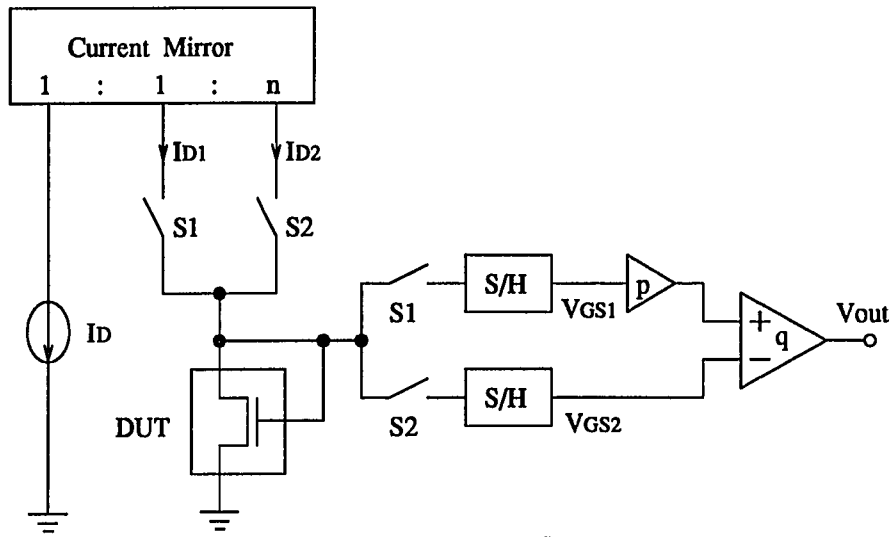


Figure 6.1: Conceptual schematic of the proposed V_T extraction scheme

where

$$I_{D2} = nI_{D1} \tag{6.3}$$

$$K = \frac{\mu C_{OX}}{2}\frac{W}{L}, \tag{6.4}$$

and where V_{GS1} and V_{GS2} are the corresponding gate-source voltages of the device under test (DUT). Equation (6.1) and (6.2) have the same K and V_T because only one test transistor is used as contrasted with other extraction schemes. Solving these equations, we readily obtain the threshold voltage V_T

$$V_T = \frac{1}{\sqrt{n}-1}(\sqrt{n}V_{GS1} - V_{GS2}) \tag{6.5}$$

Assume S1 and S2 are driven by a complimentary nonoverlapping clock. When S1 is closed, V_{GS1} is sampled and multiplied by p. When S2 is closed, V_{GS2} is sampled and subtracted from pV_{GS1}. The result is then multiplied by q. The output voltage V_{out} of Fig. 6.1 is then

$$V_{out} = q(pV_{GS1} - V_{GS2}). \tag{6.6}$$

If we select

$$p = \sqrt{n}, \tag{6.7}$$

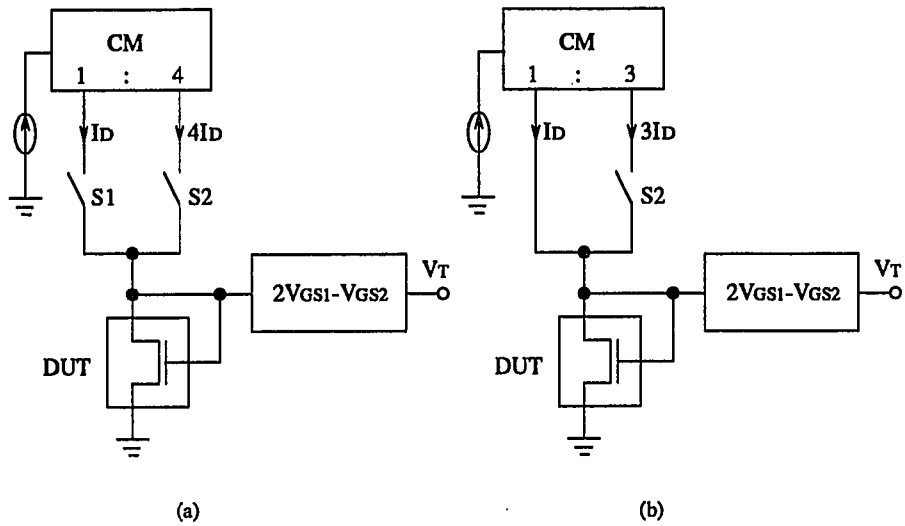


Figure 6.2: Two switching schemes to obtain currents I_D and $4I_D$

then it follows from (6.5) and (6.6) that

$$V_{out} = mV_T, \tag{6.8}$$

where

$$m = q(\sqrt{n} - 1). \tag{6.9}$$

Thus, an integer multiple of V_T can be readily obtained by choosing an integer m.

The easiest way to obtain V_T is to choose $n = 4$ and $m = 1$ resulting in $p = 2$ and $q = 1$ and thus,

$$V_{out} = 2V_{GS1} - V_{GS2} = V_T, \tag{6.10}$$

where V_{GS1} and V_{GS2} are the gate-to-source voltages of the test-transistor when the drain currents are I_D and $4I_D$, respectively. Two variants of a switching schemes that can be used to obtain currents I_D and $4I_D$ are depicted in Fig. 6.2.

The simple analog arithmetic operation, $(2V_{GS1} - V_{GS2})$, can be accurately implemented using a switched-capacitor subtracting amplifier. These kinds of switched-capacitor circuits for basic arithmetic operations are capable of providing high accuracy as indicated in [73]. Most common implementations of the current mirror and the switched-capacitor amplifier require

device and capacitor matching respectively although matching with the device under test is not required. Since both the current mirror gain and the amplifier gain have been chosen to be both small and integral, both blocks can be dynamically implemented without requiring any matching of devices or capacitors. The implementation of the blocks will be discussed in the following sections.

6.2.2 Model Error Consideration

As with most other extraction methods [103]-[106], the proposed V_T extraction scheme is also based on the assumption that MOS transistors operating in the saturation region obey the square-law. The characteristics of real MOS transistors, however, deviate from the square-law due to the nonideal effects such as channel-length modulation and mobility degradation, resulting in a discrepancy between the extracted V_T and a real V_T.

Including these nonideal effects, the drain current can be more accurately modeled by

$$I_{DS} = \left[\frac{\mu_o}{1+\theta(V_{GS}-V_T)}\right]\left[\frac{1}{L(1-\lambda V_{DS})}\right]\frac{C_{OX}W}{2}(V_{GS}-V_T)^2, \tag{6.11}$$

where λ is the channel-length modulation parameter, θ is the mobility degradation parameter, and μ_o is the zero-field mobility of carriers. The parameter λ is inversely proportional to the channel length and usually in the range of $0.004V^{-1}$ (L>50μm) to $0.3V^{-1}$ (for very short channel) [73]. Mobility degradation is caused by the increase of carrier scattering from the S_i-S_iO_2 interface as the normal channel electric field increases. The parameter θ is inversely proportional to the channel length and usually in the range of $0.001V^{-1}$ to $0.25V^{-1}$ [73].

Since the emphasis is on extracting V_T, the threshold voltage must also be discussed. It should be noted that there are three differently defined threshold voltages as follows [107]:

$\mathbf{V_{TO}}$ Zero-bias threshold voltage of a large device

$\mathbf{V_{TH}}$ Including device size effects and terminal voltage effects

$\mathbf{V_{ON}}$ Including subthreshold current effects

V_{TO} is the threshold voltage of a zero-biased very large device which is usually used as an input model parameter for SPICE simulations. V_{TO} can be extracted with high accuracy by the extraction schemes based on the square-law since large-size MOS transistors relatively well obey the law.

V_{TH} is the effective threshold voltage where device size effects (short channel and narrow width) and terminal voltage effects (V_{BS} and V_{DS}) are taken into account. V_{TO} is thus a special

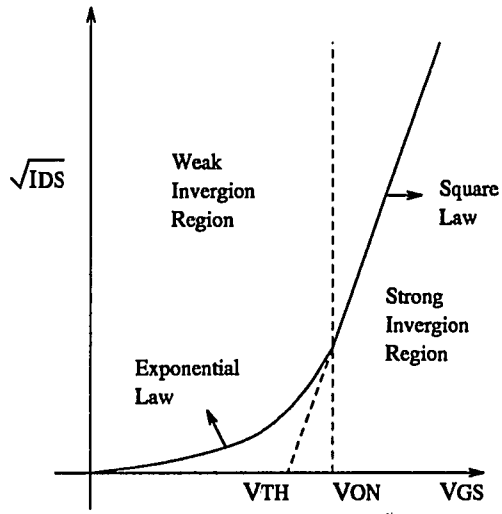


Figure 6.3: Illustration of the transition voltage V_{ON} and its difference from V_{TH}

case of V_{TH}. The effect of V_{BS} on the threshold voltage is the well-known body effect, where V_{TH} increases with $|V_{BS}|$. The threshold voltage V_{TH} is reduced as the drain-to-source voltage V_{DS} increases, which is known as the Drain-Induced Barrier Lowering (DIBL) effect [108]. The DIBL effect is not significant unless the channel length is too short. As the channel length is reduced, the discrepancy between the extracted V_{TH} and a real V_{TH} increases because the λ and θ effects become significant, thus increasing the deviation from the square-law.

V_{ON} is defined as the transition voltage between the weak inversion region and the strong inversion region [107]. The weak inversion region of operation is characterized by the fast surface states, NFS (SPICE model parameter). When the gate-to-source voltage V_{GS} reaches the transition point V_{ON}, the characteristic of the drain current I_{DS} changes from the square-law to an exponential law as shown in Fig. 6.3. The figure also shows clearly the difference between V_{TH} and V_{ON}. The threshold voltage V_{ON} can not be extracted using the extraction schemes based on the square-law. In SPICE V_{ON} is obtained by adding V_{TH} to another term which can be calculated using the parameter NFS extracted from measurements.

To investigate the influence of the model error due to the λ and θ effects on the performance of the proposed V_T scheme, equation (6.11) instead of the square-law equations (6.1) and (6.2) is used to derive V_T $n = 4$ and $m = 1$. Neglecting the second order terms of λ and/or θ, the

resulting expression is

$$\begin{aligned} 2V_{GS1} - V_{GS2} &= V_T + \frac{1}{2}(\lambda - \theta)V_{ex1}V_{ex2} \\ &= V_T + \frac{1}{4}(\lambda - \theta)V_{ex2}^2, \end{aligned} \tag{6.12}$$

where V_{ex1} is the excess voltage $V_{GS1} - V_T$ when the drain current is I_D and V_{ex2} is the excess voltage $V_{GS2} - V_T$ when the drain current is $4I_D$. The model error voltage due to the λ and θ effects is thus

$$\begin{aligned} V_{Terr} &= \frac{1}{2}(\lambda - \theta)V_{ex1}V_{ex2} \\ &= \frac{1}{4}(\lambda - \theta)V_{ex2}^2. \end{aligned} \tag{6.13}$$

Small excess voltages will help reducing the error voltage, which is an expected result because the channel-length modulation and mobility degradation effects increase with V_{DS} and V_{GS}, respectively, and the test device in our extractor is diode-connected to guarantee its saturation-region operation, resulting in $V_{DS} = V_{GS}$. It is interesting to note that the two parameters in (6.13) are in a relation of canceling each other, and fortunately, both parameters are inversely proportional to the channel length. Therefore, the variance of $\lambda - \theta$ and thus the error voltage will not increase substantially with the channel length reduction. For example, if the maximum difference value of the two parameters is $0.1V^{-1}$, if $V_T = 0.8V$, and if $V_{ex2} = 0.4V$, then the model error voltage will be less than 0.5%.

The model error in the proposed extraction scheme has been simulated for two test devices which have different geometries using SPICE Level 2 MOS models (VTO=0.924V). To determine model error effects alone, no error associated with the current mirror or the analog arithmetic block was assumed. The extracted $V_{T_{ext}}$ which has been calculated from simulated V_{GS1} (at I_D) and V_{GS2} (at $4I_D$) is compared with the threshold voltage V_{TH} (NFS=0) computed by SPICE and listed in the SPICE output file. With the assumption that the V_{TH} computed by SPICE is the actual threshold voltage, the error voltage $V_{T_{err}}$ ($V_{T_{ext}} - V_{TH}$) is plotted in Fig. 6.4(a) as a function of bias current I_D.

As expected, the error voltages for the long-channel device (W/L=200μm/40μm) are smaller at all I_D values than those for the short-channel device (W/L=20μm/4μm). It can be seen that the error increases with I_D since large I_D increases the excess voltage as shown in Fig. 6.4(b), where the error voltages are plotted as a function of the excess voltage V_{ex2} ($V_{GS2} - V_{TH}$). The figure exhibits that the error variation of the proposed scheme due to the excess voltage V_{ex2} is comparable with the variation due to the device size. It can be also seen

that the slope of the curves in Fig. 6.4 changes substantially at a small I_D or a small V_{ex2} that corresponds to the transition point V_{ON}. Therefore, the bias current I_D should be selected carefully such that the excess voltages V_{ex1} and V_{ex2} are greater than the transition point but not too big to maintain a small model error. It can be seen in Fig. 6.4(b) that if $V_{ex2} \leq 0.4$V than the error voltage due to the model error will be less than 5mV even with the short-channel device (L=4μm).

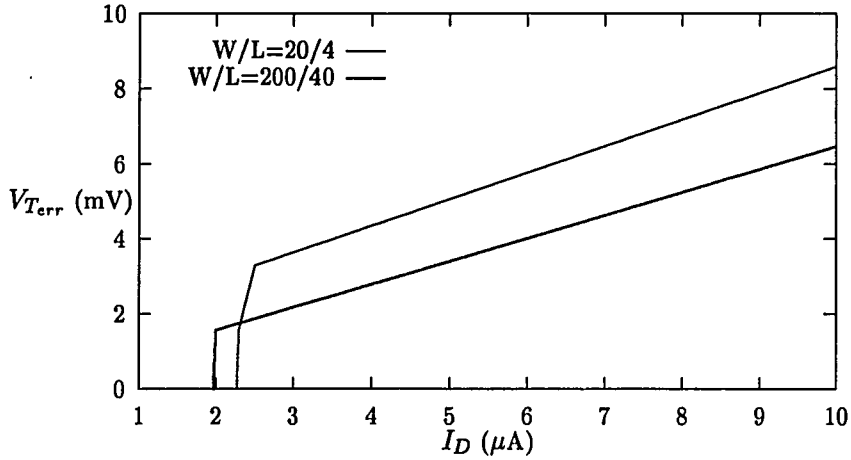
The proposed scheme has also been compared with the linear regression (LR) method [82] in Table 6.2. In the LR method, I_{DS} values are collected at 20 V_{GS} values using SPICE, so no measurement error is assumed. For consistency in excess voltages, the V_{GS} values are selected such that the highest sample value V_{GSh} is V_{GS2}, and the lowest sample value V_{GSl} is V_{GS1} as shown in Table 6.2. Since threshold voltages are functions of device terminal voltages, and their variation increases as the device size decreases, the actual threshold voltage V_{TH} of the short-channel device (L=4μm) computed by SPICE varies with V_{GS} that is equal to V_{DS} as shown in the table. At $V_{GS} = 1.066$V, $V_{TH} = 0.889$V, while at $V_{GS} = 1.525$V, $V_{TH} = 0.887$V. Thus, the V_{TH} variation is about 2mV when the V_{GS} change is 0.46V. This variation will be significant for shorter-channel devices. The V_{TH} variation of the long-channel device (L=40μm) is almost negligible. In the proposed scheme the variation in the threshold voltage is due to the two different V_{GS} values, V_{GS1} and V_{GS2}, used to extract V_{TH}, and in the LR method the variation is due to the different V_{GS} values used to grab the I_D data. The average values

$$V_{THav} = \begin{cases} (V_{TH}(V_{GS1}) + V_{TH}(V_{GS2}))/2 & \text{for the proposed scheme} \\ (V_{TH}(V_{GSl}) + V_{TH}(V_{GSh}))/2 & \text{for the LR method} \end{cases}$$

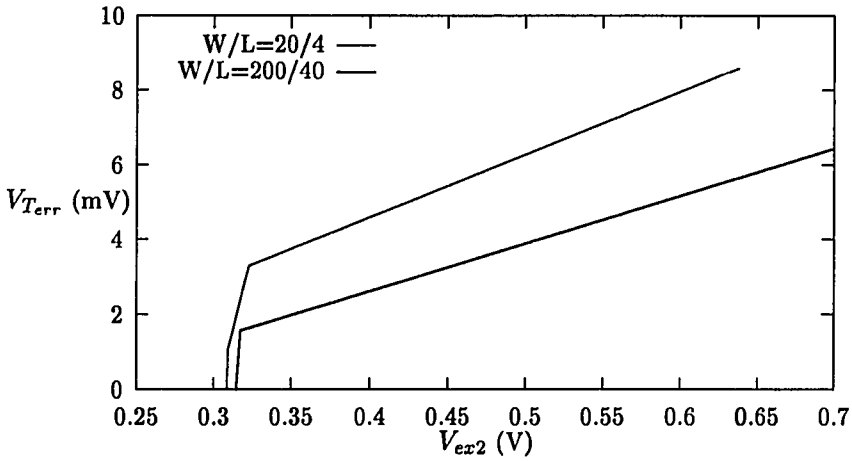
were used to calculate the error of extracted threshold voltages. It can be seen from the table that the accuracy of the LR method is similar to that of the proposed scheme, and both methods give large error when the samples are taken from large V_{GS} values.

6.3 Ratio-Independent SC Subtracting Amplifier

Because of their potential for high-precision monolithic fabrication, switched capacitor (SC) circuits have been widely used for many applications such as filters, data converters, and basic building blocks for analog signal processing. SC summing and/or subtracting amplifiers are one of the most common high-precision analog arithmetic building blocks. However, SC circuits still have some error sources associated with the components composing SC circuits such as MOS switches, capacitors, and amplifiers. Nonideal factors limiting the performance of SC circuits are as follows:



(a)



(b)

Figure 6.4: Error voltage of the extracted voltage V_{Text} from the actual threshold voltage V_{TH} computed by SPICE (a) as a function of the bias current I_D (b) as a function of the excess voltage V_{ex2}

Table 6.2: Accuracy comparison between the proposed scheme and the linear regression method

	Test device size (W/L)	Number of samples	V_{GS1} / V_{TH}	V_{GSh} / V_{TH}	Extracted V_{Text}	Error V_{Text}-V_{THav}
Linear Regression Method [1]	$\frac{20um}{4um}$	20	1.066V/0.888V	1.241V/0.888V	0.8919V	3.43mV
			1.211V/0.888V	1.525V/0.887V	0.8971V	9.58mV
	$\frac{200um}{40um}$	20	1.116V/0.921V	1.309V/0.921V	0.9230V	2.03mV
			1.275V/0.921V	1.623V/0.921V	0.9281V	7.09mV
	Test device size (W/L)	Bias Curr. I_D	V_{GS1} / V_{TH}	V_{GS2} / V_{TH}	Extracted V_{Text}	Error V_{Text}-V_{THav}
The Proposed Scheme	$\frac{20um}{4um}$	3uA	1.066V/0.888V	1.241V/0.888V	0.8921V	3.61mV
			1.211V/0.888V	1.525V/0.887V	0.8961V	8.58mV
	$\frac{200um}{40um}$	10uA	1.116V/0.921V	1.309V/0.921V	0.9233V	2.26mV
			1.275V/0.921V	1.623V/0.921V	0.9275V	6.47mV

1. Parasitic capacitances
2. Nonzero offset voltage of op-amps
3. Finite dc gain of op-amps
4. Capacitor mismatches
5. Charge injection of MOS switches

Many circuit techniques and strategies have been proposed to overcome these nonideal effects [109]-[121]. Most modern SC circuits use parasitic-insensitive structures [72, 109, 110] with which the influence of parasitic capacitances can be significantly reduced. The error due to op-amp offset voltages can also be readily reduced using various switching procedures, which is an attractive feature inherent to SC circuits [82, 111, 112]. A commonly used offset compensation scheme is to store the op-amp offset voltage on capacitors in one clock phase and to subtract it in the subsequent signal processing clock phase. The technique is known as auto-zeroing or correlated double sampling [82],[73].

In all SC circuits the performance depends on the accuracy of capacitor ratios not the individual capacitor accuracy. Although the ratios can be realized with high accuracy in modern technology, they still produce some error. The ratio error due to capacitor mismatches can, however, be eliminated using the ratio-independent concept introduced by Lee [113] where multiplication by integer N of the input voltage can be obtained independently of the capacitor ratios at the expense of more clock phases with the required number of clock phases increasing with N. Hence, the ratio-independent concept has been used primarily for high-accuracy but low to medium speed applications such as high-resolution and low-cost ADCs [113]-[116].

Recently, Nagaraj [117],[118], Lason [119] and Haug [120] have proposed techniques to reduce the error associated with the op-amp finite gain. In these techniques the finite gain error is compensated during a main-operation clock phase using the finite gain error information obtained from a preliminary operation during the previous clock phase. It has been shown that with this scheme the effective gain of the op-amp is squared, and the phase error is also reduced.

The simple arithmetic $(2V_{GS1} - V_{GS2})$ needed in our V_T extractor is realized using a SC circuit. Interestingly, the arithmetic, multiplication by two and subtraction, is also a key operation in algorithmic or cyclic ADCs [113]-[116]. To implement the SC circuit in a matching-free way and to relax the op-amp gain requirement, Lee's ratio-independent concept and Nagaraj's gain-insensitive technique are employed in our SC subtracting amplifier. Now, all the nonidealities except for the charge injection of MOS switches can be compensated.

With this approach the charge injection effect becomes the dominant factor limiting the performance of the SC circuit. The charge injection mechanisms are quite complicated. To date, the charge injection compensation schemes [131]-[136] can perform only partial compensation. The charge injection problem deserves a more detailed discussion and will be addressed in a separate section. In this section the operation principle of the proposed SC subtracting amplifier is discussed, and the residual nonideal effects are studied through both theoretical analyses and SWITCAP [122] simulations.

6.3.1 Principle of Operation

The schematic of the proposed ratio-independent SC subtracting amplifier is shown in Fig. 6.5. The circuit performs the analog arithmetic $(2V_{GS1} - V_{GS2})$ and operates in six nonoverlapping clock phases ϕ_1 to ϕ_6. Since a single test device is used, V_{GS1} and V_{GS2} can not be available at the same time. Thus, the input of the SC circuit, V_G is

$$V_G = \begin{cases} V_{GS1} & \text{for } \phi_1 \text{ and } \phi_4 \\ V_{GS2} & \text{for } \phi_2 \text{ and } \phi_3 \end{cases} \tag{6.14}$$

The current mirror in Fig. 6.2 is dynamically implemented such that it can supply the test device with I_D during ϕ_1 and ϕ_4 and with $4I_D$ during ϕ_2 and ϕ_3. The capacitors C1 and C2 are used for main operations, and C3 and C4 are the corresponding auxiliary capacitors for the preliminary operations required for compensation of the finite op-amp gain. C3 and C4 are chosen such that C3/C4=C1/C2. Capacitor Cc is used to store the finite gain error voltage.

The step-by-step operation of the circuit is described in Fig. 6.6 with ideal equations for the capacitor voltages. During phase ϕ_1 the input signal V_{GS1} is sampled onto both the sampling capacitors C1 and C3. During phase ϕ_2 the charge corresponding to $V_{GS1} - V_{GS2}$ is transferred onto C4 from C3. At this time the error voltage (ideally zero) at the inverting input terminal of the op-amp which is caused by the finite gain and the offset voltage of the op-amp is stored in Cc. The error voltage is denoted as $V_1(2)$ where the subscript denotes the node number and the number inside the parenthesis denotes the phase. The error voltage $V_1(2)$ is subtracted from $V_1(3)$ during ϕ_3 such that the virtual ground voltage level $V_2(3)$ becomes as small as $V_1(3) - V_1(2)$, while the main charge transfer is performed from C1 onto C2. The difference voltage $V_1(3) - V_1(2)$ will be very small. If the difference voltage is assumed zero, then the amount of charge stored in C2 will be exactly $C1(V_{GS1} - V_{GS2})$.

During ϕ_4 the input signal V_{GS1} is sampled again onto C1 and C3. During ϕ_5 the charge stored in C4 is transferred back onto C3. The error voltage associated with this operation is

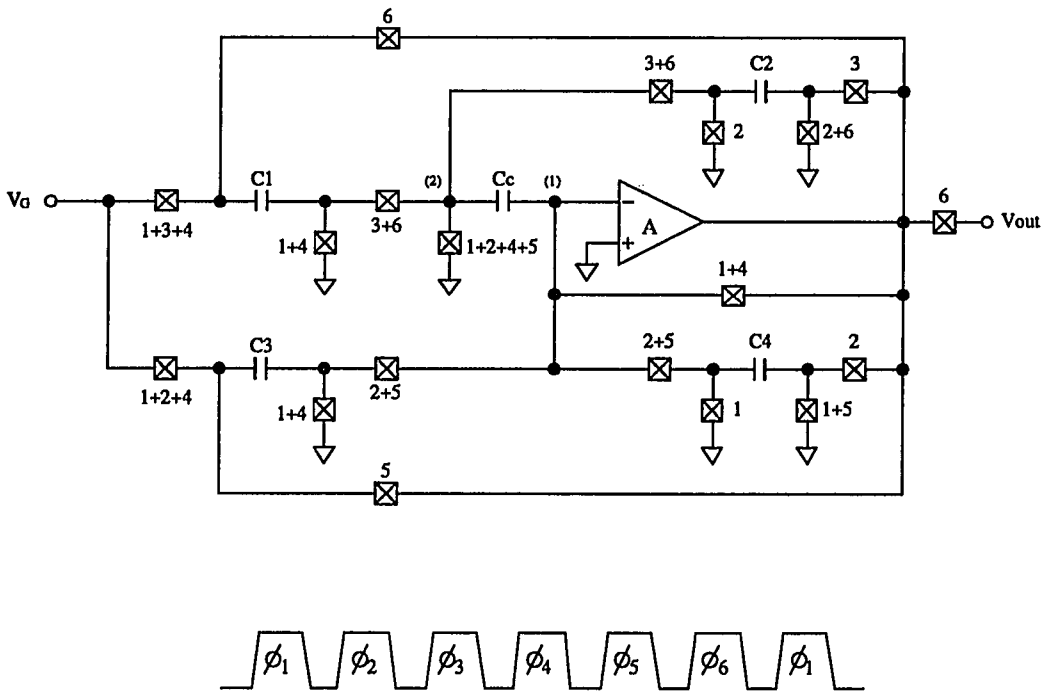


Figure 6.5: Schematic of the proposed ratio-independent SC subtracting amplifier and clock sequence

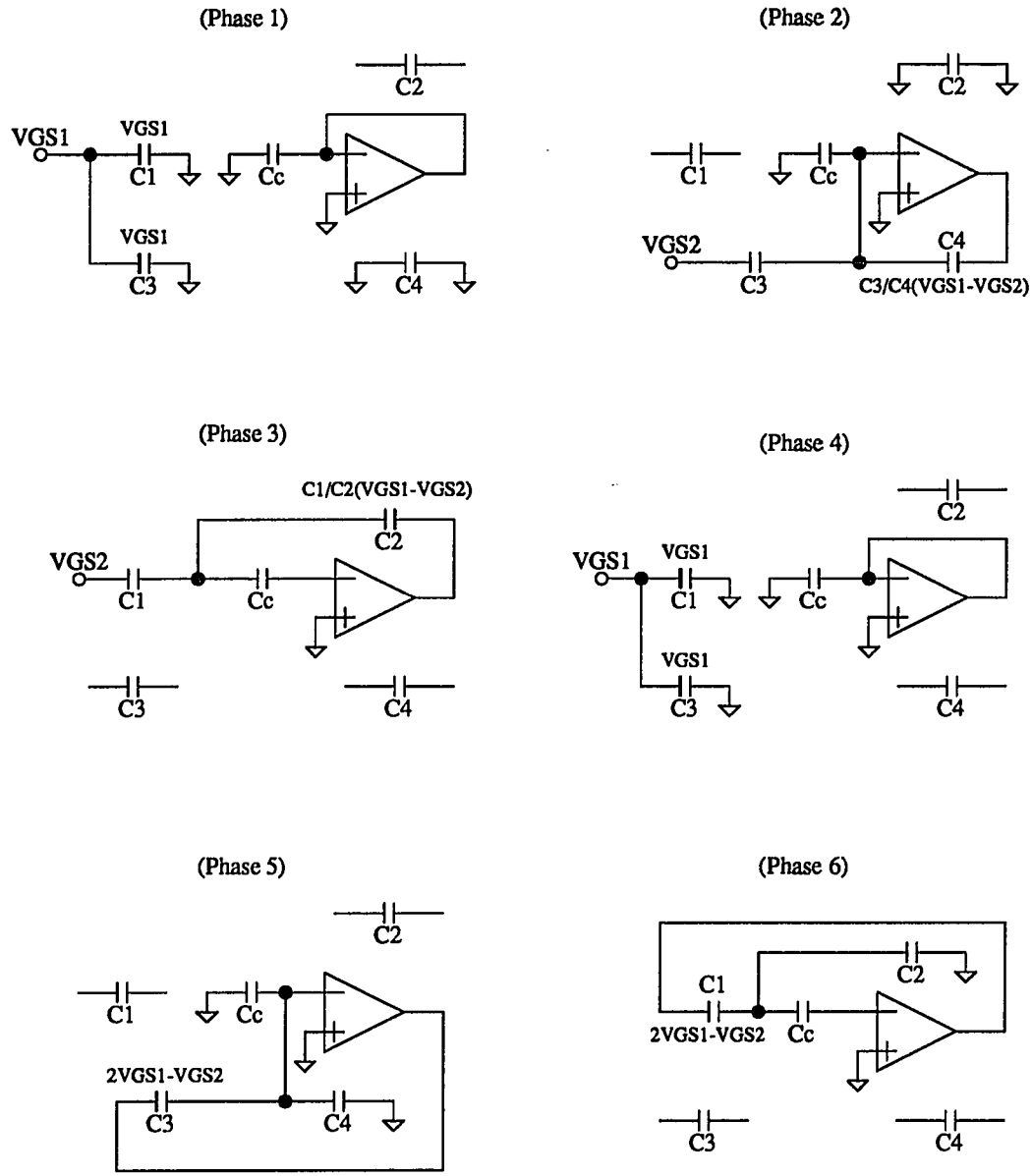


Figure 6.6: Step-by-step operation of the SC subtracting amplifier with ideal equations of the capacitor voltages

also stored in Cc. During the last clock phase ϕ_6, the charge stored in C2 is transferred back onto C1 and added to the charge stored already during ϕ_4. During this phase the virtual ground voltage $V_2(6)$ also becomes as small as $V_1(6) - V_1(5)$. Assume again $V_1(6) = V_1(5)$, then the amount of charge stored in C1 is $C1(2V_{GS1} - V_{GS2})$, hence the output voltage is $2V_{GS1} - V_{GS2}$ independently of the capacitor ratios. In reality the difference voltages $V_1(3) - V_1(2)$ and $V_1(6) - V_1(5)$, however, are not exactly zero because of the nonideal effects. In the following the effects of the nonidealities on the error voltage are investigated analytically.

6.3.2 Sensitivity to Finite Op-amp Gain and Parasitic Capacitances

The error voltage associated with the finite op-amp gain and the parasitic capacitances at internal nodes are analytically derived. To reduce complexity only two parasitic capacitances C_{p1} and C_{p2} at node (1) and (2) which are critical nodes are considered. It is assumed in this analysis that the offset voltage of the op-amp is zero and

$$\frac{C3}{C4} = \frac{C1}{C2} \tag{6.15}$$

There are two main charge transfer operations through the op-amp feedback loop during ϕ_3 and ϕ_6.

During phase ϕ_3 the voltage across C2 is given by

$$V_{out}(3) - V_2(3) = \left[\frac{1+az}{1+ay}\left(\frac{C1}{C2} + \frac{ay}{1+ax}\frac{C3}{C4}\right) - \frac{az}{1+ax}\frac{C3}{C4}\right](V_{GS1} - V_{GS2}) \tag{6.16}$$

$$= \frac{C1}{C2}(V_{GS1} - V_{GS2})\frac{1}{1+\epsilon_1} \tag{6.17}$$

where

$$\epsilon_1 \simeq a^2 x(y-z) \tag{6.18}$$

and

$$a = 1/A \quad (A \text{ is the op-amp gain}) \tag{6.19}$$

$$x = 1 + \frac{C3 + Cc + C_{p1}}{C4} \tag{6.20}$$

$$y = \left(1 + \frac{C_{p1}}{Cc}\right)\left(1 + \frac{C1 + Cc + C_{p2}}{C2}\right) - \frac{Cc}{C2} \tag{6.21}$$

$$z = 1 + \frac{C_{p1}}{Cc} \tag{6.22}$$

It can be seen that the gain error ϵ_1 is inversely proportional to A^2, and the effects of parasitic capacitances C_{p1} and C_{p2} are also divided by A^2.

During phase ϕ_6 the output voltage is given by

$$V_{out}(6) = (2V_{GS1} - V_{GS2}) - V_{err} \tag{6.23}$$

where

$$V_{err} \simeq a^2[y'(x-1)(V_{GS1} - V_{GS2}) + x'y'(2V_{GS1} - V_{GS2})] + \epsilon_1(V_{GS1} - V_{GS2}) \tag{6.24}$$

and

$$x' = 1 + \frac{C4 + Cc + C_{p1}}{C3} \tag{6.25}$$

$$y' = \left(1 + \frac{C_{p1}}{Cc}\right)\left(1 + \frac{C2 + Cc + C_{p2}}{C1}\right) - \frac{Cc}{C1} \tag{6.26}$$

The final error voltage V_{err} is also inversely proportional to A^2 where it has been shown in (6.18) that $\epsilon_1 \propto 1/A^2$.

The derived equations were verified through SWITCAP simulations. One example that shows the errors in extracting V_{TH} due to a limited finite gain and the presence of parasitic capacitances follows. With the following conditions:

$$\begin{aligned} &V_{GS1} = 1.2V \quad V_{GS2} = 1.4V \\ &C1 = C2 = C3 = C4 = Cc \\ &\frac{C_{p1}}{C1} = \frac{C_{p2}}{C1} = 0.1 \\ &A = 100 \quad (a = 0.01), \end{aligned} \tag{6.27}$$

the calculated output voltage is 0.99944V while the simulated one is 0.99946V. Since the ideal output voltage $(V_{GS1} - V_{GS2})$ is 1V, it can be seen that with even a very small op-amp gain of 100, the error voltage is as small as 0.05%.

6.3.3 Sensitivity to Capacitor Ratio Mismatches

In the previous analysis, capacitor matching as given in equation (6.15) is assumed. However, there will exist a small mismatch component between the two capacitor ratios, i.e., C3/C4 used at preliminary operation and C1/C2 used at main operation. This mismatch component is a side effect of the technique employed for compensation of the finite gain error and will slightly increase the error voltage. Effect of the capacitor ratio mismatch is now investigated.

The capacitances C1, C2, C3, and C4 can be defined as in [5] (see Chapter 3).

$$Ck = Ck_N + Ck_{R1} + Ck_{R2} \quad \text{for } k = 1, 2, 3, 4 \tag{6.28}$$

where the N subscript denotes the nominal value, the R1 subscript denotes a random component that is process dependent but does not vary from capacitor to capacitor on a wafer, and the R2 subscript denotes a random component that varies randomly from capacitor to capacitor on a wafer. The nominal values are chosen such that $C1_N = C2_N = C3_N = C4_N$. Since the random components including subscript R1 are all the same on a wafer, equation (6.28) can be rewritten by

$$Ck = C + Ck_{R2} \quad \text{for } k = 1, 2, 3, 4 \tag{6.29}$$

where $C = C1_N + C1_{R1} = C2_N + C2_{R1} = C3_N + C3_{R1} = C4_N + C4_{R1}$.

The capacitor ratio C1/C2 is then

$$\begin{aligned} \frac{C1}{C2} &= \frac{C + C1_{R2}}{C + C2_{R2}} = \frac{C(1 + C1_{R2}/C)}{C(1 + C2_{R2}/C)} \\ &\simeq \left(1 + \frac{C1_{R2}}{C}\right)\left(1 - \frac{C2_{R2}}{C}\right) \\ &\simeq 1 + \frac{C1_{R2}}{C} - \frac{C2_{R2}}{C} \end{aligned} \tag{6.30}$$

Correspondingly,

$$\frac{C3}{C4} \simeq 1 + \frac{C3_{R2}}{C} - \frac{C4_{R2}}{C} \tag{6.31}$$

From (6.30) and (6.31), the ratio C3/C4 can be expressed as

$$\frac{C3}{C4} = \frac{C1}{C2} - \frac{C1_{R2} - C2_{R2} + C4_{R2} - C3_{R2}}{C} \tag{6.32}$$

where the second term denotes the random mismatch component between C1/C2 and C3/C4.

By substituting (6.32) into (6.16), equation (6.17) should be modified as

$$V_2(3) - V_{out}(3) \simeq \frac{C1}{C2}(V_{GS1} - V_{GS2})\frac{1}{1 + \epsilon_1 + \epsilon_2} \tag{6.33}$$

where ϵ_1 is given in (18) and

$$\epsilon_2 \simeq a(z - y)\frac{C1_{R2} - C2_{R2} + C4_{R2} - C3_{R2}}{C} \tag{6.34}$$

The error ϵ_2 associated with the capacitor ratio mismatch is very small because the mismatch component is divided by the op-amp gain A as can be seen in (6.34).

Assume $C1_{R2}$, $C2_{R2}$, $C3_{R2}$, and $C4_{R2}$ are independent random variables with the same standard deviation σ_c, then the standard deviation of ϵ_2 becomes

$$\sigma_{\epsilon_2} = \frac{|z - y|}{A}\frac{2\sigma_c}{C} \tag{6.35}$$

If a 1% tracking error, i.e., $\sigma_c/C = 0.01$, is assumed, then with the conditions in (6.27) the calculated σ_{ϵ_2} is 0.26mV while the calculated deterministic gain error ϵ_1 is 0.41mV. The error due to the capacitor ratio mismatch is small even with a small op-amp gain of 100. However, ϵ_2 can be larger than ϵ_1 when the op-amp gain is large because $\epsilon_2 \propto 1/A$ while $\epsilon_1 \propto 1/A^2$. During ϕ_6 the modified final error voltage is

$$V_{err} = a^2[y'(x-1)(V_{GS1} - V_{GS2}) + x'y'(2V_{GS1} - V_{GS2})] + (\epsilon_1 + \epsilon_2)(V_{GS1} - V_{GS2}) \quad (6.36)$$

6.3.4 Sensitivity to the Op-amp Offset Voltage

With the input V_G grounded the error voltages due to the op-amp offset during ϕ_3 and ϕ_6 are approximately given by

$$V_{out}(3) - V_2(3) \simeq a(y-z)\left(1 + \frac{C3}{C4}\right) V_{OS} \quad (6.37)$$

$$V_{out}(6) \simeq a\left[y'\left(2 + \frac{C4}{C3}\right) + (y-z)\left(1 + \frac{C2}{C1}\right)\frac{C2}{C1}\right] V_{OS} \quad (6.38)$$

where V_{OS} is the op-amp offset voltage. It can be seen that the circuit is also relatively insensitive to the op-amp offset voltage. With the conditions in (6.27), $V_{out}(6) = 0.0985 V_{OS}$, where the op-amp gain is only 100. If $A = 1000$ and V_{OS}=10mV, then the error voltage due to V_{OS} will be less than 0.1mV.

6.3.5 Summary and SWITCAP Simulation Results

Taking into account all the nonideal effects except that due to charge injection, the output voltage during ϕ_6 is given in (6.23), and the final error voltage V_{err} will be

$$V_{err} = \epsilon_{gain} + \epsilon_{mis} + \epsilon_{off} \quad (6.39)$$

where

$$\epsilon_{gain} \simeq \frac{1}{A^2}[\{y'(x-1) + x(y-z)\}(V_{GS1} - V_{GS2}) + x'y'(2V_{GS1} - V_{GS2})] \quad (6.40)$$

$$\sigma_{\epsilon_{mis}} \simeq \frac{1}{A}|(z-y)(V_{GS1} - V_{GS2})|\frac{2\sigma_c}{C} \quad (6.41)$$

$$\epsilon_{off} \simeq \frac{1}{A}\left[y'\left(2 + \frac{C4}{C3}\right) + (y-z)\left(1 + \frac{C2}{C1}\right)\frac{C2}{C1}\right] V_{OS} \quad (6.42)$$

where x, y, and z are given in (6.20)-(6.22), and x' and y' in (6.25) and (6.26).

To examine which error term is dominant, the three error terms have been calculated based on the conditions in (6.27) at various op-amp gains. In this calculation $\sigma_{\epsilon_{mis}}$ is used for ϵ_{mis}

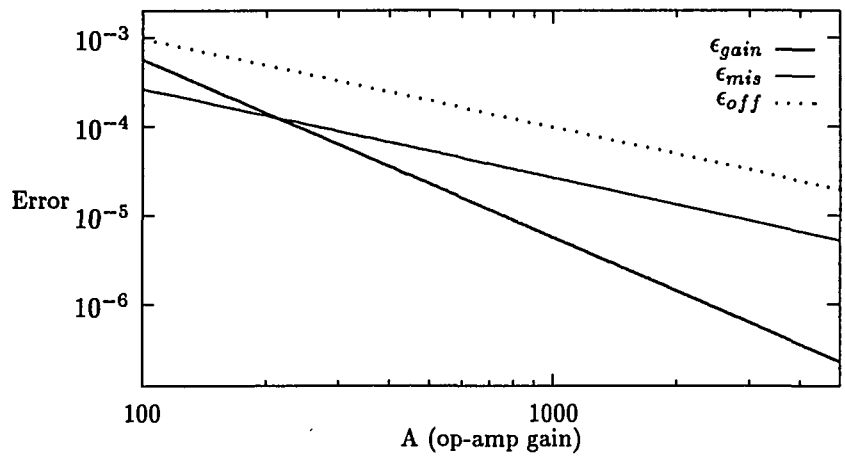


Figure 6.7: Calculated error voltages ϵ_{gain}, ϵ_{mis}, and ϵ_{off} at various op-amp gains ($\epsilon_{mis} = \sigma_{\epsilon_{mis}}$, $\sigma_c/C = 0.05$, and $V_{OS} = 10\text{mV}$)

with $\sigma_c/C = 0.05$, and $V_{OS} = 10\text{mV}$. The calculated results are shown in Fig. 6.7 where it can be seen that the error due to the op-amp offset dominates the other error terms especially at large op-amp gains.

The proposed SC subtracting amplifier has been simulated with SWITCAP. The simulated output error at different op-amp gains with V_{OS} as a parameter is shown in Fig. 6.8. The conditions in (6.27) were used again except that in addition to the parasitic capacitances at nodes (1) and (2), parasitic capacitances (10% as before) associated with all other internal nodes are also considered. It is seen that with op-amp gains greater than 500, the error becomes less than 0.05% even with $V_{OS} = -20\text{mV}$. In this simulation the capacitor ratio mismatch was not considered. Another simulation result is shown in Fig. 6.9 where $A = 500$ and $V_{OS} = -20\text{mV}$ are used. This simulation was done to determine an optimum size of the error storage capacitor Cc. It can be seen that the optimum value of Cc is around 2C (C=C1=C2=C3=C4). With Cc<C, the error is relatively big compared to that with Cc≥C.

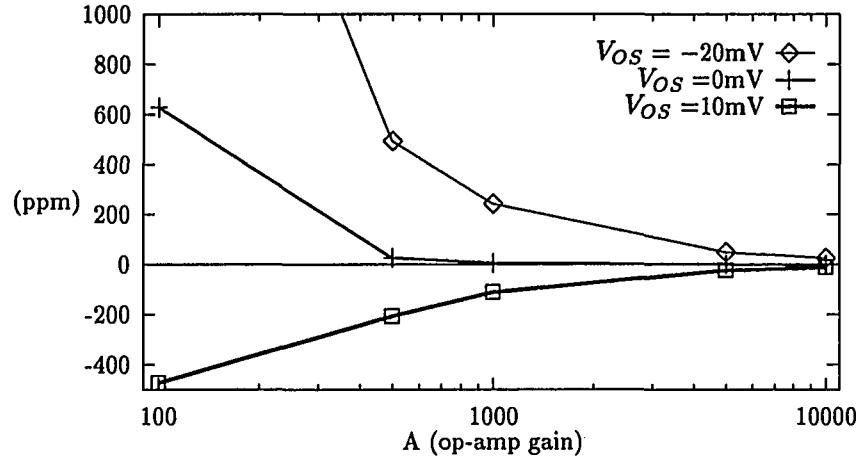


Figure 6.8: SWITCAP simulated output error voltage V_{err} at various op-amp gains with V_{OS} as a parameter

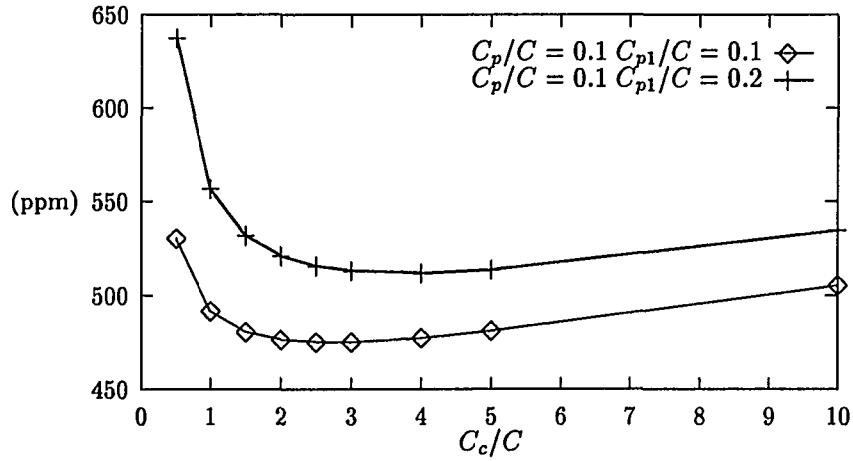


Figure 6.9: SWITCAP simulated output error voltage V_{err} with different sizes of the error storage capacitor Cc (C=C1=C2=C3=C4, A = 500, and V_{OS} = −20mV

6.4 Charge Injection Reduction Schemes

In the previous section it has been demonstrated analytically and through simulations that the proposed SC subtracting amplifier is very accurate due to its insensitivity to the nonideal error sources such as parasitic capacitances, capacitor or capacitor ratio mismatches, finite op-amp gains, and op-amp offset voltages. Now, the remaining error source is the charge injection of the MOS switches. The SC amplifier has been simulated with SPICE using a charge controlled MOS model (XQC=0.5) [107],[137], where charge conservation is guaranteed by the method of computing terminal currents. The simulated error voltage due to charge injection effects is around 10mV. This is somewhat large and thus, should also be compensated to keep the accuracy high.

The charge injection problem has received considerable attention [123]-[130] because it has become the most important factor limiting the accuracy of switched capacitor (SC) and switched current (SI) circuits. Several compensation schemes have been reported in the literature [131]-[136]. In this section, the charge injection phenomenon and its effects are briefly reviewed, and then, a strategy to reduce the charge injection effects in the proposed SC amplifier is proposed.

6.4.1 Charge Injection Phenomenon

When a MOS transistor switch is turned off, the charge stored in its channel is injected into the surrounding nodes, i.e., the source, the drain , and the substrate nodes. This phenomenon is commonly known as the charge injection effect. The effect produced by the charge flowing into the substrate is called charge pumping [130]. The charge pumping effect becomes important when the gate voltage falls very quickly. No significant charge flow has been experimentally observed for switch-off fall times of greater than 5ns [123].

In addition to the channel charge, the charge associated with the feedthrough effect of the gate-to-diffusion overlap capacitance is also injected into the surrounding nodes. The turn-off of a MOS switch consists of two phases as shown in Fig. 6.10. During the first phase, the gate-to-source voltage is higher than the threshold voltage, and thus, both the channel charge and the charge associated with the overlap capacitors are injected. When the gate voltage reaches the threshold voltage, i.e., $V_G = V_S + V_T$, the conduction channel disappears, and the transistor enters the second phase of turnoff. During this phase only the charge of the overlap capacitors is injected until the gate voltage reaches V_{GL}.

The total charge released during the switching off can be expressed as

$$Q_T = C_T(V_{GH} - V_S - V_T) + (C_{ovs} + C_{ovd})(V_T + V_S - V_{GL}) \tag{6.43}$$

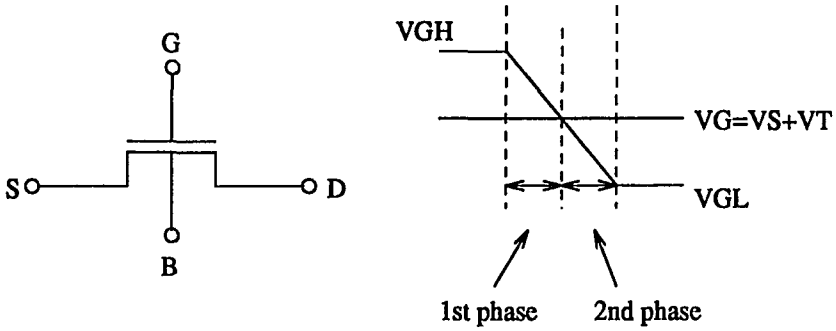


Figure 6.10: Two phases of the MOS switch turnoff

where C_{ovs} and C_{ovd} are the overlap capacitances, V_T is the threshold voltage, and C_T is the total gate capacitance given by

$$C_T = C_{ox}W_{eff}L_{eff} + C_{ovs} + C_{ovd} \tag{6.44}$$

The first term of (6.43) is the charge released during the first phase and the second term is the charge released during the second phase.

The released charge consists of signal dependent terms and signal independent terms. Since the threshold voltage is a nonlinear function of the signal V_S, the magnitude of the charge injected during switching off is also a nonlinear function of the signal. However, if the threshold voltage is assumed to depend linearly on the signal V_S as in [124], according to

$$V_T = V_{T0} + n_0 V_S \tag{6.45}$$

then equation (6.43) can be rewritten as

$$\begin{aligned} Q_T &= \{(C_{ovs} + C_{ovd} - C_T)(1 + n_0)\}V_S + \{C_T(V_{GH} - V_{T0}) + (C_{ovs} + C_{ovd})(V_{T0} - V_{GL})\} \\ &= Q_{gain} + Q_{offset} \end{aligned} \tag{6.46}$$

The first component Q_{gain} causes a gain error since it linearly depends on the signal. The second term Q_{offset} is independent of the signal and thus cause an offset error. If the assumption of (6.45) is not proper, than there will be nonlinear signal dependent terms which will be a source of distortion.

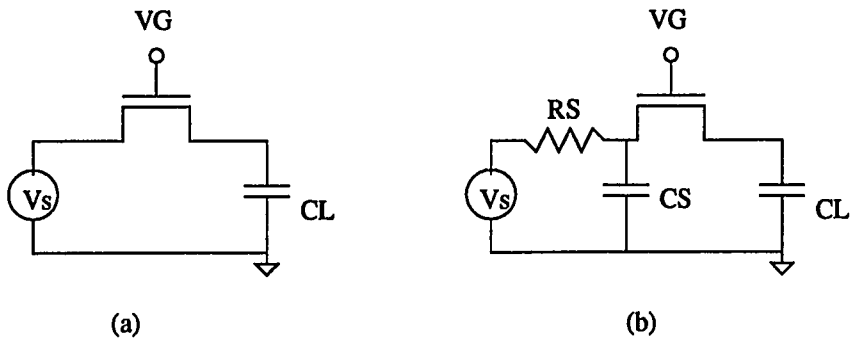


Figure 6.11: Simple sample-and-hold circuits used as test circuits for understanding the effects of charge injection

6.4.2 Effects of Charge Injection

To understand the effects of charge injection the simple sample-and-hold (S/H) circuits shown in Fig. 6.11 have been investigated analytically and experimentally by many researchers [123]-[129]. Their works indicate that the error voltages due to the charge injection are affected by the following factors:

1. Switch turnoff speed
2. Node impedances (C_S, C_L, and R_S)
3. Signal voltage level (V_S)
4. Switch transistor size (L and W)
5. Substrate voltage (V_B)

In the fast switching off conditions, the transistor conduction channel disappears very quickly, and almost equipartition of the channel charge is made independently of the node impedances. Thus, the percentage of charge injected into the data-holding node approaches to 50%. In slow switching off conditions, there is enough time to make the final voltages at both sides equal. This allows the majority of channel charge to flow out through the low impedance node. Since source resistance R_S offers a leakage path for the channel charge during the switch

turn-off period, a small source resistance reduces the amount of charge injected into the data-holding node. The above statements have been demonstrated experimentally and analytically in [123]-[126].

The gate dimension parameters L and W have an important effect on the amount of the injection charge. If the gate area is increased, the total amount of charge stored in the device is increased, as is the charge injection error voltage. It has been shown experimentally [127] and analytically [126] that the charge injection error voltage has a linear dependence on device dimensions. Theoretically, this can be attributed to the linear dependence of the inversion channel charge on the gate area. From this it can be recommended that a minimum-sized switch be used.

The magnitude of the signal V_S affects the amount of charge stored on the gate through the gate-to-source voltage. It also affects the amount of charge stored in the bulk through the source-to-substrate voltage. Linear dependence of the charge injection error voltage on V_S has been observed experimentally in a wide range of V_S [127], where the nonlinear body effect is not significant. This can validate equation (6.45). The substrate voltage V_B also has an effect on the amount of charge stored in the device. However, V_B only contributes via the nonlinear body effect. Thus the change in the charge injection error voltage with V_B is less significant than with V_S.

6.4.3 Charge Injection Compensation Schemes

The following schemes have been used to compensate the error due to charge injection.

1. Minimum-sized switches and/or large-sized storage capacitors
2. Half-sized dummy switches or specially designed dummy switches
3. Dummy capacitors
4. CMOS switches
5. Fully differential structure
6. Scheduling of the timing control sequence

The above techniques or their combination have been employed for many applications. The researchers, used technique numbers, and their applications are summarized in Table 6.3.

A simple approach to reduction of the charge injection error is to use minimum-sized switches and/or large-sized capacitors. However, this approach leads to the decreased circuit

Table 6.3: Summary of commonly used charge injection compensation schemes and their application area

Researchers	Technique Numbers	Applications
Chin [115], Poujois [133], Coln [86]	1	A/D, Amplifiers
Suarez, Gray, and Hodges [132]	2	Charge Redistribution A/D
Li, Chin, Gray, and Castello [114]	5 & 6	Algorithmic A/D
Bienstman and deMan [134]	2 & 3	D/A Converter
Yen and Gray [131]	3 & 5	SC Amplifier
Ogawa [135] and Watanabe [136]	6	SC Amplifier, S/H Circuit

operating speed. The scheme using dummy capacitors along with dummy switches to assure by symmetry that exactly half the channel charge flow into the storage capacitor has limited performance due to the source impedance [134]. Fully differential structure can do a first-order cancellation of the charge injection offset. The gain error term of the charge injection, however, can not be compensated by this structure [114],[131]. The scheme scheduling the timing control signal can effectively compensate the offset error only [135] or the gain error only [114]. In CMOS switches controlled by complementary clock signals, the two types of charge released may partially compensate each other. This scheme is not efficient since it depends on the input signal and the timing skew of the two complementary clocks, and since no real matching exists between PMOS and NMOS [124].

The technique using half-sized dummy switches may provide perfect compensation if the impedances of both sizes of the switch are identical or if the switching off time is in the order of the intrinsic carrier transit time of the switch [123]. For the above two cases equipartition of the channel charge is possible, thus compensation by the dummy switches is guaranteed even for unsymmetrical source and drain impedances if perfect matching exists between the main switches and the dummy switches. For more practical switching off time, the switch will be temporarily conductive, and an equalization of the injected charges may occur. Therefore, no perfect equipartition of the channel charge occurs and half-sized dummy switches can not fully compensate for the charge injection error. More sophisticated rules can be used for the dummy switch design [123]. In any case the unavoidable mismatch and the uncertainty of other parameters limit the achievable compensation accuracy.

6.4.4 Charge Injection Reduction Strategy for the Proposed SC Amplifier

As can be seen in the previous discussion, there does not exist any single scheme that can provide full compensation and can be applicable for all situations. Thus, it may be desirable to use a combination of several schemes. It is important to select the schemes that are most suitable for the circuit to be compensated and to apply them intelligently. For the proposed circuit, several schemes are incorporated to obtain a charge injection error voltage less than 1mV.

Since the operating speed is not critical in our circuit, small-sized switches (W/L=4μm/2μm) are used to reduce the amount of charge to be taken care of. The capacitor values have been selected such that C1=C2=C3=C4=4pF and Cc=8pF. Very large-sized capacitors can be used to reduce the error voltages due to injected charge, but this will increase the required area significantly and also reduce the operating speed. If V_{GH}=5V, V_{GL}=0V, V_S=1V, and W/L=4μm/2μm, then the total charge released during the switching off is about 26.7fC which was calculated from equation (6.43). Assuming half the total charge is injected into the capacitor (C=4pF), the error voltage will be 3.34mV. To obtain the overall circuit error voltage of less than 1mV, other compensation schemes are required. A fully differential structure is excluded because it needs a much more complicated circuit and increased area, and it can compensate the charge injection offset error only.

The scheme using half-sized dummy switches can be generally applied for any types of SC circuits if equipartition of channel charge is possible. Thus, half-sized dummy switches are used in our circuit along with a fast falling gate clock which ensures almost equipartition of channel charge such that the dummy switches can compensate it. The gate voltage falling rate should be selected carefully. If the falling rate is too fast, the charge pumping effects [130] will be significant. If the falling rate is slow, then the deviation from the equipartition will increase. A gate clock falling rate of 5V/5nsec has been selected because no significant charge pumping effects was experimentally observed down to 5nsec in [123], and SPICE simulations showed that with the switch-off fall time of 5nsec the deviation from 1:1 partition is less than 5% for most practical node impedance conditions. By using this the overall error voltage due to charge injection is expected to be greatly reduced although no perfect equipartition of the channel charge is possible, and the mismatches between the main and the dummy switches degrade the compensation accuracy.

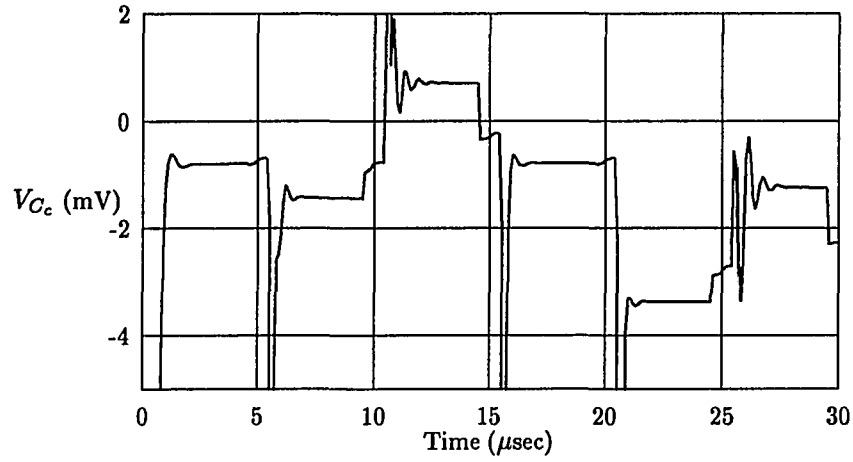
Without dummy switches the simulated charge injection error voltage was around 10mV as mentioned before where the designed op-amp has a dc gain of 800. Charge injection corrupts the signals sampled onto C1 and C3, or transferred onto C2 and C4. More seriously, due

to the charge injected into nodes (1) and (2) (see Fig. 6.5) the information stored in Cc is substantially corrupted when Cc is used for the main operations. Fig. 6.12 show the simulated voltage difference between node (2) and (1), i.e., V_{C_c} $(V2 - V1)$, during ϕ_1 to ϕ_6, where the period of one clock phase is 5μsec. The capacitor Cc stores the error voltages generated during the preliminary periods ϕ_2 and ϕ_4, and the information stored in Cc must be used without any change during the corresponding main periods ϕ_4 and ϕ_6, respectively. However, the simulated results indicate about 2mV change in V_{C_c} as shown in Fig. 6.12(a). The error voltage stored in Cc during ϕ_2 is $-1.4mV$, but this is increased to $0.7mV$ during ϕ_3. Thus, the corrupted information is used for the main operations, resulting in a degraded accuracy.

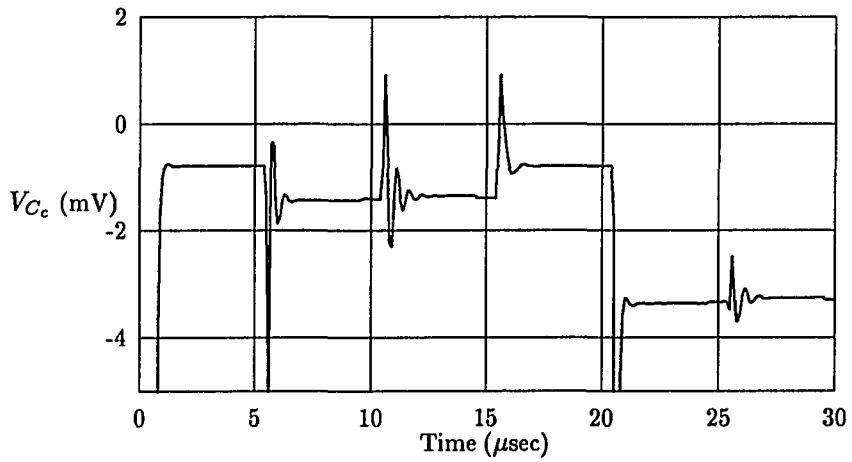
A very small change in V_{C_c} of about 0.1mV can be observed in Fig. 6.12(b) where half-sized dummy switch compensation was used. It can be seen that the dummy switches compensate the charge injected into node (1) and (2) and thus, greatly reduce the change of the information stored in Cc. When V_{GS1}=1.4V and V_{GS2}=1.8V, the simulated overall output error voltage is 0.6mV which is a greatly reduced value compared with 10mV obtained without compensation. This accuracy well satisfies our targeted accuracy of 1mV. The simulated op-amp output voltage is shown in Fig. 6.13 where a 30mV offset voltage source is inserted at the noninverting input terminal of the op-amp. In the figure the preliminary operations which are erroneous due to the offset voltage and the finite op-amp gain, and the compensated main operations can be easily distinguished. The accuracy can be degraded by the nonideal factors associated the dummy switch compensation such as mismatches between the main and the dummy switches and clock skews. These effects are investigated next.

6.4.5 Consideration of the Nonideal Effects Associated with Dummy-Switch Compensation

Besides the simple S/H circuit in Fig. 6.11, another S/H structure which is widely used in SC circuits consists of a floating sampling capacitor between two MOS switches as shown in Fig. 6.14(a). This structure has been preferably used especially in stray-insensitive SC circuits such as SC filters and SC arithmetic building blocks. The structure is also a basic element in our SC subtracting amplifier as can be seen in Fig. 6.5. The dummy-switch compensated version is shown in Fig. 6.14(b). The charge injection effects of this circuit can also be investigated analytically. However its analysis will be much more complicated compared to the simple circuits in Fig. 6.11 which were analyzed theoretically in [123]-[126]. Even for the simple circuit, no closed-form solution exists for general cases. Thus, theoretical analysis of the circuit



(a)



(b)

Figure 6.12: SPICE simulated voltage across the error storage capacitor Cc, V_{C_c} (a) Without dummy-switch compensation (b) With dummy-switch compensation

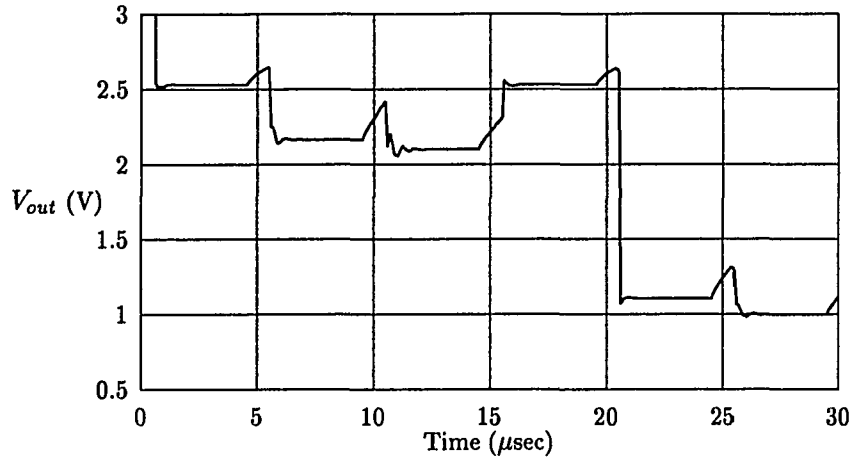
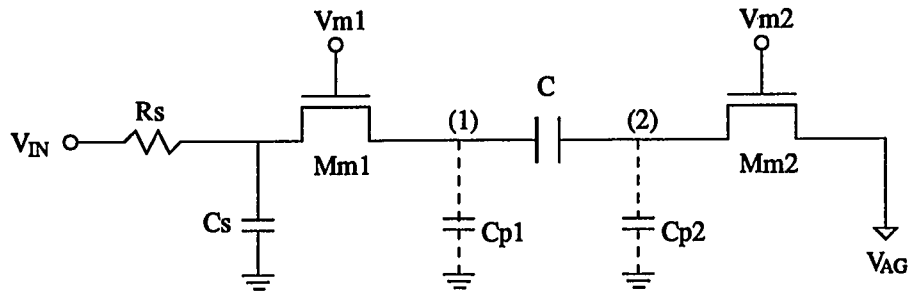


Figure 6.13: Simulated op-amp output voltage of the SC subtracting amplifier in Fig. 6.5 when V_{GS1}=1.4V, V_{GS2}=1.8V, V_{OS}=30mV, and A=800

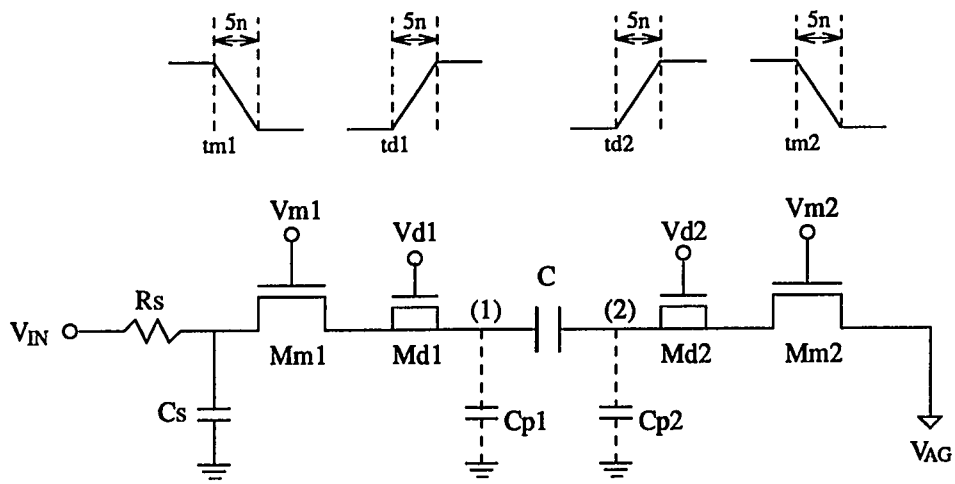
in Fig. 6.14 is avoided. Instead, SPICE is used to investigate the nonideal effects associated with the dummy-switch compensation shown in Fig. 6.14(b).

In this simulations the following parameters were used: analog ground V_{AG}=2.5V, gate high voltage V_{GH}=5V, gate low voltage V_{GL}=0V, gate voltage falling time TF=5nsec, R_S=10kΩ, and C_S=1pF. Since the sampling capacitor is floating, the parasitic capacitors at both nodes and their mismatch will affect the charge injection process. Simulated charge injection error voltages V_{Cerr}, $V_{Cerr} = (V_{IN} - V_{AG}) - (V1 - V2)$, are depicted in Fig. 6.15(a) for the uncompensated circuit of Fig. 6.14(a) and in Fig. 6.15(b) for the compensated circuit of Fig. 6.14(b) as a function of $C_{p1} - C_{p2}$. From the simulated results it can be seen that the effect of the parasitic capacitance mismatch on the charge injection error voltage is much larger for the uncompensated circuit than for the compensated one. The half-sized dummy switch compensation is still sensitive to the input signal V_{IN} as shown in Fig. 6.15(b) where simulations were done for four different input voltages from 1V to 2.5V. The input voltage range of 1.0V to 2.5V can cover most NMOS test transistors which have different geometries and different substrate bias conditions.

It can be also seen that the sensitivity to input signals decreases as $C_{p1} - C_{p2}$ increases. Fortunately, this kind of parasitic capacitance mismatch is natural because an integrated ca-

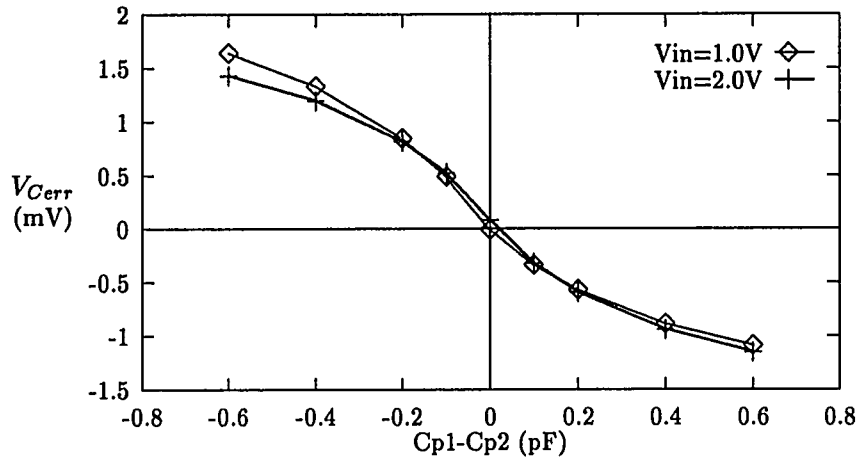


(a)

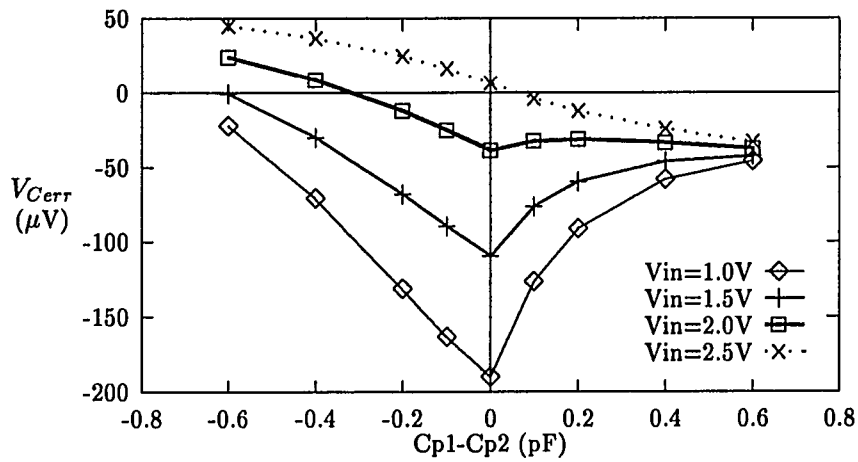


(b)

Figure 6.14: (a) A S/H structure consisting of a floating capacitor between two MOS switches (b) The half-sized dummy switch compensated version of (a)



(a)



(b)

Figure 6.15: Sensitivity of the charge injection error voltage V_{Cerr} to the mismatch of parasitic capacitances C_{p1} and C_{p2} (a) for the uncompensated case of Fig. 6.14(a) (b) for the compensated case of Fig. 6.14(b) ($V_{Cerr} = (V_{IN}-V_{AG})-(V1-V2)$)

pacitor (C) has a large parasitic capacitance (0.1C $\sim$ 0.2C) associated with the bottom plate and a small parasitic capacitance (0.01C $\sim$ 0.05C) associated with the top plate. Thus, the bottom plate, which is denoted with a thicker line in the figure, should be connected to node (1), the input signal side. If the input signal is close to the analog ground, then the error voltage variation is small regardless of the parasitic capacitance mismatch. In Fig. 6.5 the bottom plates are also denoted with thicker lines, and the signal levels associated with capacitors C2, C4, and Cc are the analog ground or the op-amp output. The op-amp output differs from the analog ground by only $V_{GS1} - V_{GS2}$ during ϕ_2 and ϕ_3. The input resistance R_S has little effect on the results because a fast falling gate clock is used. The effect of the input capacitance C_S is also negligible unless it is too large.

The effect of size mismatch between the main and the dummy switches on the compensation process has been simulated. The simulated results are shown in Fig. 6.16 where the parameter of the horizontal axis is $(A_m - 2A_d)/2A_d$, and A_m and A_d denote the area of main switches and the area of dummy switches, respectively. With 20% area mismatch, the variance is about 100μV when $V_{IN} = 2.5$V. Fig. 6.17 shows the effects of clock skews. The effects of delayed switching of the dummy switches is shown in Fig. 6.17(a) where the variance is less than 50μV for $V_{IN} = 1$V, and it becomes constant for long delay modes. Fig. 6.17(b) shows the effects of the clock skew between the two main switches. These simulation results indicate that the effects of the nonideal factors on the dummy-switch compensation process are not significant.

6.5 Dynamic Current Mirror

The current mirror block shown in Fig. 6.2 is implemented dynamically to supply accurate currents I_D and $4I_D$ to test devices. The dynamic analog techniques utilize an inherently attractive property of MOS transistors that analog information can be stored on the gate capacitor since no gate current is required in MOS transistors. Recently, this dynamic concept has been widely used to accurately implement analog circuits such as current mirrors [138]-[140], data converters [141],[142], and switched current (SI) circuits and filters [144]-[147].

By applying the dynamic concept to current mirrors, accurate current mirroring is possible without depending on the transistor matching. However, some by-products associated with the dynamic technique newly occur such as charge injection effects of MOS switches, transient effects when switching, leakage current in the sampling switches. To reduce the finite output conductance effects of the current mirror, the self-biased stacked mirror concept proposed by

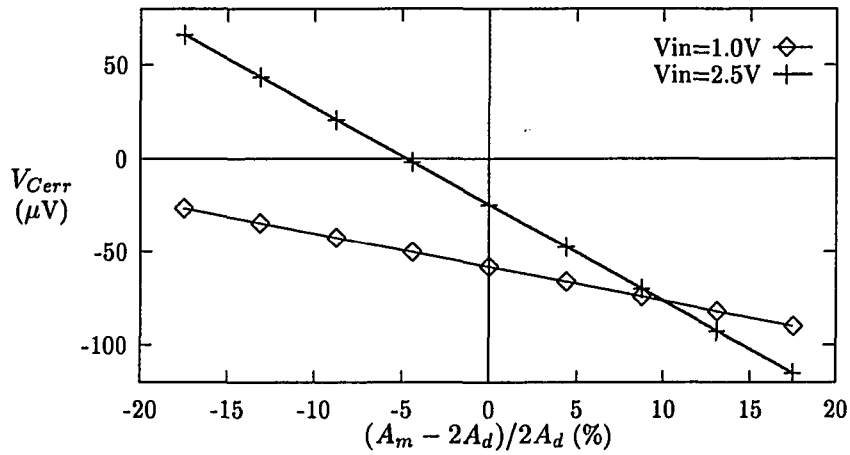
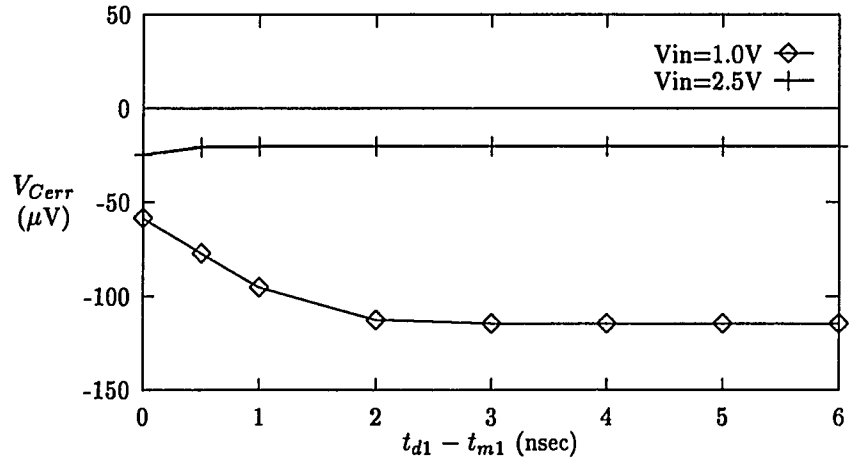
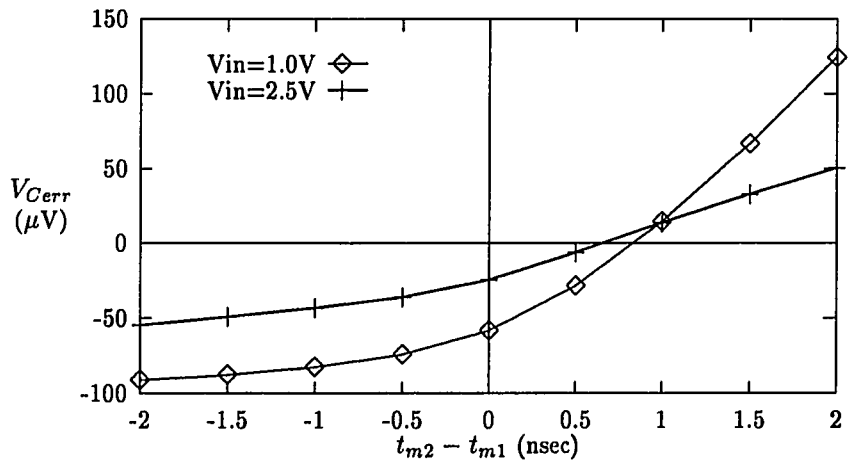


Figure 6.16: Sensitivity of the error voltage V_{Cerr} to the mismatch between the main switch area of A_m and the dummy switch area of A_d (C_{p1} = 0.6pF and C_{p2} = 0.2pF)



(a)



(b)

Figure 6.17: Sensitivity of the error voltage V_{Cerr} to the clock skew as a function of (a) $t_{d1} - t_{m1}$, where $V_{m1} = V_{m2}$ and $V_{d1} = V_{d2}$ (b) $t_{m2} - t_{m1}$, where $V_{d1} = \overline{V_{m1}}$ and $V_{d2} = \overline{V_{m2}}$ ($C_{p1} = 0.6$pF and $C_{p2} = 0.2$pF)

Wegmann and Vittoz [138]-[140] is used in our circuit.

6.5.1 Principle of Operation

The schematic of the dynamic current mirror and the required clock phases are shown in Fig. 6.18. To correlate the clock phases with those used for the SC subtracting amplifier, the first switching scheme shown in Fig. 6.2(a) was selected for implementing the current mirror.

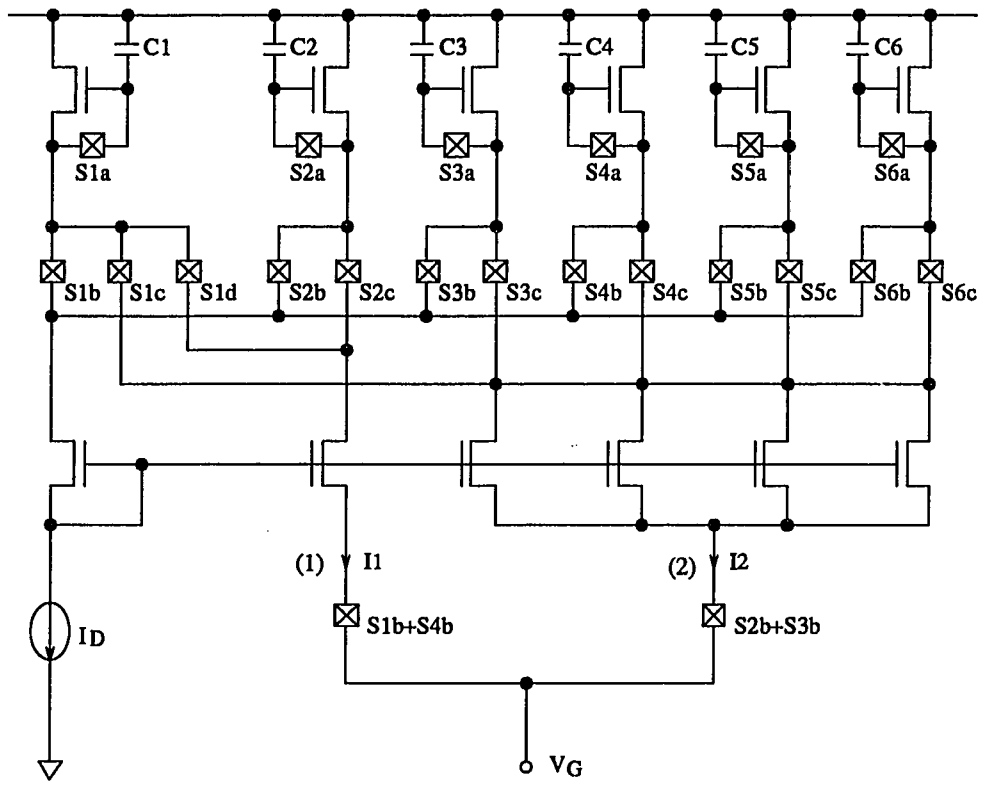
The dynamic current mirror is composed of six current copier cells proposed in [143]. Each cell consists of a sampling switch S_{ia}, ($i = 1, 2, \ldots$ or 6), a storage capacitor C_i, and a PMOS transistor. Switches S_{ib} and S_{ic} for $i = 1, 2, \ldots, 6$, are used to periodically connect the cells with the input I_D for refreshing the stored information and with the output for supplying the mirrored currents. The stacked common-gate transistors which are employed to increase the output impedance are connected such that one cell is always connected with node (1) to deliver current I_1 (ideally I_D), four cells are always connected with node (2) to deliver current I_2 (ideally $4I_D$), and remaining one cell is connected to the input bias current I_D for refreshing.

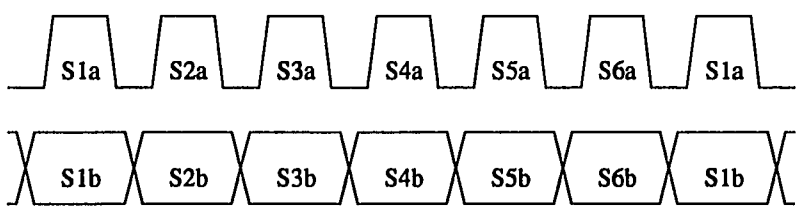
When switches S_{ia} and S_{ib} are closed to memorize the input current I_D, the sampling switch S_{ia} must be opened first as shown in the clock phase diagram in Fig. 6.18 in order not to contaminate the stored information. Once S_{ia} is open, the gate voltage is kept constant if leakage current in the sampling switch is ignored such that the drain current remains equal to I_D. When S_{ib} is opened, and S_{ic} or S_{1d} is closed, the memorized current is available at the output. The transients occurred when S_{ib} and S_{ic} are switched can be a significant error source for continuous-time applications as investigated in [138],[139]. In our circuit the transient effects is not important because the currents are required for only specific time intervals, which indicates that the dynamic current mirror is suitable for our V_T extractor.

Clock phases of the current mirror, S_{ia}, are correlated with clock phases of the SC subtracting amplifier, ϕ_i, as mentioned before such that the current mirror can supply I_D during ϕ_1 and ϕ_4 and $4I_D$ during ϕ_2 and ϕ_3 to the test device (see two switches connected to node (1) and (2) in Fig. 6.18). In fact, the switches in the current mirror are PMOS transistors, and thus the polarity should be inverted as follows:

$$\phi_i = \overline{S_{ia}} \qquad \text{for } i = 1, 2, \ldots, 6$$

The output V_G is directly connected to the input of the SC subtracting amplifier and to a diode-connected test device.





S1c=S3b+S4b+S5b+S6b, S1d=S2b

Sic = $\overline{\text{Sib}}$ for i=2,3,4,5,6

Figure 6.18: Schematic of dynamic current mirror and required clock phases

6.5.2 Accuracy Simulations

The charge injection problem from the sampling switches is also compensated by the same strategy as used in the SC subtracting amplifier. Half-sized PMOS switches are inserted between the sampling switches and the storage capacitors. The switch gates are driven by fast rising clocks. Since the leakage current problem is not important if a proper clock frequency is used, the nonideal effects limiting performance of the current mirror have been taken care of. The finite output conductance effects are reduced by using the stacked transistors, and the transients effect is not a problem in our circuit as mentioned before.

The dynamic current mirror in Fig. 6.18 was simulated when node (1) and (2) were connected with a NMOS transistor (W/L=20μm/4μm). The simulated output currents I_1 and I_2 are depicted in Fig. 6.19. The initial behaviors of the current copier cells to produce output currents of ratio 1:4 can be observed until $t = 30\mu$sec. After the initial cycle, the current mirror can supply the currents I_1 and I_2 of which the ratio is ideally 1:4 to the test device. The output current ratio accuracy is shown in Fig. 6.20 as a function of the input bias current I_D. Since the ratio error for one I_D value varies slightly at different clock phases, the maximum values are selected and shown on the figure. The ratio errors in the I_D range in interest are less than 700 ppm which produces approximately 0.5mV error in the arithmetic operation of $2V_{GS1} - V_{GS2}$. Therefore, along with the SC subtracting amplifier discussed in the previous sections the dynamic current mirror can perform the proposed V_T extraction scheme accurately.

6.6 Conclusions

An accurate real-time V_T extraction scheme which does not need matched replica of the device under test has been proposed. A ratio-independent and finite gain insensitive switched-capacitor subtracting amplifier and a dynamic current mirror have been designed to perform the proposed scheme accurately in a matching-free way. Model error associated with the proposed scheme has been investigated and compared with the linear regression method. The nonideal factors limiting the performance of the SC amplifier and the dynamic current mirror have been thoroughly investigated and their effects have been compensated in design.

Extensive simulation results show the potential of the proposed V_T extractor in accuracy. Taking into account unexpected process variations, the total error voltages associated with the designed circuit are in a few millivolt range. This error is smaller compared with the model error. To make the V_T extractor applicable to various transistors which has different geometries

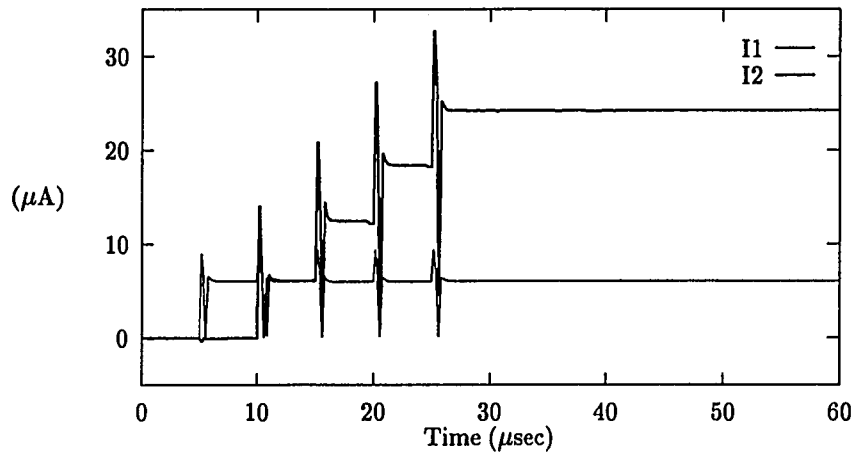


Figure 6.19: Output currents I_1 and I_2 of the dynamic current mirror of ratio 1:4

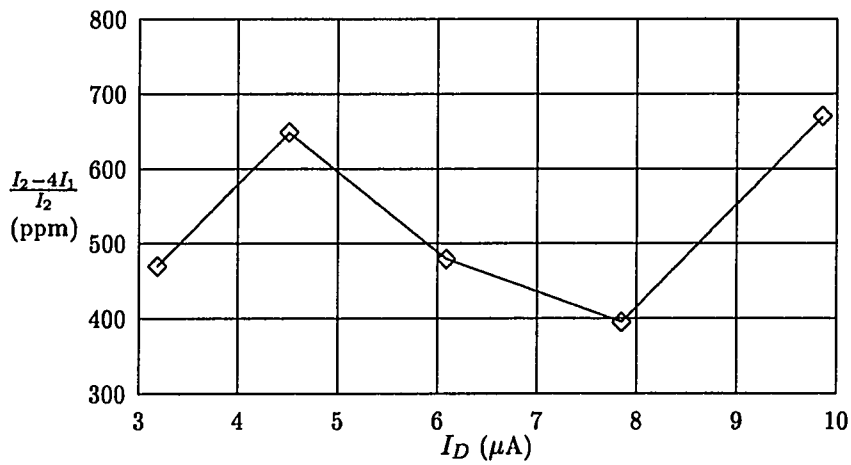


Figure 6.20: Simulated current ratio error $\frac{I_2-4I_1}{I_2}$ of the dynamic current mirror as a function of the input current I_D

and different bias conditions and to achieve a high accuracy, the model error should be always kept small. The model error can be reduced by using an adaptive biasing scheme such that the excess voltage of test transistors are always kept small.

The scheme is applicable to various applications where many V_T measurements are required. For example, the scheme can be well applied for implementation of low-voltage floating-gate MOSFET operational amplifiers presented in the previous chapter where V_T measurement of many floating-gate MOSFETs with different geometries are essential for V_T tuning [3].

CHAPTER 7. CONCLUSIONS

In this dissertation five topics are investigated which are concerned with theories and techniques for high-precision linear integrated circuit design and implementation. Since each topic has been already concluded, only brief summary are given in this chapter.

In Chapter 2, a digital tuning scheme was presented for digitally programmable/tunable continuous-time filters. The tuning scheme consists of two steps, system identification (ID) and adjustment. Among various continuous-time system ID methods two indirect methods have been investigated. One is a time-domain approach where a discrete-time model is first estimated from input-output samples, and then it is transformed into an equivalent continuous-time model. The other is a frequency-domain approach where frequency response of the filter are first measured by frequency response measurement algorithms from input-output samples, and a continuous-time model is then estimated by s-domain system ID algorithms based on the measured frequency response data. Very accurate domain transformation methods were presented. It has been shown that transformed results by the complex LS s-to-z and z-to-s methods are much more accurate than those by the well-known bilinear method. As a robust s-domain system ID method, an iterative complex LS algorithm was presented. While it has been demonstrated from extensive simulations that both approaches can be fairly well applicable to the tuning scheme, the frequency-domain approach has been combined with an adjustment algorithm to serve as the digital tuning scheme because of its applicability to high-frequency applications with low-cost data acquisition circuits. Extensive simulations and experimental results have demonstrated that the digital tuning scheme can be applicable with fairly good accuracy to high-frequency and high-Q filters as well as to various filter functions.

In Chapter 3, the common-mode rejection ratio and the offset of two-stage CMOS op-amps have been investigated. Equations representing their statistical characteristics have been derived from which the distribution, mean, and variance of the CMRR and offset can be easily obtained if the process parameter variations are given. It has been shown that the random

common-mode gain as determined by the mismatch of paired devices is comparable to the deterministic common-mode gain. It has also been shown that the probability density function of the CMRR is distributed similar to that of a Gaussian random variable, but the mean is finite and the symmetry is skewed somewhat, as contrasted to the probability density function of the offset voltage which has a Gaussian distribution with zero mean. The op-amp errors associated with finite open-loop gains, finite CMRRs, and nonzero offset voltages have been analyzed. It has been shown that a nonideal finite CMRR can actually reduce the op-amp errors caused by a finite open-loop gain.

In Chapter 4 an automatic offset compensation scheme for CMOS operational amplifiers was presented. The proposed offset reduction scheme is to use a programmable current mirror instead of a conventional one as a load of the op-amp differential input stage for control of the offset by adjusting the bias voltage of the programmable current mirror. By employing a ping-pong structure, continuous-time operation is obtained while the offset is constantly compensated which makes the scheme insensitive to time and temperature drift. The performance of the proposed scheme has been experimentally investigated. The proposed circuit has been fabricated using a 1.0-μm n-well CMOS process. The measured offset voltages of the test circuits are less than 400μV in magnitude. The resolution can be improved by increasing the number of bits of the digital-to-analog converter and using a unipolar compensation scheme. It has been experimentally shown that the transient effects associated with the ping-pong operation are not problematic. Several methods have also been proposed to further reduce the transient effects.

In Chapter 5, a threshold voltage tunable op-amp structure that can be operated with a very low power supply has been presented. Floating gate MOS transistors have been employed as the basic op-amp circuit elements. By reducing the threshold voltages of the floating gate MOS transistors, the op-amp circuit can operate with a very low power supply. Detailed circuit implementation methods have been discussed. Good matching can also be achieved by tuning the threshold voltages. A two-step threshold voltage tuning scheme has been presented. Its functionality has been demonstrated through simulations. The basic low-voltage methodology can be extended to other analog circuits as well as digital applications. Thus, implementation, characterization, and extension of the proposed low-voltage scheme can be a good future research topic.

An accurate threshold voltage extraction scheme for MOS transistors has been presented in Chapter 6. The proposed scheme differs from conventional methods in that it does not need matched replica of the transistor under test and thus, can be applied more easily and accurately than any others to real-time on-chip applications where threshold voltage measurement are required for many transistors with various geometries and bias conditions. The proposed circuit has been designed in a matching-free way using a ratio-independent switched-capacitor subtracting amplifier and a dynamic current mirror. Nonideal effects associated with these circuits have been thoroughly investigated. Simulation results have shown that the error associated with the designed V_T extraction circuit is in a few millivolt range. The scheme can also be well applied to the low-voltage circuits presented in Chapter 5 where V_T measurement of many floating-gate MOS transistors with different geometries are essential for V_T tuning.

REFERENCES

[1] C. G. Yu and R. L. Geiger, "An Accurate and Matching-Free Threshold Voltage Extraction Scheme for MOS Transistors," V_T Extractor," submitted to *IEEE International Symposium on Circuits and Systems,* 1994.

[2] C. G. Yu and R. L. Geiger, "An Automatic Offset Compensation Scheme with Ping-Pong Control for CMOS Operational Amplifiers," has been accepted for publication in *J. of Solid State Circuits.*

[3] C. G. Yu and R. L. Geiger, "Very Low Voltage Operational Amplifiers Using Floating Gate MOS Transistors," *IEEE International Symposium on Circuits and Systems,* Vol. 2, pp. 1152-1155, May 1993.

[4] C. G. Yu and R. L. Geiger, "Frequency Response Measurement algorithms," *Second Annual Midwest Electro-Technology Conference,* pp. 56-59, April 1993.

[5] C. G. Yu and R. L. Geiger, "Nonideality Consideration for High-Precision Amplifiers – Analysis of Random Common-Mode Rejection Ratio," *IEEE Trans. on Circuits and Systems Part I,* Vol. 40, No. 1, pp. 1-12, Jan. 1993.

[6] C. G. Yu and R. L. Geiger, "Precision Offset Compensated OP-Amp with Ping-Pong Control," *Government Microcircuit Applications Conference,* pp. 189-190, Dec. 1992.

[7] C. G. Yu and R. L. Geiger, "Analysis of Random Common-Mode Rejection Ratio in Operational Amplifiers," *35th Midwest Symposium on Circuits and Systems,* pp. 949-952, Aug. 1992.

[8] C. G. Yu, W. G. Bliss and R. L. Geiger, "A Tuning Algorithm for Digitally Programmable Continuous-Time Filters," *IEEE International Symposium on Circuits and Systems,* pp. 1436-1439, June 1991.

[9] C. G. Yu, W. G. Bliss and R. L. Geiger, "A Frequency Domain Parametric System Identification Algorithm for Continuous–Time Filters," *IEEE International Symposium on Circuits and Systems*, pp. 1725-1728, June 1991.

[10] C. G. Yu, W. G. Bliss and R. L. Geiger, "Performance of a High–Precision Digital Tuning Algorithm," *34th Midwest Symposium on Circuits and Systems*, pp. 356-359, May 1991.

[11] R. L. Geiger, P. E. Allen, and N. R. Strader, *VLSI Design Techniques for Analog and Digital Circuits*, McGraw-Hill, New York, 1990.

[12] Y. Tsividis and P. Antognetti, *Design of MOS VLSI Circuits for Telecommunications*, Prentice-Hall, Englewood Cliffs, NJ, 1985.

[13] B. Nauta, "A CMOS transconductance-C filter technique for very high frequencies," *IEEE J. of Solid-State Circuits*, Vol. 27, pp. 142-153, Feb. 1992.

[14] V. Gopinathan, Y. P. Tsividis, K. S. Tan, and R. K. Hester, "Design considerations for high-frequency continuous-time filters and implementation of an antialiasing filter for digital video," *IEEE J. of Solid-State Circuits*, Vol. 25, pp. 1368-1378, Dec. 1990.

[15] C. S. Park and R. Schaumann, "Design of a 4-MHz analog integrated CMOS transconductance-C bandpass filter," *IEEE J. of Solid-State Circuits*, Vol. SC-23, pp. 987-996, August 1988.

[16] C. S. Park and R. Schaumann, "Design of an eighth-order fully integrated CMOS 4MHz continuous-time bandpass filter with digital/analog control of frequency and quality factor," in *Proc. Int. Symp. CAS*, pp. 754-757, May 1987.

[17] P. M. Vanpeteghem and R. Song, "Tuning strategies in high-frequency integrated continuous-time filters," *IEEE Trans. on Circuits and Systems*, Vol. 36, pp. 136-139, January 1989.

[18] H. Khorramabadi and P. R. Gray, "High-frequency CMOS continuous-time filters," *IEEE J. of Solid-State Circuits*, Vol. SC-19, pp. 939-948, December 1984.

[19] Y. Tsividis, M. Banu and J. Khoury, "Continuous-time MOSFET-C filters in VLSI," *IEEE Trans. on Circuits and Systems*, Vol. CAS-33, pp. 125-138, February 1986.

[20] F. Krummenacher and N. Joel, "A 4-MHz CMOS continuous-time filter with on-chip automatic tuning," *IEEE J. of Solid-State Circuits*, Vol. SC-23, pp. 750-758, June 1988.

[21] M. Banu and Y. Tsividis, "An elliptic continuous-time CMOS filter with on-chip automatic tuning," *IEEE J. of Solid-State Circuits*, Vol. SC-20, pp. 1114-1121, December 1985.

[22] D. Senderowicz, D. A. Hodges, and P. R. Gray, "An NMOS integrated vector-locked loop," in *Proc. IEEE Int. Symp. CAS*, pp. 1164-1167, 1982.

[23] K. S. Tan and P. R. Gray, "Fully integrated analog filters using bipolar-JFET technology," *IEEE J. of Solid-State Circuits*, Vol. SC-13, pp. 814-821, December 1978.

[24] R. L. Geiger, P. E. Allen and T. Ngo, "switched resistor filters– A continuous-time approach to monolithic MOS filter design," *IEEE Trans. on Circuits and Systems*, Vol. 29, pp. 306-315, May 1982.

[25] Y. Tsividis, "Self-tuned filters," *Electron Lett.*, Vol. 17, pp. 406-407, June 1981.

[26] D. E. Hocevar and T. N. Trick, "Automatic tuning algorithms for active filters," *IEEE Trans. on Circuits and System*, Vol. CAS-29, pp. 448-458, July 1982.

[27] D. E. Hocevar, *Automatic tuning algorithm and statistical circuit design*, Ph.D. Dissertation, Univ. Illinois, Urbana-Champaign, 1982.

[28] P. V. Lopresti, "Optimum design of linear tuning algorithm," *IEEE Trans. on Circuits and System*, Vol. CAS-24, pp. 144-151, March 1977.

[29] P. V. Lopresti and K. R. Laker, "Optimum tuning of multi-loop feedback active-*RC* filters," in *Proc. IEEE Int. Symp. Circuits Syst.*, pp. 812-815, April 1980.

[30] C. J. Alajajian, T. N. Trick, and E. I. El-Masry, "On the design of an efficient tuning algorithm," in *Proc. IEEE Int. Symp. Circuits Syst.*, pp. 807-811, April 1980.

[31] E. Gleissner, "Some aspects on the computer aided tuning of networks," in *Proc. IEEE Int. Symp. Circuits Syst.*, pp. 726-729, April 1976.

[32] T. D. Shockley and C. F. Morris, "Computerized design and tuning of active filters," *IEEE Trans. on Circuit Theory*, pp. 438-441, July 1973.

[33] R. Gejji, "A method for relaxing tolerance requirements in *RC* filters," *IEEE Trans. on Circuits and System*, Vol. 35, pp. 1311-1317, October 1988.

[34] J. F. Pinel, "Computer-aided network tuning," *IEEE Trans. on Circuit Theory*, pp. 192-194, January 1971.

[35] J. W. Bandler and P. C. Liu, "Automated network design with optimal tolerances," *IEEE Trans. on Circuits and System*, Vol. CAS-21, pp. 219-222, March 1974.

[36] S. W. Director and G. D. Hachtel, "The simplical approximation approach to design centering," *IEEE Trans. on Circuits and System*, Vol. CAS-24, pp. 363-371, July 1977.

[37] E. Lueder and G. Malek, "Measure-predict tuning of hybrid thin-film filters," *IEEE Trans. on Circuits and System*, Vol. CAS-23, pp. 461-465, July 1976.

[38] K. H. Loh, D. L. Hiser, W. J. Adams and R. L. Geiger, "A robust digitally programmable and reconfigurable monolithic filter structure," *IEEE International Symposium on Circuits and Systems*, Vol. 1, pp. 110-113, May 1989.

[39] K. H. Loh, D. L. Hiser, W. J. Adams and R. L. Geiger, "A versatile digitally controlled continuous-time filter structure with wide range and fine resolution capability," *IEEE Trans. on Circuits and Systems Part II*, Vol. 39, pp. 265-276, May, 1992.

[40] K. H. Loh, *On the tuning of digitally programmable continuous-time filters*, Ph.D. Dissertation, Texas A&M University, College Station, August 1991.

[41] D. L. Hiser, *A programmable digitally tuned CMOS OTA-C 2nd-order bandpass filter architecture*, Ph.D. Dissertation, Texas A&M University, College Station, August 1990.

[42] A. P. Nedungadi, *Design of linear transconductance elements in CMOS technology*, Ph.D. Dissertation, Texas A&M University, College Station, Texas, December 1987.

[43] W. J. Adams, A. Nedungadi and R. L. Geiger, "Design of a programmable OTA with multi-decade transconductance adjustment," in *Proc. Int. Symp. CAS*, pp. 663-666, May 1989.

[44] T. C. Hsia, "On least squares algorithm for system parameter identification," *IEEE Trans. Automatic Control,* vol. AC-17, pp. 104-108, Feb. 1976.

[45] N. K. Sinha and B. Kuszta, *Modeling and Identification of Dynamic Systems,* Van Nostrand Reinhold Company, New York, CH. 3, 1983.

[46] K. J. Astrom and P. Eykhoff, "System identification - A survey," *Automatica,* vol. 7, pp. 123-162, Mar. 1971.

[47] V. Strejc, "Least squares parameter estimation," *Automatica,* vol. 16, pp. 535-550, Sep. 1980.

[48] V. Strejc, "Trends in identification," system parameter identification," *Autamatica,* vol. 17, pp. 7-22, Jan. 1981.

[49] L. Ljung and T. Soderstrom, *Theory and Practice of Recursive Identification,* The MIT Press, Cambridge, Massachusetts, 1983.

[50] T. Soderstrom, "Convergence properties of the generalized least squares identification method," *Automatica,* vol. 10, pp. 617-626, 1974.

[51] D. W. Clarke, "Generalized-least-squares estimation of parameters of a dynamic model," *IFAC Symposium-Identification in Automatic Control Systems,* Prague, 1967.

[52] G. C. Goodwin and R. L. Payne, *Dynamic System Identification,* Academic press, New York, 1977.

[53] G. F. Franklin, J. D. Powell, and M. L. Workman, *Digital Control of Dynamic Systems,* Addison-Wesley, Reading, MA, CH. 8, 1990.

[54] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing,* Cambridge University Press, Cambridge, 1988.

[55] T. C. Hsia, *System Identification,* Lexington Books, Lexington, MA, 1977.

[56] G. Berchin and M. A. Soderstrand, "A total least squares approach to frequency domain system identification," pp. 709-711, 1990.

[57] G. Berchin, R. Roberts, and M. A. Soderstrand, "New model-based parameter estimation technique for use in deconvolution," *Twentieth Annual Asilomar Conference on Signals, Systems, and Computers,* pp. 666-670, 1987.

[58] N. K. Sinha and G. J. Lastman, "Identification of continuous time multivariable systems from sampled data," *Int. J. Control,* Vol. 35, pp. 117-126, 1982.

[59] N. K. Sinha and G. P. Rao, *Identification of Continuous-Time Systems,* Kluwer Academic Publishers, Dordrecht, 1991.

[60] N. K. Sinha, "Estimation of transfer function of continuous system from sampled data," *Proc. IEE,* Vol. 119, pp. 612-614, 1972.

[61] K. Steiglitz and L. E. McBride, "A technique for identification of linear systems," *IEEE Trans. on Automatic Control,* vol. At-10, pp. 461-464, Oct., 1965.

[62] E. V. Bohn, "Methods using Walsh functions," in *Identification of Continuous-Time Systems,* (N. K. Sinha and G. P. Rao), Kluwer Academic Publishers, Dordrecht, pp. 139-157, 1991.

[63] H. Dai and N. K. Sinha, "Use of numerical integration methods," in *Identification of Continuous-Time Systems,* (N. K. Sinha and G. P. Rao), Kluwer Academic Publishers, Dordrecht, pp. 259-289, 1991.

[64] S. Sagara and Z. Y. Zhao, "Application of digital filtering techniques," in *Identification of Continuous-Time Systems,* (N. K. Sinha and G. P. Rao), Kluwer Academic Publishers, Dordrecht, pp. 291-325, 1991.

[65] D. Graupe, *Time Series Analysis, Identification and Adaptive Filtering,* Robert E. Krieger Publishing Company Inc., Malabar, FL, 1984.

[66] G. E. Box and G. M. Jenkins, *Time Series Analysis forecasting and control,* Holden-Day Inc., Oakland, CA, 1976.

[67] K. H. Loh and R. L. Geiger, "Frequency response measurements of high-frequency linear systems by least squares identification techniques," in *ECCTD* proceedings, pp. 719-728, 1991.

[68] J. Ramirez and E. Sanchez, "Active compensation of operational transconductance amplifier filters using partial positive feedback," *IEEE J. of Solid-State Circuits*, Vol. SC-25, pp. 1024-1028, August 1990.

[69] K. A. Kozma, D. A. Johns and A. S. Sedra, "An adaptive tuning circuit for integrated continuous-time filters," in *Proc. Int. Symp. CAS*, pp. 1163-1166, May 1990.

[70] G. S. Moschytz, *Linear Integrated Networks Design,* Van Nostrand Reinhold Company, New York, 1975.

[71] P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, Wiley, New York, 1984.

[72] R. Gregorian and G. C. Temes, *Analog MOS Integrated Circuits For Signal Processing*, Wiley, New York, 1986.

[73] R. Unbehauen and A. Cichocki, *MOS Switched-Capacitor and Continuous-Time Integrated Circuits and Systems*, Springer-Verlag, Berlin, 1989.

[74] J. I. Brown, "Differential Amplifiers That Reject Common-Mode Currents," *IEEE J. of Solid-State Circuits*, Vol. SC-6, No. 6, pp. 385-391, December 1971.

[75] G. Meyer-Brotz and A. Kley, "The Common-Mode Rejection of Transistor Differential Amplifiers," *IEEE Trans. on Circuits Theory*, Vol. CT-13, No. 2, pp. 171-175, June 1966.

[76] P. M. Vanpeterghem and J. F. Duque-Carrillo, "A General Description of Common-Mode Feedback in Fully-Differential Amplifiers," in *Proc. IEEE Int. Symp. CAS*, pp. 3209-3212, 1990.

[77] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 1984.

[78] M. Abramowitz and I. E. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables,* National Bureau of Standards, AMS 55, December 1972.

[79] K. R. Lakshmikumar, R. A. Hadaway, and M. A. Copeland, "Characterization and Modeling of Mismatch in MOS Transistors for Precision Analog Design," *IEEE J. of Solid-State Circuits*, Vol. SC-21, No. 6, pp. 1057-1066, December 1986.

[80] M. J. Pelgrom, A. C. Duinmaijer, and A. P. Welbers, "Matching Properties of MOS Transistors," *IEEE J. of Solid-State Circuits*, Vol. 24, No. 5, pp. 1433-1439, October 1989.

[81] M. Degrauwe, E. Vittoz, and I. Verbauwhede, "A micropower CMOS-Instrumentation Amplifier," *IEEE J. of Solid-State Circuits*, Vol. SC-20, No. 3, pp. 805-807, June 1985.

[82] P. E. Allen and D. R. Holberg, *CMOS Analog Circuit Design*, HRW, New York, 1987.

[83] E. Vittoz, *Dynamic analog techniques*, in *Design of MOS VLSI circuits for telecommunications*, edited by Y. Tsividis and P. Antognetti, Prentice-Hall, Englewood Cliffs, NJ, 1985.

[84] S. L. Wong and C. Andre T. Salama, "A switched differential op-amp with low offset and reduced 1/f noise," *IEEE Trans. on Circuits and Systems*, Vol. CAS-33, No. 11, pp. 1119-1127, Nov. 1986.

[85] C. C. Enz, E. A. Vittoz, and F. Krummenacher, "A CMOS chopper amplifier," *IEEE J. Solid State Circuits*, Vol. SC-22, No. 3, pp. 335-342, June 1987.

[86] M. Coln, "Chopper stabilization of MOS operational amplifiers using feed-forward techniques," *IEEE J. Solid State Circuits*, Vol. SC-16, No. 6, pp. 745-748, Dec. 1981.

[87] R. J. Wiegerink, E. Seevinck, and W. D. Jager, "Offset canceling circuit," *IEEE J. Solid State Circuits*, Vol. 24, No. 3, pp. 651-658, June 1989.

[88] B. J. Mccarroll, C. G. Sodini, and H. S. Lee, "A high-speed CMOS comparator for use in an ADC," *IEEE J. Solid State Circuits*, Vol. 23, No. 1, pp. 159-165, Feb. 1988.

[89] J. H. Atherton and H. T. Simmonds, "An offset reduction technique for use with CMOS integrated comparators and amplifiers," *IEEE J. Solid State Circuits*, vol. 27, no. 8, pp. 1168-1175, Aug. 1992.

[90] R. L. Geiger and P. E. Allen, "Switched resistor filters," *24th Midwest Symposium on Circuits and Systems*, pp. 87-91, June 1981.

[91] R. L. Geiger, P. E. Allen, and T. Ngo, "Switched resistor filters–A continuous-time approach to monolithic MOS filter design," *IEEE Trans. on Circuits and Systems*, Vol. 29, pp. 306-315, May 1982.

[92] A. P. Chandrakasan, S. Sheng, and R. W. Brodersen, "Low-power CMOS digital design," *IEEE J. of Solid-State Circuits*, Vol. 27, No. 4, pp. 473-484, April 1992.

[93] Y. Nakagome, H. Tanaka, K. Takeuchi, E. Kume, Y. Watanabe, T. Kaga, Y. Kawamoto, F. Murai, R. Izawa, D. Hisamoto, T. Kisu, T. Nishida, E. Takeda, and K. Itoh, "An experimental 1.5-V 64-Mb DRAM," *IEEE J. of Solid-State Circuits*, Vol. 26, No. 4, pp. 465-470, April 1991.

[94] A. Sekiyama, T. Seki, S. Nagai, A. Iwase, N. Suzuki, and M. Hayasaka, "A 1-V operating 256-kb full-CMOS SRAM," *IEEE J. of Solid-State Circuits*, Vol. 27, No. 5, pp. 776-782, May 1992.

[95] M. Nagata, "Limitations, innovations, and challenges of circuits and devices into a half micrometer and beyond," *IEEE J. of Solid-State Circuits*, Vol. 27, No. 4, pp. 465-472, April 1992.

[96] M. Aoki, T. Ishii, T. Yoshimura, Y. Kiyota, S. Iijima, T. Yamanaka, T. Kure, K. Ohyu, T. Nishida, S. Okazaki, K. Seki, and K. Shimohigashi, "0.1μm CMOS devices using low-impurity-channel transistor (LICT)," in *IEDM Tech. Dig.*, pp. 939-941, 1990.

[97] C. Kuo, J. R. Yeargain, W. J. Downey, K. A. Ilgenstein, J. R. Jorvig, S. L. Smith, and A. R. Bormann, "An 80ns 32K EEPROM using the FETMOS cell," *IEEE J. of Solid-State Circuits*, Vol. SC-17, No. 5, pp. 821-827, October 1982.

[98] M68HC11 Reference Manual, Motorola, 1991

[99] J. Sweeney and R. L. Geiger, "Very high precision analog trimming using floating gate MOSFETs," *European Conference on Circuit Theory and Design (ECCTD)*, Brighton UK, pp.652-655, September 1989.

[100] J. F. Dickson, "On-chip high-voltage generation in MNOS integrated circuits using an improved voltage multiplier technique," *IEEE J. of Solid-State Circuits*, Vol. SC-11, No. 3, pp. 374-378, June 1976.

[101] H. G. Lee, S. Y. Oh, and G. Fuller, "A simple and accurate method to measure the threshold voltage of an enhancement-mode MOSFET," *IEEE Trans. on Electron Devices*, Vol. ED-29, No. 2, pp. 346-348, Feb. 1982.

[102] S. T. Wang, "On the I-V characteristics of floating-gate MOS transistors," *IEEE Trans. on Electron Devices*, Vol. ED-26, No. 9, pp. 346-348, Sep. 1979.

[103] M. G. Johnson, "An input-free V_T extractor circuit using a two-transistor differential amplifier," *IEEE J. Solid-State Circuits*, vol. 28, pp. 704-705, June 1993.

[104] Z. Wang, "Automatic V_T extractors based on an $n \times n^2$ MOS transistor array and their application," *IEEE of Solid-State Circuits*, vol. 27, pp. 1057-1066, Sep. 1992.

[105] Y. P. Tsividis and R. W. Ulmer, "Threshold voltage generation and supply-independent biasing in CMOS integrated circuits," *Electronic Circuits and Systems,* vol. 3, pp. 1-4, Jan. 1979.

[106] R. Alini, A. Baschirotto, R. Castello, and F. Montecchi, "Accurate MOS threshold voltage detector for bias circuitry," *IEEE International Symposium on Circuits and Systems*, pp. 1280-1283, San Diego, May 1992.

[107] A. Vladimirescu and S. Liu, "The simulation of MOS integrated circuits using SPICE2," Memo UCB/ERL M80/7, Electron. Res. Lab., Univ. California, Berkeley, CA 1980.

[108] H. B. Bakoglu, *Circuits, interconnections, and packaging for VLSI*, Addison-Wesley, Reading, MA, 1990.

[109] R. Gregorian and W. Nicholson, "CMOS switched-capacitor filters for a PCM voice CODEC," *IEEE J. of Solid-State Circuits*, SC-14 pp. 970-980, Dec., 1979

[110] K. Martin and S. Sedra, "Strays-insensitive switched-capacitors filters based on bilinear z-transform," *Electron Lett.*, vol. 15, pp. 365-366, June 1979.

[111] W. H. Ki and G. C. Temes, "Offset-compensated switched-capacitor integrators," *IEEE International Symposium on Circuits and Systems*, pp. 2829-2831, 1990.

[112] G. C. Temes and K. Haug, "Improved offset-compensated schemes for switched-capacitor circuits," *Electron Lett.*, vol. 20, pp. 508-509, June. 1984.

[113] C. C. Lee, "A new switched-capacitor realization for cyclic analog-to-digital converter," *IEEE International Symposium on Circuits and Systems*, pp. 1261-1265, 1983.

[114] P. W. Li, M. J. Chin, P. R. Gray, and R. Castello, "A ratio-independent algorithmic analog-to-digital conversion technique," *IEEE J. Solid-State Circuits*, vol. SC-19, no. 6, pp. 828-836, Dec. 1984.

[115] S. Y. Chin and C. Y. Wu, "A ratio-independent and gain-insensitive algorithmic analog-to-digital converter," *IEEE International Symposium on Circuits and Systems*, pp. 1200-1203, 1993.

[116] H. Onodera, T. Tateishi, and K. Tamaru, "A cyclic A/D converter that does not require ratio-matched components," *IEEE J. Solid-State Circuits*, vol. 23, no. 1, pp. 152-158, Feb. 1988.

[117] K. Nagaraj, T. R. Viswanathan, K. Singhal, and J. Vlach, "Switched-capacitor circuits with reduced sensitivity to amplifier gain," *IEEE Trans. on Circuits and Systems*, vol. CAS-34, pp. 571-574, May 1987.

[118] K. Nagaraj, K. Singhal, T. R. Viswanathan, and J. Vlach, "Reduction of the finite gain effect in switched-capacitor filters," *Electron Lett.*, vol. 21, pp. 644-645, July. 1985.

[119] L. E. Larson and G. C. Temes, "Switched-capacitor building blocks with reduced sensitivity to finite amplifier gain, bandwidth, and offset voltage," *IEEE International Symposium on Circuits and Systems*, pp. 334-338, 1987.

[120] K. Haug, F. Maloberti, and G. C. Temes, "Switched-capacitor integrators with low finite gain sensitivity," *Electron Lett.*, vol. 21, pp. 1156-1157, Nov. 1985.

[121] G. Fischer and G. S. Moschytz, "SC filters for high frequencies with compensation for finite gain effect," *IEEE Trans. on Circuits and Systems,* vol. CAS-32 pp. 1050-1056, Oct. 1985.

[122] K. Suyama and S. C. Fang, "Users' manual for SWITCAP2 version 1.1," Dept. of Electrical Eng. and Center for Telecommunications Research, Columbia University, Feb. 1992.

[123] C. Eichenberger and W. Guggenbuhl, "On charge injection in analog MOS switches and dummy switch compensation techniques," *IEEE Transactions on Circuits and Systems*, vol. 37, pp. 256-264, Feb. 1990.

[124] G. Wegmann, E. A. Vittoz, and F. Rahali, "Charge injection in analog MOS switches," *IEEE J. of Solid-state Circuits,* vol. SC-22, no. 6, pp. 1091-1097, Dec. 1987.

[125] J. H. Shieh, M. Patil, and B. J. Sheu, "Measurement and analysis of charge injection in MOS analog switches," *IEEE J. of Solid-state Circuits,* vol. SC-22, no. 2, pp. 277-281, April 1987.

[126] B. J. Sheu and C. Hu, "Switch-induced error voltage on a switched capacitor," *IEEE J. of Solid-state Circuits,* vol. SC-19, no. 4, pp. 519-525, Aug. 1984.

[127] W. B. Wilson, H. Z. Massoud, E. J. Swanson, R. T. George, and R. B. Fair, "Measurement and modeling of charge feedthrough in n-channel MOS analog switches," *IEEE J. of Solid-state Circuits,* vol. SC-20, no. 6, pp. 1206-1213, Dec. 1985.

[128] J. B. Kuo, R. W. Dutton, and B. A. Wooley, "Turn-off transients in circular geometry MOS pass transistors," *IEEE J. of Solid-state Circuits,* vol. SC-21, no. 5, pp. 837-844, Oct. 1986.

[129] D. Macquigg, "Residual charge on a switched capacitor," *IEEE J. of Solid-state Circuits,* vol. SC-18, no. 6, pp. 811-813, Dec. 1983.

[130] J. Brugler and P. Jespers, "Charge pumping in MOS devices," *IEEE Transactions on Electron Devices,* vol. ED-16, pp. 297-302, March 1969.

[131] R. C. Yen and P. R. Gray, "A MOS switched-capacitor instrumentation amplifier," *IEEE J. of Solid-state Circuits,* vol. SC-17, no. 6, pp. 1008-1013, Dec. 1982.

[132] R. E. Suarez, P. R. Gray, and D. A. Hodges, "All-MOS charge redistribution analog-to-digital conversion techniques–Part II," *IEEE J. of Solid-state Circuits,* vol. SC-10, no. 6, pp. 379-385, Dec. 1975.

[133] R. Poujois and J. Borel, "A low drift fully integrated MOSFET operational amplifier," *IEEE J. of Solid-state Circuits,* vol. SC-13, Aug. 1978.

[134] L. Bienstman and H. J. deMan, "An eight-channel 8bit micro-processor compatible NMOS converter with programmable scaling," *IEEE J. of Solid-state Circuits,* vol. SC-15, pp. 1051-1059, Dec. 1980.

[135] S. Ogawa and K. Watanabe, "Clock-feedthrough compensated switched-capacitor circuits," *International Symposium on Circuits and Systems,* pp. 1195-1198, 1992.

[136] K. Watanabe and S. Ogawa, "Clock-feedthrough compensated sample/hold circuits," *Electron. Lett.,* vol. 24, pp.1226-1227, 1988.

[137] D. E. Ward and R. W. Dutton, "A charge oriented model for MOS transistor capacitances," *IEEE J. of Solid-state Circuits,* vol. SC-13, pp. 703-708, October 1978.

[138] G. Wegmann and E. A. Vittoz, "Analysis and improvements of accurate dynamic current mirrors," *IEEE J. of Solid-State Circuits,* vol. 25, no. 3, pp. 699-706, June 1990.

[139] G. Wegmann and E. A. Vittoz, "Basic principles of accurate dynamic current mirrors," *IEE Proceedings,* vol. 137, pt. G, no. 2, pp. 95-100, April 1990.

[140] G. Wegmann and E. A. Vittoz, "Very accurate dynamic current mirrors," *Electron. Lett.,* vol. 25, pp. 644-646, May 1989.

[141] W. Groeneveld, H. Schouwenaars, and H. Termeer, "A self calibration technique for monolithic high-resolution D/A converters," *ISSCC Digest of Technical Papers,* pp. 22-23, Feb. 1989.

[142] D. G. Nairn and C. A. T. Salama, "Current mode analog-to-digital converters," *IEEE International Symposium on Circuits and Systems,* pp. 1588-1591, 1989.

[143] S. J. Daubert, D. Vallancourt, and Y. P. Tsividis, "Current copier cells," *Electron. Lett.*, vol. 24, no. 25, pp. 1560-1562, Dec. 1988.

[144] D. Vallancourt, and Y. P. Tsividis, "Sampled-current circuits," *IEEE International Symposium on Circuits and Systems*, pp. 1592-1595, 1989.

[145] T. S. Fiez, G. Liang, and D. J. Allstot, "Switched-current circuit design issues," *IEEE J. of Solid-State Circuits*, vol. 26, no. 3, pp. 192-202, March 1991.

[146] T. S. Fiez and D. J. Allstot, "CMOS switched-current ladder filters," *IEEE J. of Solid-State Circuits*, vol. 25, no. 6, pp. 1360-1367, Dec. 1990.

[147] J. B. Hughes, N. C. Bird, and I. C. Macbeth, "Switched currents - A new technique for analog sampled-data signal processing," *IEEE International Symposium on Circuits and Systems*, pp. 1584-1587, 1989.