RATIO-INDEPENDENT ALGORITHMIC ANALOG TO DIGITAL CONVERSION TECHNIQUES

by

P-W. Li

Memorandum No. UCB/ERL M84/66

20 August 1984

RATIO-INDEPENDENT ALGORITHMIC ANALOG TO DIGITAL CONVERSION TECHNIQUES

by

P-W. Li

Memorandum No. UCB/ERL M84/66

20 August 1984

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Research sponsored by the National Science Foundation Grant ECS-8310442, and by the Intel Corporation.

RATIO-INDEPENDENT ALGORITHMIC ANALOG TO DIGITAL CONVERSION TECHNIQUES

Ph. D. Ping-Wai Li Department of EECS

Chairman of Committee

Abstract

This thesis reports on an investigation on algorithmic analog to digital conversion techniques in achieving high resolution and linearity A/D converters, implemented using metal-oxide-semiconductor large scale integration. With this technique, high resolution converters can be realised within a small area.

A simple ratio-independent switching algorithm has been devised to overcome the gain error of MOS gain blocks that employs capacitor for the gain-setting. Using a delayed timing control scheme and a differential circuit implementation to cancel the charge injection from MOS transistors, high monotonicity and linearity can be achieved without the need for capacitor matching.

As a test vehicle an experimental integrated circuit has been fabricated using a 5-micron CMOS process. The converter showed monotonicity of 12 bits operating at a sampling frequency up to 8kHz. The integral linearity error was 3.2lsb. Total chip area for the analog circuit part of the converter was less than $2400 mil^2$.

TABLE OF CONTENTS

RATIO-INDEPENDENT ALGORITHMIC ANALOG TO DIGITAL CONVERSION TECHNIQUES

by

P-W. Li

Memorandum No. UCB/ERL M84/66

20 August 1984

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Research sponsored by the National Science Foundation Grant ECS-8310442, and by the Intel Corporation.

RATIO-INDEPENDENT ALGORITHMIC ANALOG TO DIGITAL CONVERSION TECHNIQUES

Ph. D. Ping-Wai Li Department of EECS

Chairman of Committee

Abstract

This thesis reports on an investigation on algorithmic analog to digital conversion techniques in achieving high resolution and linearity A/D converters, implemented using metal-oxide-semiconductor large scale integration. With this technique, high resolution converters can be realised within a small area.

A simple ratio-independent switching algorithm has been devised to overcome the gain error of MOS gain blocks that employs capacitor for the gain-setting. Using a delayed timing control scheme and a differential circuit implementation to cancel the charge injection from MOS transistors, high monotonicity and linearity can be achieved without the need for capacitor matching.

As a test vehicle an experimental integrated circuit has been fabricated using a 5-micron CMOS process. The converter showed monotonicity of 12 bits operating at a sampling frequency up to 8kHz. The integral linearity error was 3.2lsb. Total chip area for the analog circuit part of the converter was less than $2400 mil^2$.

TABLE OF CONTENTS

CHAPTER 0

Introduction

The rapid development of communication and data acquisition systems has demanded the integration of more analog signal processing circuitry on the same silicon chip. As a result, more area-efficient realization of given circuit functions have to be developed. Recent advances in analog MOS integrated circuits have fundamentally influenced the design approach for these circuits. Whereas resistor and transistor ratios are employed in the analog bipolar technology to define the accuracy of a circuit function, their functions are substituted by the MOS capacitor in analog MOS circuit. Not only does the MOS capacitor give better matching within the same area, it also possesses better stability and less voltage dependence than the monolithic resistor. This makes it very suitable for the integration of precision analog function on the silicon chip. One example of this new development is the switched capacitor filter. Instead of relying on accurate RC time constants, which is generally not available in the integrated circuit technology to define the frequency characteristics of the filter, it uses an accurately defined clock frequency and capacitors to realize precise time constants. The external clock frequency can be very accurate if it is derived from a crystal. With proper precaution, the capacitors can be matched to within 0.1%. The frequency response of switched capacitor filters are therefore relatively precise and they have found wide acceptance in voice band and even higher frequency communication systems. Another application of the ratioed capacitor in the data acquisition system is the charge redistribution technique which uses an array of ratioed capacitors to affect an analog to digital conversion. By

switching the binary ratioed capacitors between ground and the reference voltage, charges are subtracted or added to the signal charge and following a successive approximation procedure the digital representation of the signal can be found. This type of converter is capable of achieving 10 bit resolution and linearity with a conventional MOS technology.

Because of the matching requirement in the capacitors, they require comparably large geometrical areas. To this there are three disadvantages. Firstly the corresponding area of the circuit increases. Secondly the large capacitance slows down the circuit due to effects such as the RC discharge time constant or the settling time of MOS operational amplifiers. Finally and most importantly, the accuracy of the circuit is seriously limited by the matching accuracy of capacitors, which due to lithographical and etching error in a modern MOS process, often lies in the 8-9 bit region. This dissertation is aimed at the realization of precise analog functions using unratioed capacitors. Efforts have been concentrated in developing techniques which allow exact integral gain values to be obtained from MOS gain blocks without the used of matched capacitors. Other effects that have been investigated are related to circuit errors arising from charge injection from MOS transistors. As a verification of the techniques developed, a high resolution algorithmic analog to digital converter has been implemented and fabricated in CMOS technology. Here a ratio-independent switching algorithm is utilized to correct for the capacitor mismatch error in setting exact integral gain values with MOS gain blocks. Also it is shown how a combination of control signal timing and differential circuit can be used beneficially to cancel the charge injection from MOS transistor. A CMOS operational amplifier with a balanced common mode feedback stage ensures fast settling time and therefore moderately high speed conversion rate of the converter.

This dissertation is organized as follows. Chapter 1 describes in detail the principle and operation of the algorithmic A/D converter. The algorithmic conversion technique is analyzed in context to the successive approximation conversion technique. Following this, the various limitations or sensitivity problems of the converter are studied based on an earlier implementation.

In chapter 2, a practical implementation of the ratio-independent multiplication technique is given. This technique is essential to the operation of the algorithmic converter since it provides gain accuracy to the converter gain blocks previously unattainable by ratioed passive elements. However it is then shown that nonidealities still exist which perturbs the accuracy of the circuit. Amongst them the most critical is the charge injection effect of MOS transistor switches. To alleviate this effect, the differential circuit configuration and a delayed timing control scheme are developed. They allow a first order cancellation of the charge injection effect.

Chapter 3 depicts the design of the high speed high gain CMOS operational amplifier used for the converter. The advantages of using a single gain stage architecture are explained. An improved common mode feedback circuit featuring two balanced MOS differential pairs is used.

In chapter 4 the full implementation of the converter is given together with a detailed description of the individual stages involved. For experimental purpose, the analog circuit is controlled externally by a microprogrammable sequencer. But an on-chip ROM control scheme is proposed which uses minimum area. The noise property of the converter is also briefly discussed.

Chapter 5 gives some experimental results on a test circuit used to demonstrate the design concepts. It is an algorithmic A/D converter fabricated with a 5μm CMOS technology. It achieves 12 bit resolution and 10 bit linearity at a conversion rate of 8 kHz.

CHAPTER 1

Algorithmic Analog To Digital Conversion Fundamentals

1.1. Introduction

In this chapter the concept of the algorithmic A/D converter is reviewed. The successive approximation technique for analog to digital conversion is introduced and compared to the algorithmic method. An example of an implementation of the algorithmic A/D converter in a metal gate CMOS technology will be given. With this implementation example, some of the pertinent variables in the converter are studied and effects of their nonidealities on the characteristics the converter investigated. Before we embark upon the task of describing the algorithmic A/D conversion method, we shall first explain the operation of the successive approximation method for A/D conversion. It will become apparent that the algorithmic A/D conversion method is nothing more than a successive approximation in disguise.

1.2. Successive Approximation Method for A/D Conversion

The successive approximation method of analog to digital conversion is widely used for high resolution and medium to high speed conversion[1]. The converter consists of a comparator, a digital to analog converter (DAC), a voltage reference, a successive approximation register (SAR) and control logic as shown in Fig 1.1. The conversion is achieved by comparing successively the input signal V_{in} with the sum of binary fractions of the reference generated by the DAC. The conversion usually starts by setting the most

significant bit (MSB) of the DAC and comparing the output with V_{in}. If this output is larger than V_{in} then the most significant bit of the input is 1, otherwise the MSB of the DAC is returned to 0 and the corresponding MSB of the input is 0 as well. This process is continued for each of the bits sequentially, keeping all the previously set or unset bits, until the desired accuracy has been reached. The bit pattern of the DAC then gives the digital representation of the input signal. Thus to decide n bits of digital data, n basic cycles are required. The basic cycle consists of a settling time for the DAC, the decision time for the comparator and the delay time through the logic. Mathematically, the procedure can be described as follows. Let the digital representation of an analog signal V_{in} be $d_0 d_1, \cdots, d_{n-1}$ where $d_i \in \{0,1\}$. The analog and digital representations are related by the equation:

$$V_{in} = \left(\frac{d_0}{2} + \frac{d_1}{2^2} + \ldots + \frac{d_{n-1}}{2^n}\right) V_{ref} + \epsilon_n \tag{1.1}$$

where

V_{in} : input voltage

$d_0 d_1, \cdots, d_{n-1}$: digital representation of the signal ; assumes the value 0 or 1

V_{ref} : reference voltage

ϵ_n : quantization error

This equation expresses quantatively how the analog input signal can be represented with respect to the reference voltage by a string of binary digits. The last term ϵ_n, commonly known as the quantization error , gives the error between the digital representation of the signal and the actual value of the signal. Its magnitude is

$$-\frac{V_{ref}}{2^{n+1}} < \epsilon_n < \frac{V_{ref}}{2^{n+1}}$$

so that the error of the digital approximation is always bounded.

To derive the MSB, we subtract the quantity $\frac{V_{ref}}{2}$ from both sides of equation (1.1).

$$V_{in} - \frac{V_{ref}}{2} = (\frac{(d_0 - 1)}{2} + \frac{d_1}{2^2} + \cdots + \frac{d_{n-1}}{2^n})V_{ref} + \epsilon_n \tag{1.2}$$

Now it is true that

$$\frac{1}{2} > \frac{1}{2^2} + \frac{1}{2^3} + \cdots + \frac{1}{2^n} \tag{1.3}$$

Therefore if d_0 is 0, the right hand side of equation (1.2) would become negative. If we apply this argument to all the bits the digital representation can be easily obtained.

An example of an actual conversion is shown in the following figure. It depicts the output of the DAC as a function of time during a conversion. Take for example a signal of magnitude equal to 1/10 of the reference and it is required to convert to 5 bit accuracy. In the first cycle, the most significant bit of the DAC is set and the resultant output voltage $V_{o1} = \frac{V_{ref}}{2}$ is compared to the signal. Since it is larger than the input signal, this bit is not kept in the DAC. In the next cycle, the next MSB is set, changing the output voltage of the DAC to $\frac{V_{ref}}{4}$. This process is continued until the 4th bit is set and tested to be smaller than the input voltage. It is therefore kept and the 5th bit set and also kept in the last cycle. Thus all 5 bits have been obtained and the result is equal to 00011.

1.3. Algorithmic Analog to Digital Conversion Technique

The algorithmic, also known as the cyclic or recirculating analog to digital conversion technique has been realized in various forms since the 1960's. It was first constructed

in a partially integrated form by Tom Hornak in 1975, using an external transformer to realize the multiplication of two[2]. Then McCharles *et al* integrated the analog portion of the converter using a metal gate CMOS technology[3] in 1978. Using matched capacitors, 8-9 bit of accuracy could be achieved. The first high precision converter using this technique was done by Harris Semiconductor where precise gain values were set using trimming technique on thin film resistors[4].

A block diagram of the implementation of the algorithmic analog-to-digital converter is shown in Fig. 1.3. It consists of an analog loop which contains:

1) sample/hold block
2) multiply-by-two block
3) reference subtraction circuit
4) comparator

There are also two switches S_1 and S_2. S_1 switches the input of the sample/hold block between the input path and the loop path. S_2 controls (symbolically) the selection between reference and ground for the subtraction circuitry. The signal to be converted is first introduced into the loop by means of S_1 and the sample/hold. From here, it would be cycled through the analog loop as often as the number of bits are required from the conversion.

From this implementation, the following characteristics of the algorithmic A/D converter can be identified:

1) the associated hardware is relatively simple. Instead of having to derive the n exactly ratioed references, only two precision elements are needed, *i.e.* both the sample/hold block and the multiply-by-two blocks must have exact gains of 1 and 2 respectively.

2) The converter possesses inherently the sample/hold function.

3) It possesses the so called floating point capability. The input signal can be amplified by a factor of 2^m before the A/D conversion begins with m being an integer. This is done by cycling the signal in the analog loop, multiplying it repeatedly without doing an A/D conversion. This property makes the algorithmic A/D converter valuable *e.g.* in multiplexed operations where signal with different dynamic ranges must be converted into digital data.

1.3.1. Description of the Conversion Technique

The conversion technique will be described using the restoring algorithm. At the start of the conversion, the input signal is sampled onto the sample/hold amplifier through the switch S_{in}. This signal is then passed onto the precision gain block where it is multiplied by the exact factor of 2. To extract the digital information from this analog signal, the loop signal, denoted by V_x in Fig. 1.3, is compared to the reference. If it is larger, then the first bit is set to an 1 and the reference is subtracted off from this signal. Otherwise the first bit is set to an 0 and the loop signal is passed on unchanged for the second bit conversion. Thus the A/D conversion process is continued in this manner until the desired number of bits is obtained. It can be seen that the digital information comes out in a serial pattern, starting with the most significant bit.

To see that this process indeed performs the analog-to-digital conversion that is wanted, consider the same equation:

$$V_{in} = \left(\frac{d_0}{2} + \frac{d_1}{2^2} + \cdots + \frac{d_{n-1}}{2^n - 1} \right) V_{ref} + \epsilon_n$$

In a successive approximation determination of the digital information, the signal is compared successively to different terms on the right hand side of equation (1.1) beginning with the first or numerically largest term, *i.e.* $d_0\frac{V_{ref}}{2}$. If the result is positive, the difference is kept for the next comparison. Otherwise, the original signal is kept for the next comparison. By repeating this process, the bits are determined individually until the desired resolution is reached.

By rearranging equation (1.1) in another form, a description of the algorithmic A/D conversion technique can be derived. This is done by multiplying equation (1.1) on both side by 2 and moving the term d_0 to the left side of the equation.

$$2(V_{in} - \frac{V_{ref}}{2}) = ((d_0 - 1) + \frac{d_1}{2} + \cdots + \frac{d_{n-1}}{2^{n-2}})V_{ref} + 2\epsilon_n \qquad (1.4)$$

the first bit can be determined if the sign of left hand side of equation can be tested. If it is positive, then d_o is 1 and the equation is kept in the same form. Otherwise it is 0, in which case the subtraction will be annulled. In the same equation the quantization error or equivalently the comparator window error is also doubled. This has the effect that as the conversion proceeds, comparator error has less and less effect on the characteristics of the conversion. Similarly for the second bit,

$$2(2(V_{in} - d_0\frac{V_{ref}}{2}) - \frac{V_{ref}}{2}) = ((d_1 - 1) + \frac{d_2}{2} + \cdots + \frac{d_{n-1}}{2^{n-3}})V_{ref} + 2^2\epsilon_n \qquad (1.5)$$

The bits are extracted following this scheme of multiplication by two and compared until d_{n-1} is obtained through the equation:

$$2(2(\ldots 2(V_{in} - d_0\frac{V_{ref}}{2}) - d_1\frac{V_{ref}}{2})\ldots) - d_{n-2}\frac{V_{ref}}{2}) - d_{n-1}V\frac{ref}{2}) = 2^n\epsilon_n \qquad (1.6)$$

Therefore the algorithmic A/D conversion technique requires 1 comparison per bit of

conversion, completely analogous to the successive approximation method. Also it is less sensitive to comparator error during the conversion. However there is an additional time delay involved in moving the signal through the analog loop.

An example of the waveform of the multiply-by-two amplifier as a function of time is shown in Fig 1.4. In the first cycle, the input signal is multiplied by 2 and the reference voltage is subtracted off. If the result is positive, then this loop voltage is kept for the next conversion sequence. The conversion continues until the bit pattern is again obtained.

1.4. An implementation example

Fig. 1.5 illustrates an implementation of the algorithmic A/D converter with a metal gate CMOS technology[5]. The sample/hold and the multiply-by-two blocks are realized using MOS operational amplifiers and matched capacitors as precision feedback network. The reference subtraction circuit has been incorporated into the multiply-by-two amplifier by an extra capacitor C_5 attached to the summing node of the operational amplifier. NMOS transistor switches $T_1 - T_6$ form the interconnection paths closing the analog loop. They are be controlled by simple logic. With this circuit configuration, the signal is compared to the reference by first forming their difference and comparing to zero or ground. One of the advantages is that either the restoring or the non-restoring algorithms can be applied. Another is that it simplifies the design of the comparator because usually the reference is located very near to the voltage supply and if the signal is compared directly to the reference, the comparator must have a large input common mode voltage range. In this case, the large common mode voltage swing is avoided.

1.4.1. Mode of operation

Assuming that the operational amplifiers are ideal, the input signal is introduced into the loop by turning on switch T_1 and T_2. T_2 puts operational amplifier 1 into a unity feedback mode which establishes a quiescent voltage on the negative summing node equal to ground. At the same time, the input signal voltage is stored across C_1. Next, $T_1 T_2$ are turned off, isolating the summing node of operational amplifier 1. Simultaneously $T_6 T_3$ are turned on. This nulls out operational amplifier 2. At the same time, one plate of capacitor C_1 is being connected to ground. Thus the signal is transferred to the output of operational amplifier 1 which charges up capacitor C_3. Now instead of first multiplying the signal by two and then compare it to the reference voltage, the reference voltage is subtracted simultaneously during the multiplication process. The comparator would then decide whether the original signal should be restored or the difference signal $2V_{in} - V_{ref}$ be kept. This sequence of operations can be accomplished by turning on T_5, which displaces a charge through C_4 representing the reference voltage. The comparator now makes the decision. If the signal is to be restored, the plate of C_5 would be returned to ground from V_{ref}. Otherwise, the capacitor plate is kept at V_{ref}. In either case, the loop signal have been stored on C_1 and by simply nulling operational amplifier 2 again, this signal can then be holded in operational amplifier 1 for the next conversion cycle. The rest of the conversion can be carried out in a similar fashion. Each bit conversion requires 2 or 3 operational amplifier settling times τ. Let τ determine the clock cycle, the worst case conversion time for n bits is $3n\ \tau$.

1.5. Sources of Error in Algorithmic A/D converter

In the previous discussion and the implementation example, it has been assumed that the building blocks of the algorithmic converter were ideal. In practice, nonidealities occur which seriously affect the transfer characteristics of the converter. The nature of these nonidealities depends on the technique and the technology used to implement the converter. In the following section, a generalized method for quantifying the analog loop property shall be given which is similar to the analysis of the distortion in a nonlinear circuit by means of power series[6]. An understanding of this method gives some insight into the operation of the algorithmic converter and can provide some correlation between the error sources and experimental results obtained by the code density test[7].

1.5.1. Quantizing The Analog Loop

The quantization method consists in regrouping the building blocks into another set of building blocks as shown in Fig. 1.6 . The block designated by z^{-1} is an analog delay block which has a unity gain and provides one clock cycle of delay to the analog signal. This is necessary because a direct connection between the input and output of an analog gain block in this case is not possible. The other block describes the analog loop transfer function $T(V_I)$. It describes the combined transfer characteristics of the multiply-by-two and the sample/hold amplifier. It is defined as

$$T(V_I) = \frac{V_o}{V_I} \tag{1.7}$$

and it can be expanded in a Taylor series

$$T(V_I) = \frac{e_o}{V_I} + (2 - e_1) + e_2 V_I + e_3 V_I^2 + \cdots \tag{1.8}$$

where e_i is the ith order error in the loop transfer function. The effect of noise has not

been included in this loop transfer function. It can be described by the addition of a stochastic function to the series. By using this function together with a conversion algorithm, various error characteristics dependent on the coefficients $e_0, e_1, \cdots, e_n$ can be obtained.

1.5.2. Sources of Error in an MOS Implementation

In an MOS technology, probably the easiest way to implement an algorithmic A/D converter is to use MOS operational amplifiers and capacitors. An example of an implementation was given earlier. With this example, the limitation towards achieving high accuracy A/D conversion can be studied. The loop transfer function equation gives a framework with which the various error terms can be identified.

1.5.2.1. Zeroth Order Error

The zeroth order error in the loop transfer function is identical to the offset in a circuit. Offset voltage in the analog loop can be generated by various mechanisms. Referring to Fig. 1.5, it is assumed that the operational amplifiers have no input voltage offset. However MOS operational amplifiers usually have input offset voltages ranging from 2 to over 10 mV depending on the design and variations in process. These offset voltages are added to the loop signal as it passes around the loop, resulting in a zeroth order error in the loop transfer equation. It is therefore referred to as the loop offset error.

In addition to this contribution to the offset term, charge injection effect from MOS transistor switches also contribute to an offset voltage. To obtain a magnitude for this error, consider a minimum size transistor connected to a capacitor of 1pF as shown in Fig. 1.7. The transistor has dimensions of l=4μm, w=4μm and the threshold voltage is V_t =0. The gate oxide has a thickness of 50nm. If the transistor is turned off then the channel

charge at a gate bias voltage of 5V is injected into a capacitor of 1pF, the voltage change on the capacitor is

$$\delta V = \frac{Q_{ch}}{C} = C_{ox} \frac{V_g}{C} = 59mV \tag{1.9}$$

which is a significant error term for precision analog circuit operation. This injection voltage error appears as an offset term to the converter.

1.5.2.2. First Order Error

The first order error contains three components:

1) capacitor mismatch

2) charge injection error

3) capacitor voltage coefficient

The capacitor matching in an MOS technology is limited by various process steps and typically attains an accuracy of 9-10 bits. Therefore if the gain of the sample/hold and multiply-by-two blocks are set by capacitor ratios, they cannot be more accurate than 0.1% which means that the worst case gain error for the analog loop is 0.2%. This introduces an first order error term into the loop transfer function.

Similarly, insufficient gain in the operational amplifier can cause a gain accuracy problem. A rough estimate of the open looped gain of the operational amplifier required to achieve a certain gain accuracy can be made with the equation

$$\delta \frac{A}{A} = \frac{\delta \frac{a}{a}}{1 + T} \tag{1.10}$$

where

δA : change in closed loop gain
A : closed loop gain
δa : change in open loop gain
a : open loop gain
T : loop gain

For example, a closed loop gain accuracy to 13 bit requires a open loop gain of at least 20000 or 86dB.

In addition to the capacitor ratio problem, it can be shown in chapter 3 that charge injection from MOS switches can alter the gain of a MOS gain circuit. This is caused by charge injection effect which is proportional to the magnitude of the input signal.

Finally MOS capacitors are voltage dependent. When they are used in MOS gain stages, the voltage gain becomes a nonlinear function of the input. MOS capacitor are usually formed between two heavily doped polysilicon layers or between a heavily doped polysilicon layer and a diffusion layer in the substrate. The formation of accumulation or depletion layers in the silicon or polysilicon layers whose widths are dependent on the voltage introduces a voltage dependent capacitance effect. This effect contributes a significant amount to the first order error term.

1.5.2.3. Higher Order Terms

Higher order terms in the analog loop equation are usually due to the nonlinear voltage dependence of the MOS capacitor and perhaps the nonlinear input-output transfer characteristics of the operational amplifier. Their effect is usually insignificant compared to the first two terms in the equation except *e.g.* when the operational amplifier has very low gain. The small loop gain would fail to suppress the highly nonlinear output characteristics of the operational amplifier.

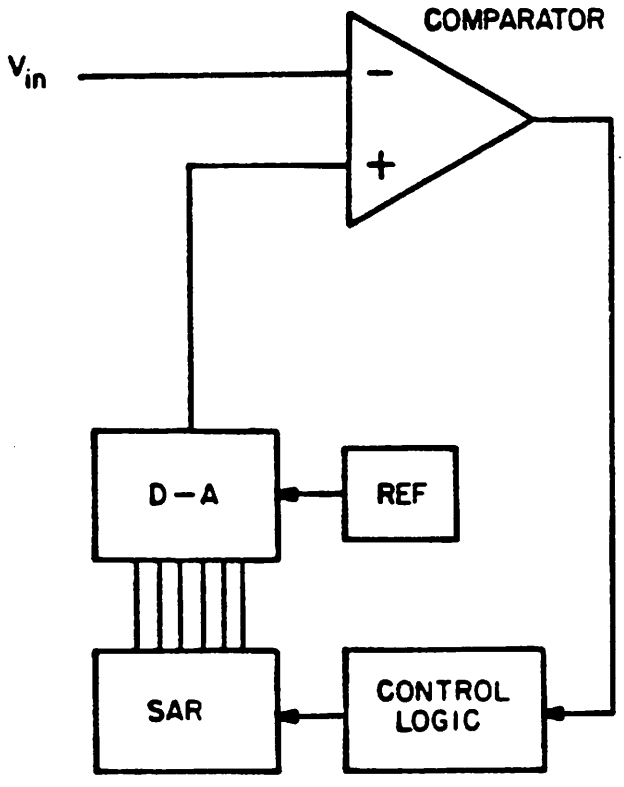


Figure 1.1

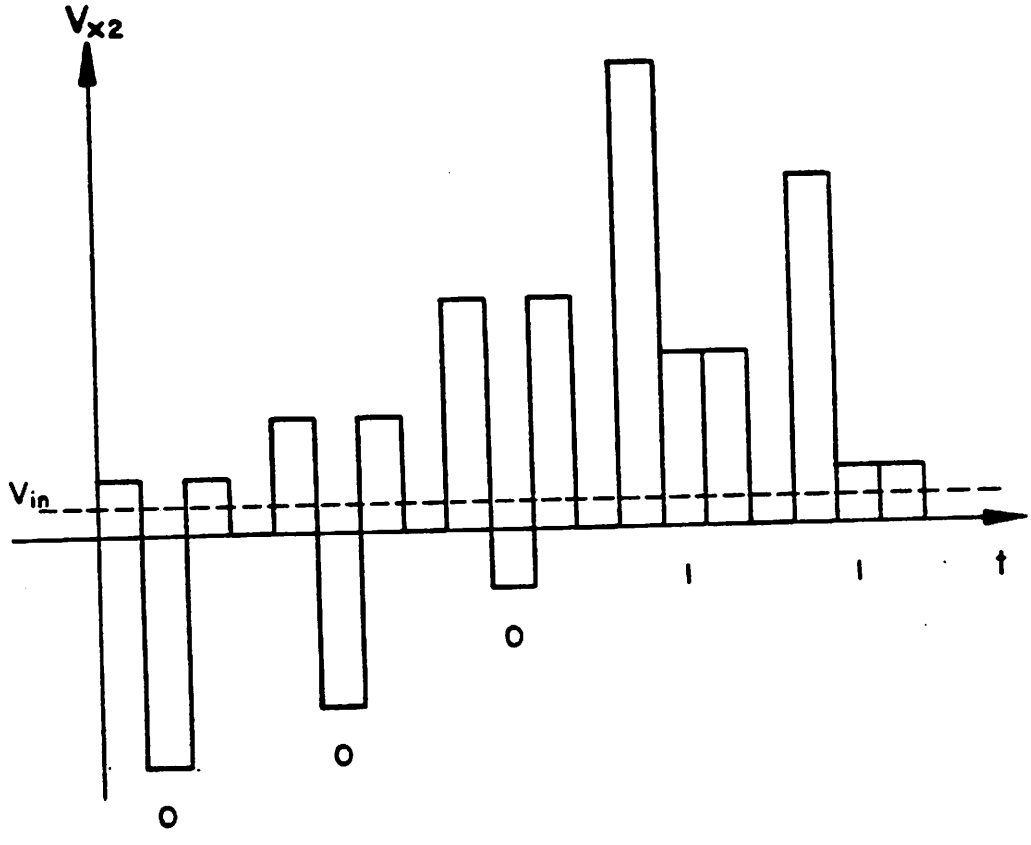


Figure 1.2

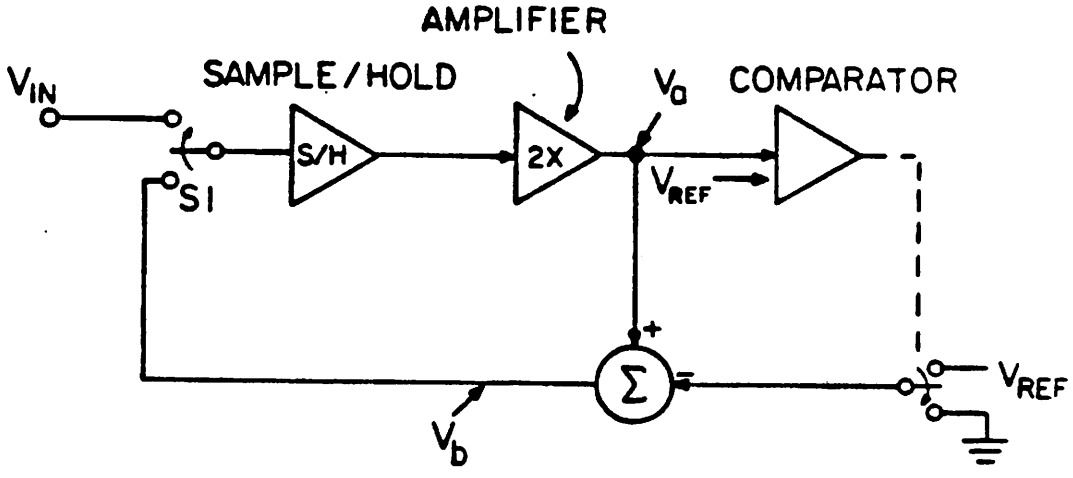

AMPLIFIER
SAMPLE / HOLD
COMPARATOR
V_{IN}
S/H
2X
V_o
V_{REF}
S1
+
Σ
−
V_{REF}
V_b

Figure 1.3

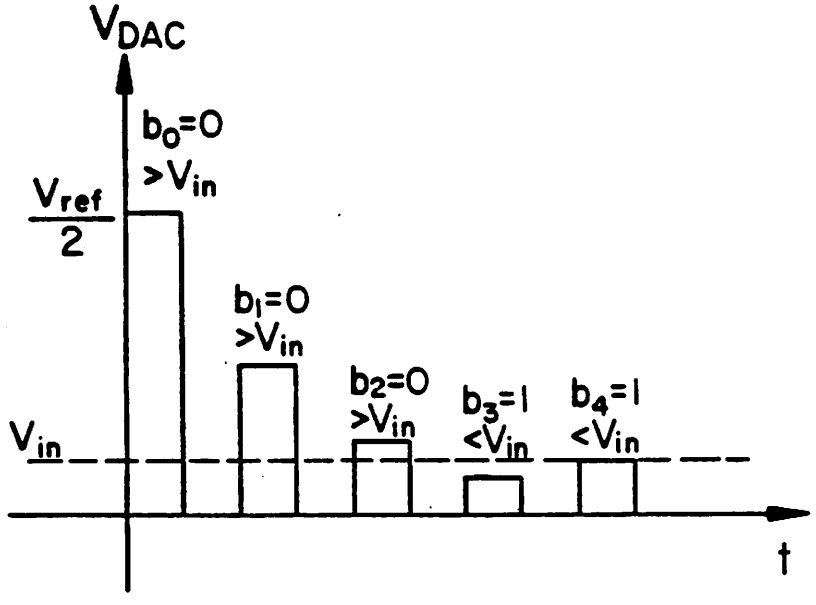
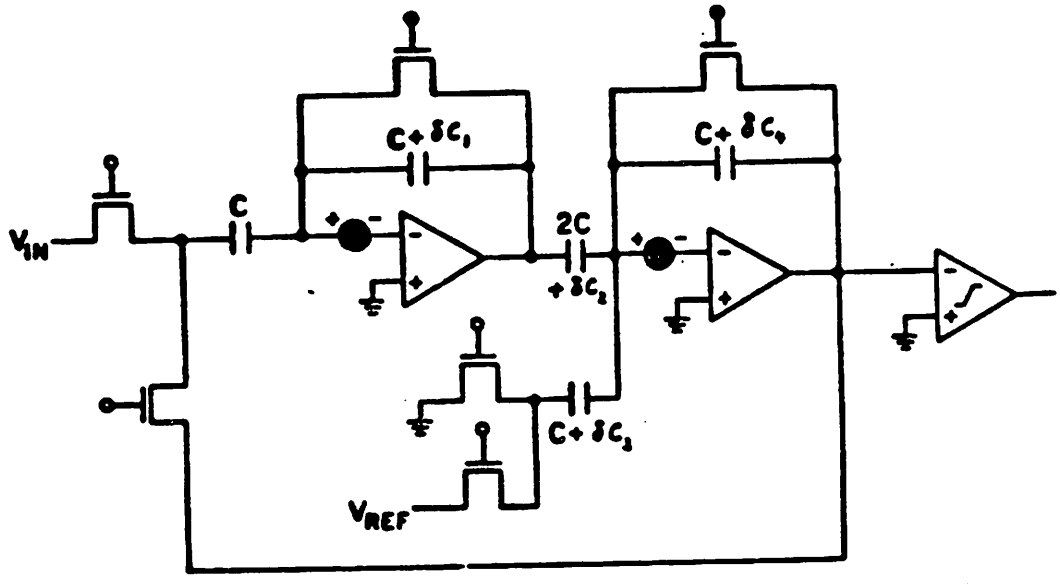


Figure 1.4



- OP AMP OFFSET ~10 mV
- CHARGE INJECTION OFFSET ~10 mV
- CAPACITOR MISMATCH ~8-9 BIT ACCURACY

Figure 1.5

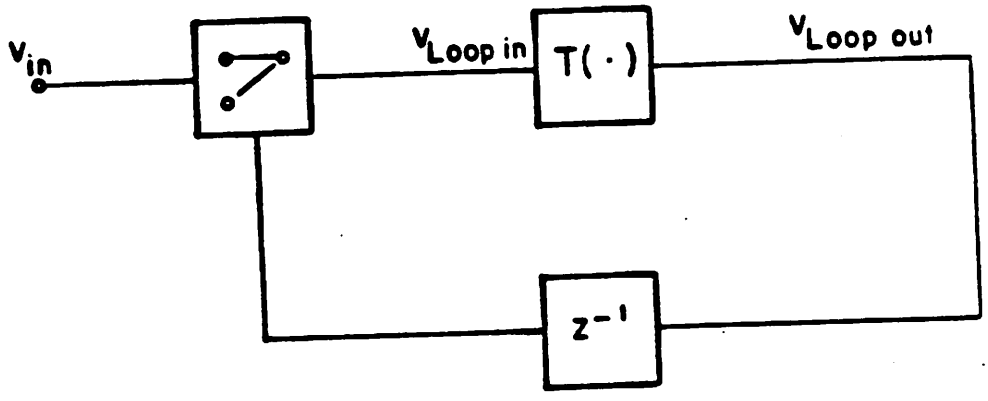


Figure 1.6

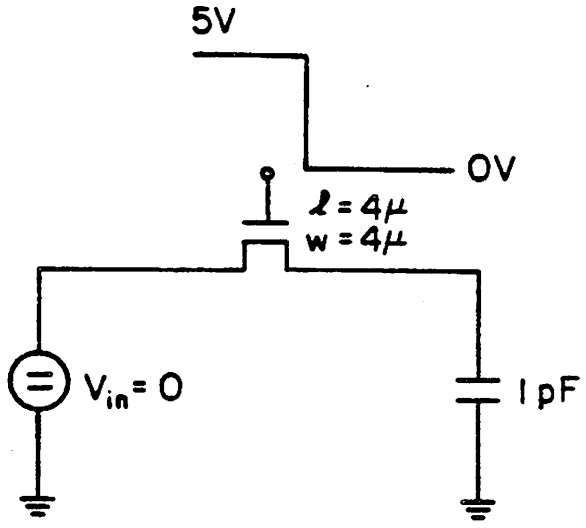


Figure 1.7

CHAPTER 2

Ratio-Independent Circuits

2.1. Introduction

One of the prevalent method for obtaining well defined and stable gain is to apply negative feedback around an amplifier using a passive network. If the gain of the amplifier is very large, the overall gain of the circuit is equal to the reciprocal of the passive network transfer function. This gain setting technique makes the gain accuracy to depend on the accuracy of the feedback network transfer function which is much more readily achievable. In integrated circuit technology, the feedback network usually consists of resistors or for MOS technology, of capacitors. With a modern processing technology, these passive elements can be controlled typically within 10% of their designed value and the ratio between neighboring units can reach 0.1% when care is taken to avoid adverse processing effects. Thus 0.1% or correspondingly 10 bit accuracy represents the typical upper limit to the gain accuracy which can be obtained with a monolithic technology. To achieve higher gain accuracy in monolithic circuits, various techniques have been applied. Extra processing steps for obtaining precision elements for example have been developed. Thus laser trimming used on thin metallic film can achieve accuracy of 13 bit level or more. Also the fuse link trimming of resistor network has been used with relative success. But the disadvantage of these methods are the high manufacturing cost associated with the additional process steps.

In this chapter, we deal with circuit techniques that can deliver precise gain values with unmatched capacitors in a MOS integrated circuit technology. This has been seen by many other as a viable alternative to the use of more expensive technological tools in realizing precision circuits. Recently circuit techniques that can realized accurate analog-to-digital conversion without utilizing matching of passive elements have been introduced. One example is the dynamic element matching method developed by Van de Plaasche [9]. This method uses time-multiplexed current sources to achieved exact binary weighing of current sources for a DAC. The current glitches that are generated from the switching is smoothed out by attaching capacitors to the current sources. Linearity up to 15 bits can be achieved. Another method is the self-calibrating method [10] developed by H.S.Lee at U.C. Berkeley where known properties of the converter are used to calibrate the passive elements and later by correcting for these errors, an extremely linear analog to digital conversion can be made. The method that are developed in this thesis uses neither of these techniques but is rather a self-correcting technique in which the error in the gain setting capacitors are corrected in every cycle by means of a switching algorithm.

As it becomes apparent in the last chapter, among the factors that limit the linearity and resolution of the algorithmic analog-to-digital converter are the loop gain error and the loop offset error. Since the loop gain error represents the sum of the gain error in the multiply-by-two amplifier and the sample/hold amplifier, it may have a worst case minimum value of 0.2%. This means that if the algorithmic A/D converter is realized with the conventional gain-setting technique by using ratioed capacitors, linearity higher than 9-10 bit is very difficult to attain. Since this gain setting technique by using ratioed capacitors is inadequate for high resolution algorithmic A/D conversion, a technique for obtaining an exact integral multiplication of a voltage has been developed and is described in section 2.2. This technique involves a sequence of switching operations with a

switched capacitor integrator.

2.2. Ratio-independent Multiplication of a Signal

Conceptually, it is not difficult to obtain an integral multiplication of a signal by capacitors which have arbitrary values. In Fig. 2.1a, n of such capacitors with different capacitances $C_1, ..., C_n$ are connected in parallel to the same voltage source V_{in}. To obtain nV_{in}, the capacitors are disconnected from the voltage source and stacked on top of each other as shown in Fig. 2.1b. The voltage at the top plate of C_n should be nV_{in}, independent of the capacitance values. In practice, however, the capacitors have small parasitic capacitances associated with their terminals. As a consequence, charge sharing would take place between the capacitors and the parasitic capacitances and the actual voltage at the top plate of C_n is smaller than nV_{in}.

Instead of stacking capacitors on top of each other, the ratio-independent multiplication technique used in the present algorithmic A/D converter uses a switched capacitor integrator to sum the signal charge and then redeposit it onto the same sampling capacitor with which the signal charge is originally sampled. These operations are shown in Fig. 2.2a-f. An ideal operational amplifier is assumed in the discussion that follows. Fig. 2.2a shows the initialization phase of the integrator. The operational amplifier is nulled by the switch M_5, while the sampling capacitor C_s is connected to the voltage V_{in} by M_1 and M_2. In Fig. 2.2b the signal charge $Q_{in} = C_s V_{in}$ is transferred to the integration capacitor C_f. While conserving the charge Q_{in} on C_f, another sample of Q_{in} is taken on C_s and integrated onto C_f. These operations are shown in Fig. 2.2c and d. After n such cycles, the total charge on C_f is nQ_{in} and the output voltage of the integrator is $n\frac{C_s}{C_f}V_{in}$. Obviously, the output voltage is still dependent on the ratio of the two capa-

citors. This is because the charge nQ_{in} is residing on C_f which bears no relationship to the sampled charge Q_{in}. Now if the position of the two capacitors are exchanged as shown in Fig. 2.2f by means of the transistor switches M_6 and M_7, the charge residing on C_f is redistributed back onto C_s and the output of the integrator becomes nV_{in} which is the desired multiplication. Because the operational amplifier maintains a virtual ground point at one plate of C_f and drives the other plate with its active output stage, this multiplication scheme is insensitive to parasitics on C_s and C_f. This scheme has been independently proposed by C.C.Lee[11].

2.3. Problems

It was tacitly assumed that all the components that build up the MOS integrator were perfect. In reality a number of effects exist that limit the performance of the ratio-independent-multiply circuit:

1) finite operational amplifier gain
2) finite operational amplifier offset
3) charge injection from MOS transistors
4) capacitance voltage coefficient

The effects of each of these non-ideal behavior on the circuit are analyzed in the following sections and ways to counteract or compensate them, if they are appropriate, are explained. The introduction of these compensation techniques leads to the development of a fully differential integrator with a modified timing control scheme adapted in a implementation of the algorithmic A/D converter.

2.3.1. Finite Operational Amplifier Gain

In order to have complete charge transfer between the capacitors in the ratio-independent-multiply circuit, the operational amplifier used must possess enough gain. The minimum required gain as a function of the gain accuracy for a feedback amplifier is given by

$$a_{min} = \frac{A_{ideal}(A_{ideal} - \Delta)}{\Delta}$$

where A_{ideal} is the ideal feedback gain value and Δ is the error between the actual and the ideal gain.

The actual gain that is needed is double the value because there is a cascade of two amplifiers in the loop. As explained in chapter 1, the loop gain accuracy must be comparable to the linearity of the conversion. Therefore to obtain 1/2 lsb of linearity for a 12 bit converter for example, the gain must be larger than 20000 or 86 dB for a ideal gain of 1.

Another consideration for using high gain in the operational amplifier is the suppression of operational amplifier output nonlinearity. Enough gain is required such that the *variation* in gain would not seriously disturb the gain accuracy.

In a NMOS technology, high operational amplifier gain cannot be achieved easily because of the lack of a satisfactory load device. In a CMOS technology it is easier to construct high gain operational amplifiers. With a conventional design, it would have called for a two stages architecture, which is not very desirable in terms of compensated bandwidth and slewing performance. These considerations lead to the design of an high gain single stage CMOS operational amplifier to be described in chapter 3. Therefore the problem of low dc gain in operational amplifier does not seem to be a fundamental

limitation to the ratio-independent multiplication scheme.

2.3.2. Operational Amplifier Input Voltage Offset

In spite of the inherent matching advantage of monolithic devices, small mismatches still exist between two adjacent, identical transistors of a differential stage. In addition to this device mismatch offset, which is mainly due to limitations in process technology such as photolithography and etching, certain circuit configurations may give rise to a natural imbalance in their output voltage or current, thus generating a systematic offset. Due to the low transconductance of the MOS transistor, such offset errors often take on offset voltages close to 2~10 mV expressed equivalently at the input of an MOS operational amplifier. In precision analog MOS circuits, care must be taken to prevent this offset from overriding the signal.

A very effective method of canceling this offset in switched capacitor circuit is to store the offset voltage on capacitors. This is shown with the MOS gain block in Fig. 2.3a. During the initialization phase, the operational amplifier is nulled and the capacitors connected to the signal source and ground as shown. The offset voltage V_{offset} is stored on capacitor C_f . This can be understood if C_f is disconnected from ground and connected to the output of the operational amplifier as in Fig. 2.3b. In this case the output of the operational amplifier is 0. Subsequently, as the input capacitor C_s is connected to ground, displacing the charge into C_f , the output voltage becomes $C_s \frac{V_{in}}{C_f}$ independent of V_{offset}. One drawback of this approach is that charge injection from the nulling switch M_3 contributes to another offset error. But since this charge injection offset error is not related to the operational amplifier input offset error, the latter can be considered as completely canceled.

2.3.3. Charge Injection Effect

The MOS transistor functions as an analog switch in switched capacitor circuits, *i.e.* it acts as a short-circuit or an open circuit path for transmitting or blocking an analog signal between two circuit positions. A model of the transistor is illustrated in Fig. 2.4. The extrinsic MOS transistor is shown as the intrinsic MOS transistor together with the gate-source and gate-drain overlap capacitances. Such a partition is especially useful for considering the charge injection effect, because it separates the linear elements (the overlap capacitances) from the nonlinear elements (the intrinsic MOS transistor).

When an MOS transistor switch is turned on, a quantity of charge is stored in its channel. Subsequently, this charge is injected into the surrounding circuit nodes when the transistor is turned off. This phenomena is commonly known as the charge injection effect. The magnitude of this charge can be expressed in a first order equation as follows:

$$\begin{aligned} Q &= C_{ox}\,(V_g - V_t) - C_{ox}\,V_s \quad sp\ 1 \\ &= Q_1 - Q_2 \end{aligned} \tag{2.1}$$

where

Q : charge stored in channel

C_{ox} : gate capacitance of the transistor

V_g : gate voltage

V_t : threshold voltage of MOS transistor

$V_s = V_d$: source or drain voltage

$Q_1 = C_{ox}\,(V_g - V_t)$

$Q_2 = C_{ox}\,V_s$

For simplicity, the back gate bias dependence of the threshold voltage has been neglected and the gate-source and gate-drain overlap capacitances are taken to be external elements. The difference between this injected charge and the charge injected form the overlap capacitances is that the transistor does not inject its channel charge continuously, but only until its gate voltage reaches the value $V_t + V_s$.

Notice that this charge has been split into two components, Q_1 and Q_2. The first component Q_1, which shall henceforth be called the charge injection offset, is independent of the drain/source voltage and potentially gives rise to an offset error in a circuit. The second component Q_2, which is called the voltage dependent charge injection, is dependent on the source/drain voltage and can give rise to a gain error in a circuit if proper measures are not taken to eliminate it.

2.3.3.1. Effect of Charge Injection

To illustrate the effect of charge injection on a circuit, consider a simple MOS gain block in the transition from the sample to the hold mode. Fig. 2.5 shows the gain block which consists of an MOS operational amplifier, a sampling capacitor C_1, a feedback capacitor C_2 and three n-channel MOS transistor switches M_1-M_3. During the transition, transistors M_1 and M_3 must be turned off and M_2 turned on. For controlling the transistors, the gates of M_1 and M_3 are usually tied to the same control line or the same switching waveform. Now if the input voltage V_{in} is at a potential higher than the ground potential, transistor M_1 will be turned off first, at time t_1 as shown in Fig. 2.6a, causing the channel charge to be injected back into the voltage source node A and the operational amplifier summing node B. The state of the circuit between time t_1 and t_0 where M_3 turns off is illustrated in Fig. 2.6b. The partition of the channel charge in M_1 into source and drain is a complex function of the control voltage waveform and the

node impedances[12][13], but in this analysis, it is assumed that the charge is split into two equal halves. Since transistor M_3 is still conducting when the charge injection takes place, the charge injection will displace from the sampling capacitor an amount of charge equal to

$$Q_{inj} = -\frac{1}{2}C_{ox}(V_g - V_t) + \frac{1}{2}C_{ox}V_{in} \tag{2.2}$$

so that the resultant sampling charge Qs becomes

$$\begin{aligned} Q_s &= C_1 V_{in} + \frac{C_{ox}}{2}V_{in} - \frac{C_{ox}}{2}(V_g - V_t) \\ &= (C_1 + \frac{C_{ox}}{2})V_{in} - \frac{C_{ox}}{2}(V_g - V_t) \end{aligned} \tag{2.3}$$

Subsequently, when M_2 is turned on, Q_s is transferred onto C_2 and causes an output voltage equal to

$$\begin{aligned} V_o &= (\frac{C_1}{C_2} + \frac{C_{ox}}{2C_2})V_{in} - \frac{C_{ox}}{2C_2}(V_g - V_t) \\ &= A'V_{in} + V_{offset} \end{aligned} \tag{2.4}$$

Ideally, if there were no charge injection, the gain of the circuit A would be $\frac{C_1}{C_2}$. Thus through the charge injection effect, the gain becomes A' and there is an additional offset term V_{offset}. The alteration of the gain can be attributed to the signal dependent charge injection component Q_2.

In this analysis the charge injection from M_3 has not been taken into account. It can be easily deduced that it would contribute another offset component to the gain block. When the operational amplifier is in the unity gain feedback mode as it is when M_3 is on, the source/drain voltage of M_3 is ideally zero thus independent of V_{in}. This means that under all circumstances the charge injected by M_3 will be a constant.

2.3.3.2. Cancellation of Signal Dependent Charge Injection

By proper scheduling of the timing control sequence, the signal dependent charge injection component and therefore the gain error can be eliminated. The timing diagram for achieving this is depicted in Fig. 2.7a. Notice that by delaying the switching off of M_1, M_3 can be controlled to switch off first at time t_0. The state of the circuit between t_0 and t_1 is shown in Fig. 2.7b. There is indeed still charge injection from M_3, but because the operational amplifier in negative feedback places M_3 at a potential (0 for ideal operational amplifier) independent of the input signal, the displacement charge that flows into the operational amplifier summing node is

$$Q_{inj} = -\frac{C_{ox}}{2}(V_g - V_t). \tag{2.5}$$

and depending on the speed of the operational amplifier and the node impedances of the circuit, this charge may distributes itself between the sampling capacitor C_1 and the feedback capacitor C_2 with varying proportions. But once M_2 is turned on, the feedback action of the operational amplifier forces all charge to reside on C_2. As a result, the output voltage during the hold mode becomes

$$V_o = \frac{C_1}{C_2}V_{in} - \frac{C_{ox}}{2C_2}(V_g - V_t) \tag{2.6}$$

and the gain is equal to the ideal gain.

2.3.3.3. Cancellation of Charge Injection Offset

The use of fully differential circuitry in switched capacitor filters [14] has contributed both to the increase of dynamic range and ease of design. In addition, it leads to a first order cancellation of the charge injection offset. This can be explained with Fig. 2.8, which shows the differential implementation of the MOS gain block that was described

earlier. It consists of two sets of capacitors C_1C_3 and C_2C_4 which constitute the two signal paths. The MOS transistor switches are duplicated in the two paths. In the transition from the sample to the hold mode, the transistor pairs M_1M_3 and M_2M_4 are turned off first. But instead of connecting the terminals of the capacitors to ground, they are connected together by means of the transistor M_5. In making use of the proper timing sequence explained in the last section, transistors M_2 and M_4 both inject charge into the operational amplifier summing nodes and displace charges in C_1 and C_2. This charge injection is however common to both channels of the differential circuit and appears therefore as a common mode signal to the operational amplifier which is then suppressed by the common mode feedback circuit in the operational amplifier. The output is not affected if the transistors M_3 and M_4 have the same gate capacitance and overlap capacitances. In actuality, the transistors can be mismatched by as much as 10-20% in their capacitances and the cancellation is only approximate. The resultant offset error voltage in this case is

$$V_{offset} = \delta C_{ox} (V_g - V_t) \tag{2.7}$$

where

δC_{ox} : difference in the gate capacitances

This represents a significant reduction in the error due to the charge injection from MOS transistor.

2.3.4. Capacitance Voltage Coefficient Cancellation

The capacitor used in an MOS process is usually voltage dependent. This voltage dependence can be expressed as a Taylor series

$$C(V) = C_o (1 + \alpha_o V + \alpha_1 V^2 + \cdots) \tag{2.8}$$

where

C_o : zero voltage capacitance

$\alpha_o, \alpha_1, \cdots$: capacitance voltage coefficients

Normally, this capacitance voltage dependence introduces nonlinearities into the circuit which are intolerable for high precision circuit operations. However, the same differential circuit which is used to cancel charge injection effect also gives a cancellation of odd order capacitance voltage coefficients of the MOS capacitors. This is because an increase of voltage or capacitance in one channel is accompanied by an equal but opposite decrease of voltage or capacitance in the other channel of the circuit. These changes combine to give overall capacitance voltage coefficients equal to the difference of the odd order capacitance voltage coefficients of the capacitors in the two channels. The even order capacitance voltage coefficients are summed together. Since $\alpha_1, \alpha_2, \ldots$ are much smaller than α_o, they can be neglected in a first order analysis and the net result is a voltage dependence of $\delta\alpha_o$.

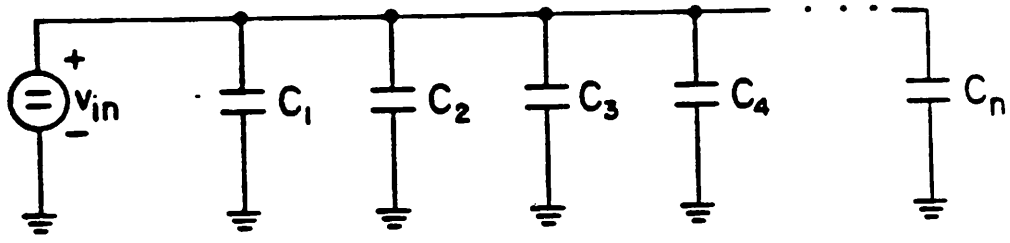
2.4. Ratio-independent Multiply-by-Two Switching Algorithm

Summarizing the foregoing discussions, the ratio-independent multiply-by-two function can be realized using a differential gain block or integrator with a switching sequence as shown in Fig. 9a-d. Assuming that the input differential voltage $V_+ - V_-$ is constant during the whole operation, it is first sampled onto the sampling capacitor C_1 and C_2 in the first clock cycle (Fig. 2.9a). This signal charge is then transferred onto the integrating capacitors C_3 and C_4 during the second clock cycle by turning on the switch M3 (Fig. 2.9b). Subsequently, C_3 and C_4 have to be separated from the operational amplifier feedback loop to prepare for the second signal sampling. To accomplish this task without introducing signal dependent charge injection, the capacitors are

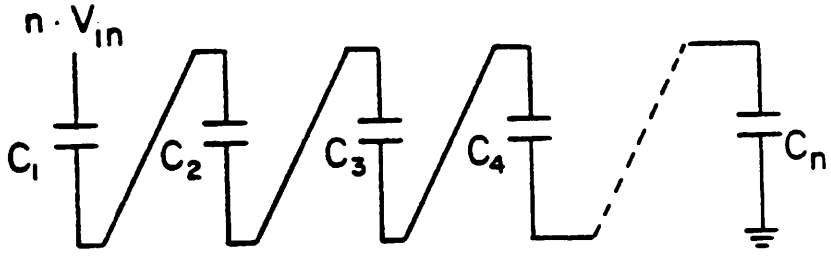
separated from the operational amplifier input nodes. Another signal sample is then taken (Fig. 2.9c), and during the fourth or the last clock cycle, the first signal charge is redistributed back onto C_1 and C_2 from C_3 and C_4 (Fig. 2.9d). The operation takes four operational amplifier settling time periods which determine the length of the basic clock cycle.

2.5. Ratio-independent Sample/Hold

A ratio-independent sample/hold function can be implemented very simply by an MOS operational amplifier and two capacitors. An intermediate storage capacitor is not needed. The operation is described in Fig. 2.10. Using a fully differential implementation again, the differential signal is sampled onto the capacitors as shown in fig. 2.10a. Any offset voltage on the operational amplifier would be stored onto the capacitors as well. Following this, the switches M_3 and M_4 are turned off and the capacitors connected to the output by means of switches M_5 and M_6. The negative feedback then forces the operational amplifier outputs to assume the same potentials as the input voltages.



a



b

Figure 2.1

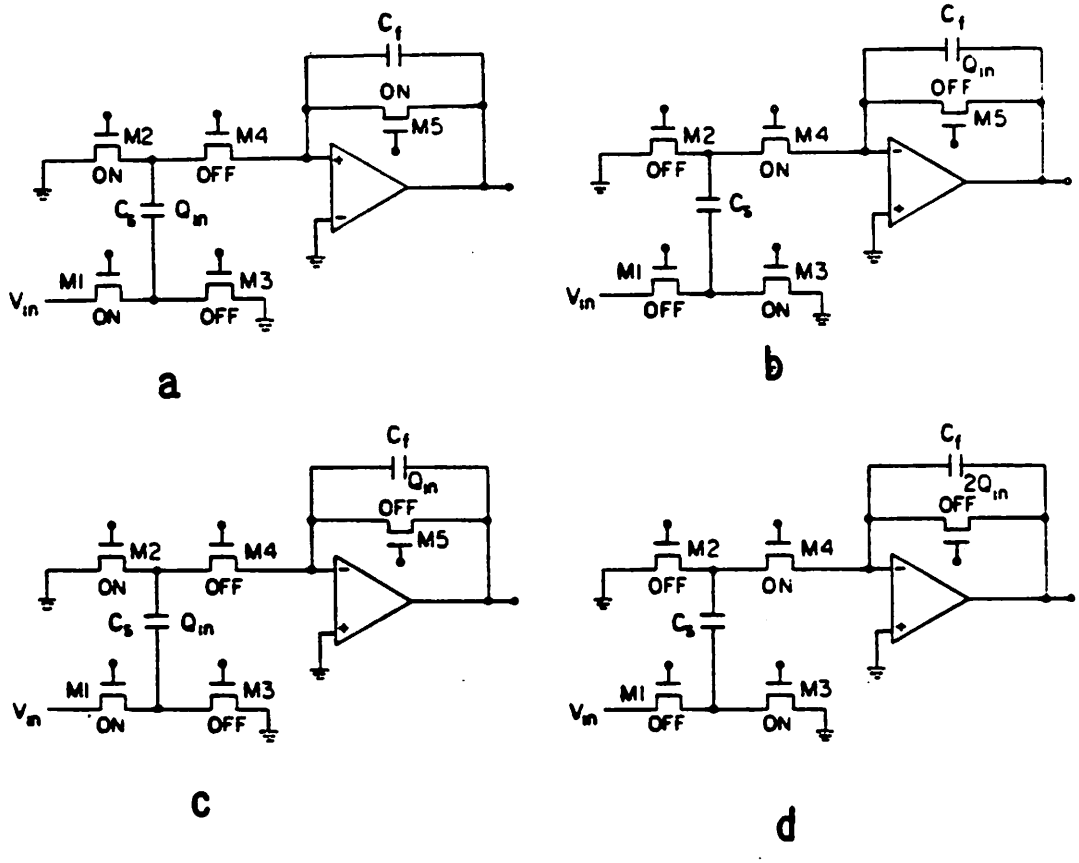

C_f
ON
M5
M2
M4
ON
OFF
C_s
Q_in
M1
M3
V_in
ON
OFF
a
C_f
Q_in
OFF
M5
M2
M4
OFF
ON
C_s
M1
M3
V_in
OFF
ON
b
C_f
Q_in
OFF
M5
M2
M4
ON
OFF
C_s
Q_in
M1
M3
V_in
ON
OFF
c
C_f
2Q_in
OFF
M5
M2
M4
OFF
ON
C_s
M1
M3
V_in
OFF
ON
d

Figure 2.2

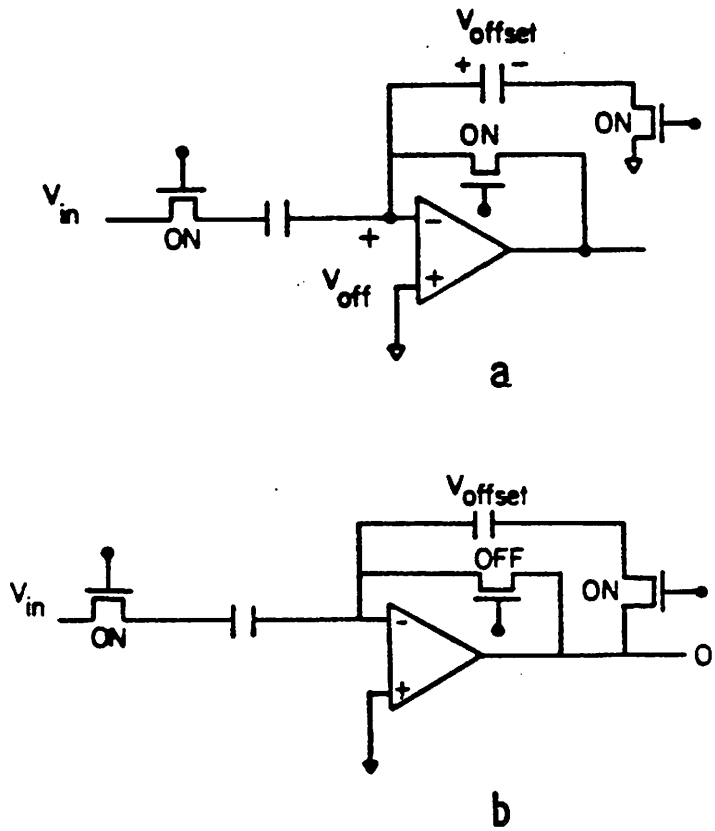


Figure 2.3

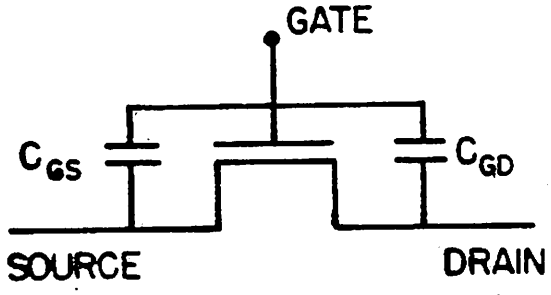


Figure 2.4

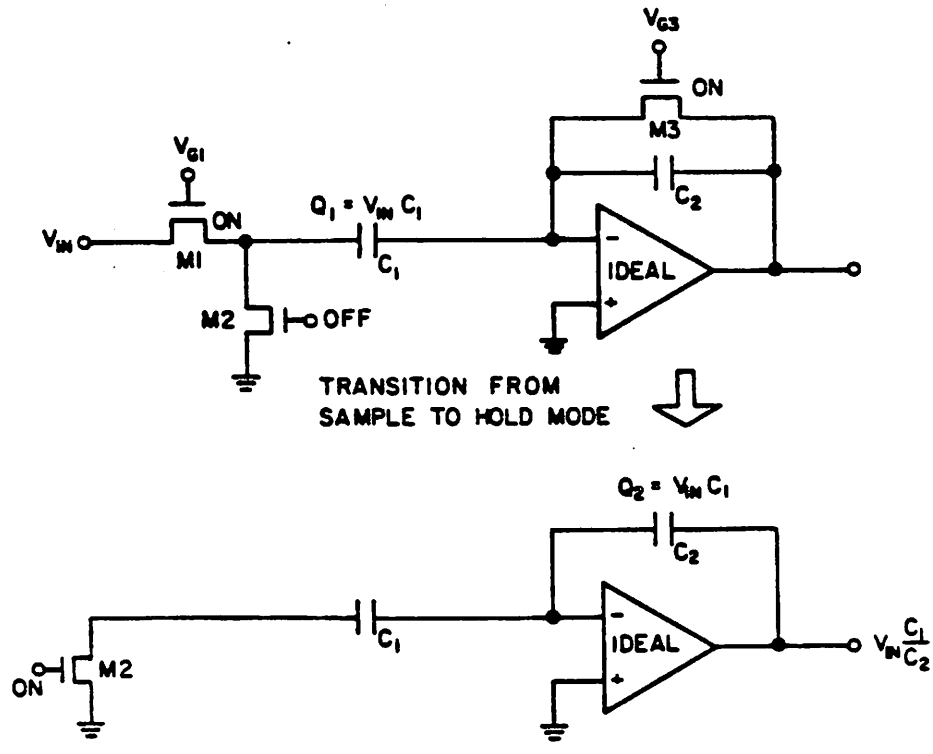

V_{G3}
ON
M3
C_2
V_{G1}
$Q_1 = V_{IN} C_1$
V_{IN}
ON
M1
C_1
IDEAL
M2
OFF
TRANSITION FROM
SAMPLE TO HOLD MODE
$Q_2 = V_{IN} C_1$
C_2
C_1
IDEAL
$V_{IN} \frac{C_1}{C_2}$
ON
M2

Figure 2.5

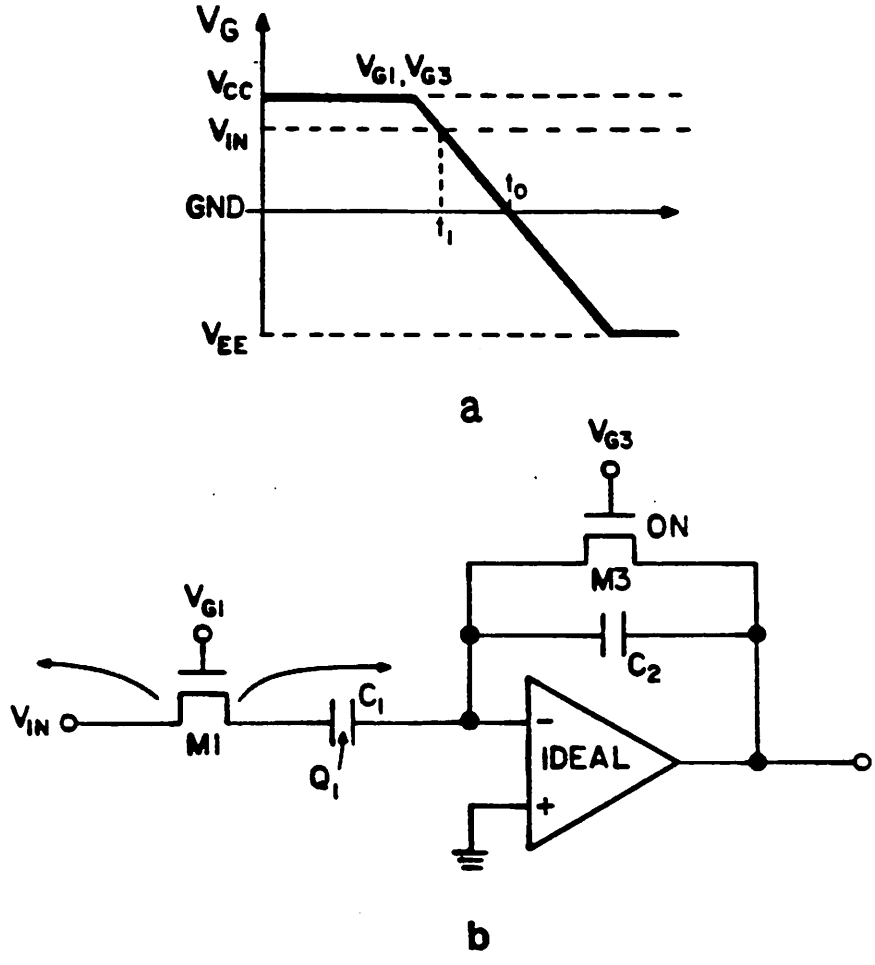

V_G
V_{CC}
V_{G1}, V_{G3}
V_{IN}
GND
t_0
t_1
V_{EE}
a
V_{G3}
ON
M3
C_2
V_{G1}
V_{IN}
M1
C_1
Q_1
IDEAL
b

Figure 2.6

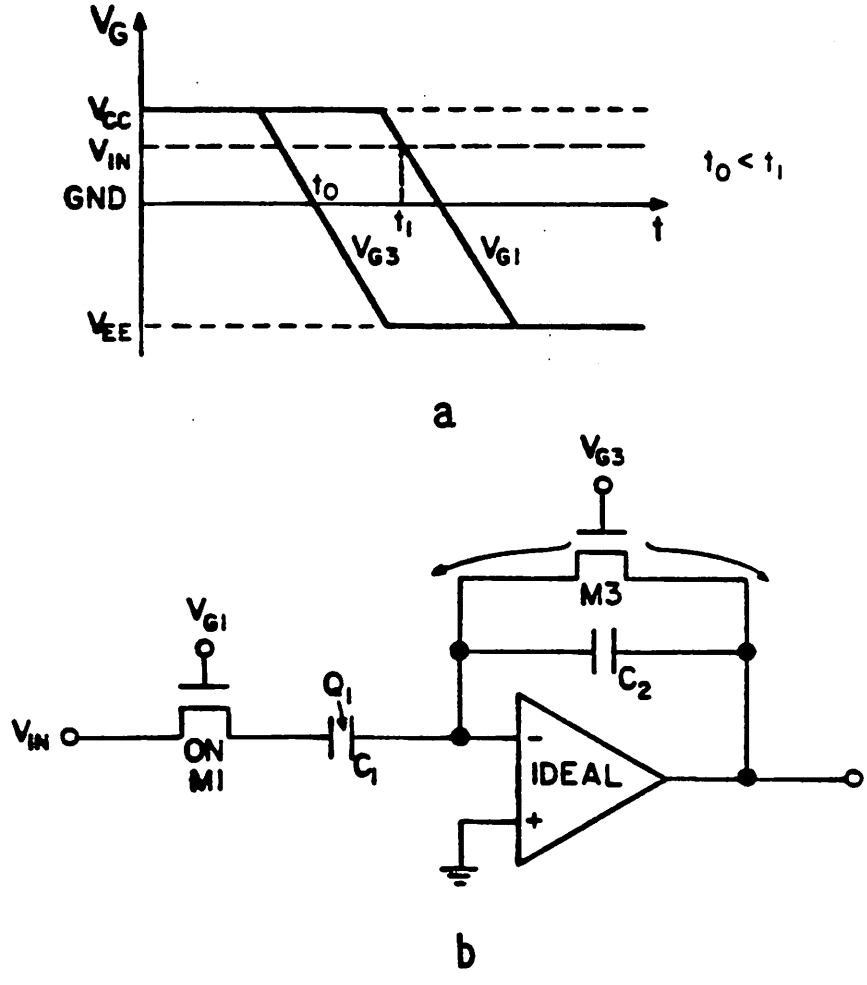

V_G
V_{CC}
V_{IN}
GND
V_{EE}
t_0
t_1
V_{G3}
V_{G1}
t
$t_0 < t_1$
a
V_{G3}
M3
V_{G1}
C_2
Q_1
V_{IN}
ON
M1
C_1
IDEAL
-
+
b

Figure 2.7

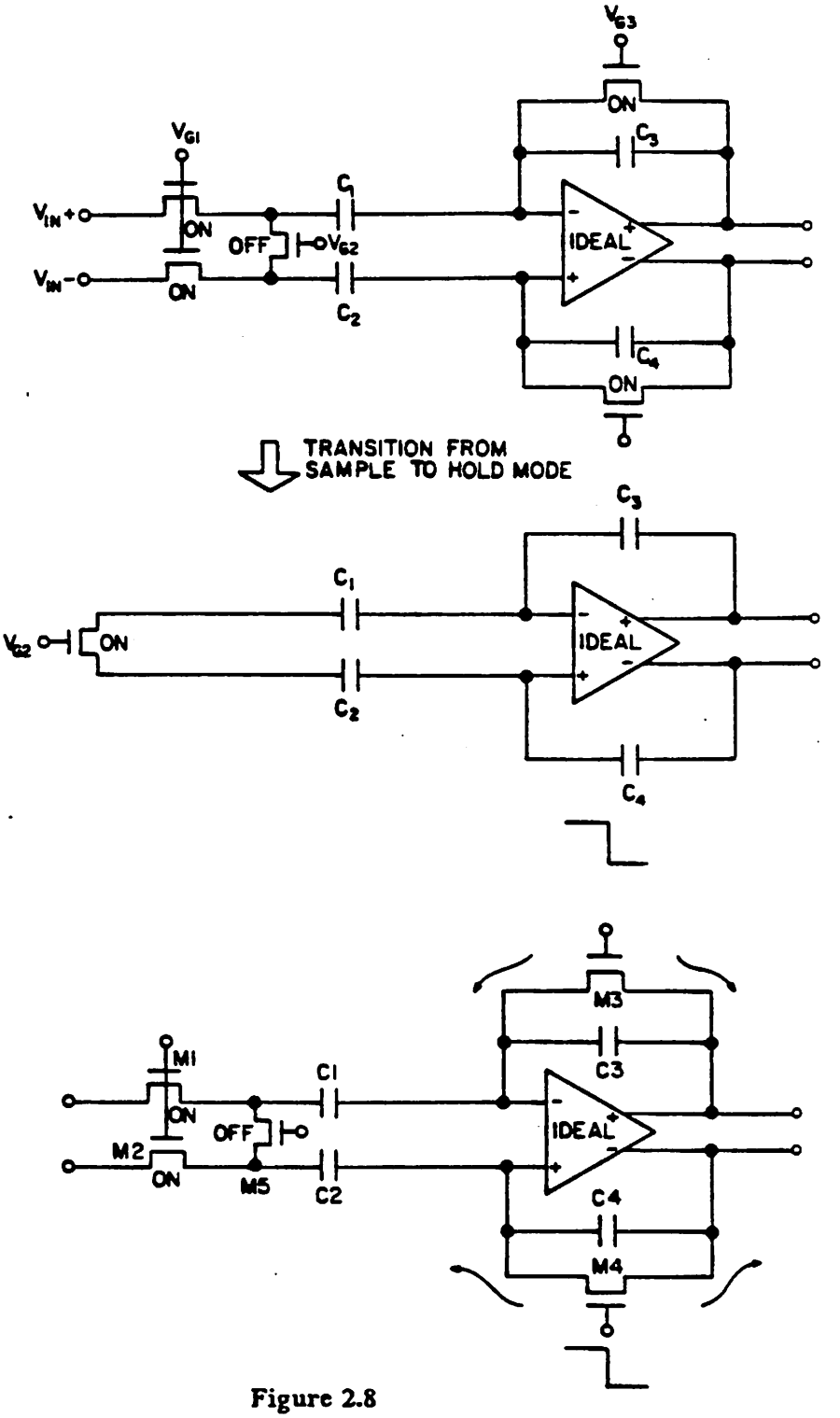

V_{G3}
ON
C_3
V_{G1}
C_1
$V_{IN}+$
ON
OFF
V_{G2}
IDEAL
$V_{IN}-$
ON
C_2
C_4
ON
TRANSITION FROM
SAMPLE TO HOLD MODE
C_3
C_1
V_{G2}
ON
IDEAL
C_2
C_4
M3
C3
M1
C1
ON
OFF
IDEAL
M2
ON
M5
C2
C4
M4

Figure 2.8

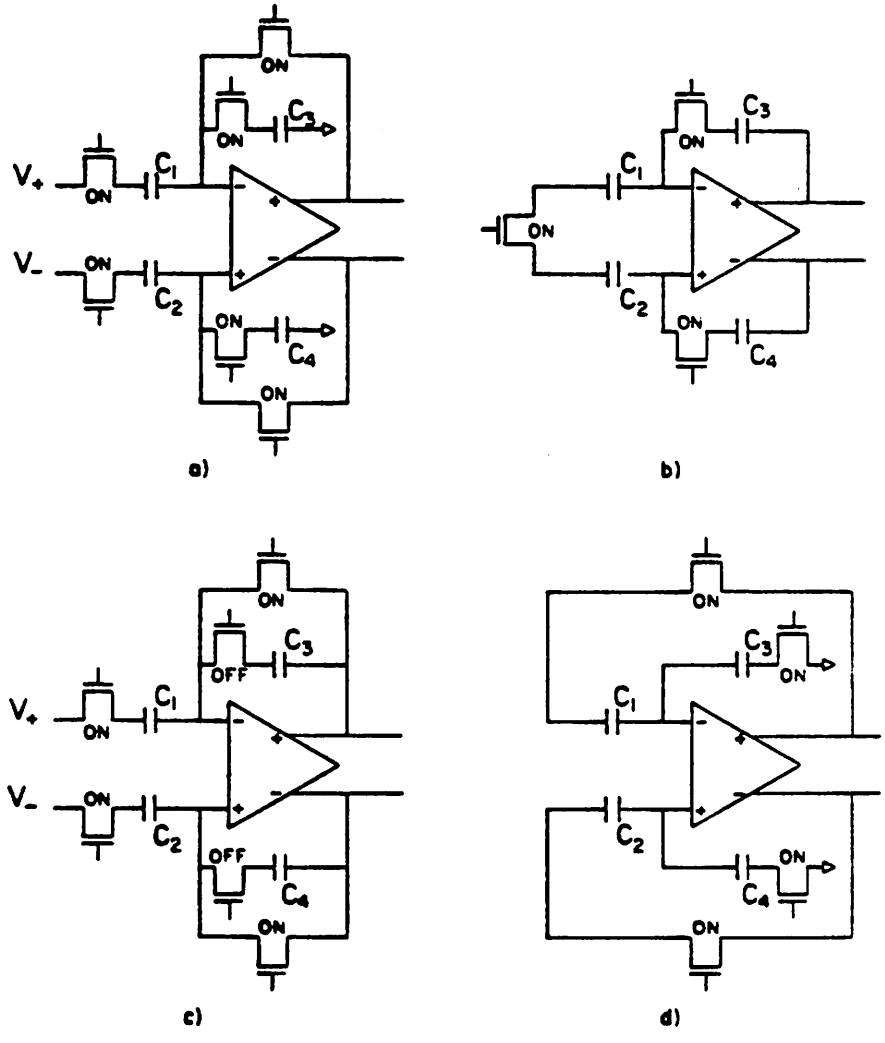
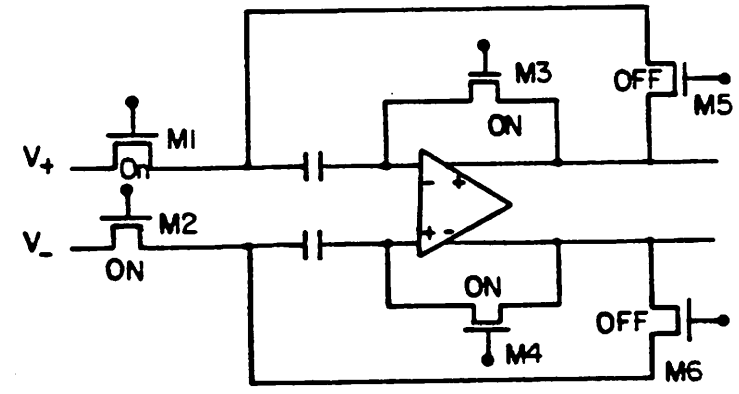
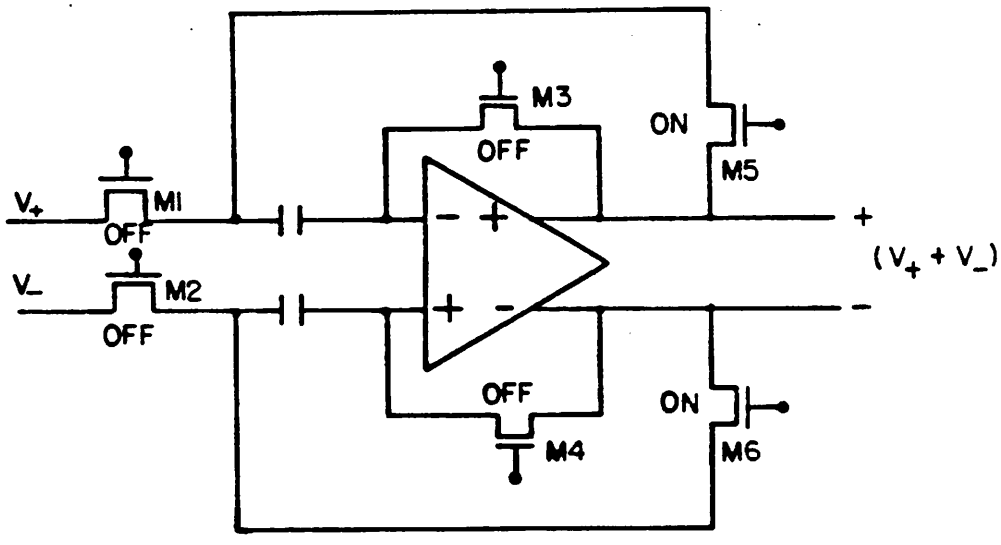


Figure 2.9



a



b

Figure 2.10

CHAPTER 3

Operational Amplifier Design

3.1. Introduction

In the design of high accuracy sampled-data switched capacitor circuits, the operational amplifier constitutes the most important building block in the system. Without optimal performance from such an amplifier, accuracy and speed can not be achieved on the basis of an theoretical switching algorithm only. We have seen that accuracy is synonymous to high dc gain in the operational amplifier. Speed in this context is generally identical to how fast an operational amplifier in a particular feedback configuration can settle to a certain percentage of its final output value in the presence of a input signal. Thus the speed is not only a function of the small signal bandwidth of the operational amplifier but also one of its slew rate. For the algorithmic analog to digital converter there are special requirements that have to be fulfilled. Some sundry items concerning the design of operational amplifiers in CMOS technology will be discussed here.

3.2. Motives

The designs of operational amplifiers for the application in a switched capacitor filter and in a algorithmic A/D converter have one fundamental difference. In a switched capacitor filter, the operational amplifier output only need to swings small voltage steps whereas in the algorithmic A/D converter, the output voltage can cover the

maximum signal dynamic range in the same amount of time. This particular application for the algorithmic converter necessitates an operational amplifier without slew rate limitation since a large part of the time needed for the output to converge to its final value is spent in the slewing period. For the MOS operational amplifier, two aspects of this slew rate limitation are apparent. First, for a single stage operational amplifier, the load capacitor acts as the compensation capacitor and determines the maximum slew rate and the small signal bandwidth of the operational amplifier. By a single stage operational amplifier it is meant an operational amplifier that has only one high impedance node. This is also true for the two stage amplifier. Second, for 2 stage operational amplifier designs, because of the finite driving capability of the second stage, the maximum slew rate at the output is the smaller of the two slew rates resulting from the internal compensation capacitor or the external loading capacitor. Additionally, it may contribute to another high frequency pole which affects the high frequency stability of the closed loop amplifier. Thus the configuration of the operational amplifier has an important bearing on various performance parameters.

In the present design, the sample, load and feedback capacitors need not be matched and are generally in the range of 2-4 pF. Therefore the operational amplifier must have a dc gain of up to 50000 and be capable of settling to 0.01% of the final accuracy within $1\mu s$ for a differential voltage step of 4-6V while driving such capacitances. The specification for this speed translates approximately into a small signal bandwidth of larger than 5MHz and a slew rate of larger than $25V/\mu s$ in both directions.

3.3. Design Considerations

Before we decide what configuration the operational amplifier should take, a thorough understanding of the $g_m r_{out}$ or the gain product of a transistor is necessary.

The gain product is a measure of the maximum gain that can be obtained from a single device assuming it is biased by ideal current sources.

3.3.1. Gain Product of A Transistor

Fig. 3.1 shows a plot of the gain product of a MOS transistor as a function of the drain current operating with a fixed drain voltage. It can be seen that this product increases with decreasing current until it reaches the subthreshold operating region of the transistor where it flattens off. This occurs because the transconductance of the device g_m is proportional to the square root of the drain current I_d

$$g_m = \sqrt{2 I_d \mu C_{ox} (W / L)} \tag{3.1}$$

and the output impedance is inversely proportional to the drain current

$$r_{out} = (\frac{1}{I_d})(\frac{1}{\lambda}) \tag{3.2}$$

so that their product becomes

$$g_m r_{out} = (\frac{1}{\sqrt{I_d}})(\frac{1}{\lambda}) \sqrt{2 \mu C_{ox} (W / L)} \tag{3.3}$$

As the device enters the subthreshold region, the transconductance becomes directly proportional to the drain current

$$g_m = \frac{I_d}{n V_T} \tag{3.4}$$

where V_T is the temperature voltage and n is the slope factor of the device and is equal to

$$n = 1 + \frac{C_d}{C_{ox}} \tag{3.5}$$

thus making the gain product independent of the drain current:

$$g_m r_{out} = \frac{1}{n \lambda V_T} \tag{3.6}$$

In a conventional MOS process, this maximum gain product is above 2000 for a long channel (*e.g.* $> 6\mu m$ electrical length) N-channel MOS device. Practically, because the transistor has to be biased so that it can achieve reasonable transconductance, the average value achievable is below 100. The implication of this result is that for the single transistor gain stage with active load, the maximum voltage gain that can be obtained with good frequency response is below 200 for long channel ($> 6\mu m$) transistors, given that the active load has about the same r_{out} as the driving device. This statement is of course fully arbitrary because we have not specified what a good frequency response is. Perhaps a good point to remember is that in the present state of the art MOS analog circuit design, the MOS transistor is biased to get at most 10-30% of the maximum available transconductance.

Therefore if higher gain is to be obtained, several single transistor stages each with a low gain product can be cascaded or some circuit of obtaining higher gain products from the basic transistor in a single stage can be used One way of doing this is by the cascoding of transistors.

3.3.2. Cascoding of Transistors

With the cascode circuit, it is possible to achieve very high output impedance and therefore high voltage gain in a single amplifying stage. The two transistor circuit is shown in Fig. 3.2. It can be seen that since the drain to source capacitance in M_2 and the

drain-gate overlap capacitance in M_1 are extremely small, the high frequency feedback from the output back on the input is negligible. Also because the voltage gain from the input to the source of the upper transistor M_2 is low:

$$\frac{V_1}{V_{in}} = -\frac{1}{1+\lambda}\frac{g_{m1}}{g_{m2}} \tag{3.7}$$

where λ is the backgate bias factor and is equal to

$$\lambda = \frac{\gamma}{2\sqrt{-V_{bs}+2\phi_F}}$$

Thus for transistors of the same size, the gain is actually smaller than 1, which means that the Miller effect due to the drain gate overlap capacitance in M_1 is reduced. This fact is important when the input transistor is driven with another diode connected transistor as in the current mirror. When looking into the drain of M_2, the output impedance is

$$r_{out} = g_{m2}r_{out2}(1+\lambda_2)r_{out1} \tag{3.9}$$

As a result the voltage gain available from such a stage is

$$A = \frac{V_{out}}{V_{in}} = g_{m1}r_{out1}g_{m2}r_{out2}(1+\lambda) \tag{3.10}$$

We see that this gain A is composed of the product of two $g_m r_{out}$'s together and is therefore proportional to the inverse of the drain current.

This idea of cascoding transistors can be extended to include three transistors from which a double cascode circuit is obtained. The circuit is shown in Fig. 3.3. The maximum gain product of this circuit can be calculated to be

$$A = g_{m\,1} r_{out\,1} g_{m\,2} r_{out\,2} g_{m\,3} r_{out\,3} (1 + \lambda_2)(1 + \lambda_3) \tag{3.11}$$

in which case is proportional to $\frac{1}{I_d^{\frac{3}{2}}}$. This circuit configuration gives rise to extremely high gain even if the transistor bias current is large. However this is bought with an corresponding in the gain sensitivity with respect to the bias current and the transistor gate length. As was mentioned earlier, the gain product of the transistor decreases very rapidly as the gate length is shrunken from $4-5\mu m$ to $1-2\mu m$. With the gain proportional to the cube power of this product, the reduction in gain could become substantial. Also since the "upper" transistors have to be biased so that the "next" transistor underneath it is in the saturation region of operation at all times, the output voltage range is limited to a narrow range. In fact, this is the major limitation to the application of the cascode circuit to low voltage supply circuits.

The frequency behavior of the double cascode circuit is quite simple. In addition to the dominant pole p_1 formed by the nodal capacitance on the drain of M_3 and the output impedance there are 2 nondominant poles. They are formed by the nodal capacitances on the sources of M_3 and M_2 and the impedances at those nodes.

$$\begin{aligned} p_2 &= \frac{g_{m\,2}}{(1 + \lambda_2) C_2} \\ p_3 &= \frac{g_{m\,3}}{(1 + \lambda_3) C_3} \end{aligned} \tag{3.12}$$

The capacitances C_2 and C_3 are composed of the sum of the gate capacitances from the individual transistor and the parasitic capacitance connected with the nodes such as drain/source junction capacitance or interconnect capacitance.

$$C_3 = \frac{2}{3}C_{ox\,3} + C_{parasitics\,3} \tag{3.13}$$

Thus M_2 and M_3 are operating very near to their intrinsic speed. To reduce C_2 and C_3 the gates of those transistors can be put very near to each other as shown in Fig. 3.4a for a single polysilicon MOS process. For a double polysilicon MOS process, the parasitic node capacitances can be eliminated by overlapping the two gates of the transistors which are formed by different layers of polysilicon[15]. This layout is shown in Fig. 3.4b.

In principle, the cascode transistor circuit can be extended to N transistors, with the corresponding composite gain product equal to

$$A = (g_m r_{out})^N (1 + \lambda)^N \tag{C3.14}$$

However, for reasons to become apparent later, repeated cascoding cannot be used, both in the bipolar as well as the MOS technologies. The double cascode configuration is not used in bipolar circuit design because the collector-base resistance r_μ in the cascode transistor would limit the maximum obtainable impedance. This resistance would bleed current away from the negative feedback path formed by the cascode circuit into the biasing voltage source, in contrast to the output impedance of the cascode transistor which conducts a current only to the next lower transistor in the feedback path (Fig. 3.5). Thus it acts effectively as a shunt resistor from the high impedance node to ground.

IN MOS transistors such a resistance does not exist due to the insulating nature of the gate. The maximum obtainable impedance at a node is limited by other effects. One of them is the reverse biased leakage in the drain junction. This resistance of a $10\mu m X\ 100\mu m$ junction can easily reach several $G\ \Omega$ for a well controlled MOS process. Another one is the surface leakage currents between two conductors. This leakage depends on the surface cleanliness and is usually negligible if the wafer is passivated

with an additional passivation layer such as CVD oxide or nitride. Nevertheless the effect of impact ionization in the channel of an MOS transistor eventually this impedance level.

3.3.3. Impact ionization in MOS Transistors

It has long been recognized that channel current in the high field region near the drain causes low-level avalanche multiplication effect thus resulting in a substantial substrate current. This problem is not limited to the short channel transistors alone but appears also in longer channel devices. Fig. 3.6a shows a plot of the substrate current for a $100\mu m$ / $10\mu m$ device as a function of the gate voltage for several drain voltages. Fig. 3.6b is a plot for a $100\mu m$ / $2\mu m$ device. This substrate current can have several effects on the operations of analog integrated circuits. Firstly, a transistor can have different characteristics if they are located near to an adjacent transistor which injects substrate current. Usually, in the presence of a large substrate current, the drain current increases for the same bias voltages. As a result, the operating bias currents of some circuit paths may differ significantly from the designed values and cause parts of the circuit to enter into undesired operating states. Secondly, the strong dependence of the substrate current and therefore the drain current on the gate and the drain voltages represents a shunting resistance from the channel to the substrate. This would lower the impedances of high impedance circuit nodes and make them highly nonlinear with respect to the node potential.

From the circuit designer's viewpoint, the impact ionization current of the transistor should best be eliminated. But this is not always possible because of the tendency of modern MOS process in going to shorter channel lengths which further aggravates the problem. In seeking to understand and solve this problem, there is the problem of modeling which have to be overcome. It is now feasible to model the substrate current of a

MOS transistor by relating it to the maximum electrical field in the device by an exponential function which has the form

$$I_{sb} = I_{sbo}\, e^{\frac{a}{E_{max}}} \tag{3.15}$$

so that the shunting effect of the substrate current can be accounted for[16]. The proximity effect of substrate current injection however cannot be modeled without exact knowledge of the layout and technological parameters. In most cases such calculations are too tedious to carry out so that the design still has to rely on the experience of the designer.

A effective method of reducing the impact ionization current in transistor experiencing large drain voltage excursions is shown in Fig. 3.7. Another transistor which operates in the linear region is added on top of the original device. Because there is no large electrical field at the drain region for such a device, the impact ionization current loss to the substrate becomes negligible. The original device is prevented from entering the high voltage operation by the cascode device.

3.3.4. One Gain Stage Versus Two Gain Stage Design

The accepted practice for designing operational amplifiers with high dc gain is to used a two gain stage configuration. Fig. 3.8 shows an CMOS operational amplifier with such a configuration, which is a classical design very similar to that of the bipolar technology. It consists of a differential input stage together with the next stage which is either a class A or class AB inverting stage. Each stage is designed to have gain ranging from 30-100 so that the overall gain is from 900-10000. The operational amplifier is frequency compensated by a pole splitting capacitor C_c and a resistor r_Z .

Sufficiently high gain can also be obtained in a single gain stage configuration by using a cascading of transistors in the current source. Fig. 3.9 shows the single gain stage folded cascode configuration. The gain in this case is approximately proportional to $(g_m r_{out})^2$. With $g_m r_{out}$ close to 30 -100 the total gain is between 900-10000. Thus the obtainable gain in both configurations are comparable. The single stage operational amplifier is compensated by the load capacitor C_L attached to the output.

More gain can be obtain in both configurations by appending cascode transistors to the gain stage(s). In the single stage configuration, the output voltage swing can be severely limited by such a measure. For example for a supply voltage of +/-5V, double cascode output appears to be the ultimate limit if any useful voltage swing is required. For the two gain stage configuration, providing that the second gain stage can supply enough gain, which usually is the case, the first gain stage can be cascaded to any degree as long as the cascode transistors can be kept in their saturation operations. Therefore, for low power supply voltages, the two gain stage configuration seems to possess an inherent advantage over the single gain stage configuration in high dc gain designs.

In terms of the frequency response, there are no consensus as to which configuration can give the largest bandwidth. Although the heuristic rule states that the simpler the circuit, the faster it will become. Referring to Fig. 3.10, which shows the small signal equivalent circuit of the operational amplifier, the uncompensated two stage configuration has two low frequency poles which contribute excessive phase shifts above the unity gain frequency and is often compensated by pole-splitting. This involves the addition of a Miller capacitance to the second gain stage such that through the interaction of the poles and the feedforward zero. In MOS operational amplifiers, this is done by the addition of a capacitor plus a resistor or an active stage. With the resistor, the feedforward zero is moved from the right real axis to the left half axis and thereby draw out the two

complex secondary poles to higher frequencies. Alternatively, an active feedback stage can be used to eliminate the feedforward zero, creating only the Miller capacitance which is necessary for the compensation. One such realization, originally due to Read and Weiser, is shown in Fig. 3.11. With such a compensation method, it is claimed that the second pole can be pushed out to extremely high frequency. Following the analysis of Ahuja[18], the pole locations are

$$p_1 = \frac{1}{(g_{m2}R_2)C_c R_1}$$
$$p_2 = g_{m2}\frac{C_c}{C_1(C_c + C_L)} \tag{3.16}$$

Thus with a small load capacitance C_c, the second pole approaches

$$p_2 = \frac{g_{m2}}{C_1} \tag{3.17}$$

which is near the maximum operating frequency of the second stage transistor. Allowing for a certain phase margin at the unity gain frequency, the operational amplifier can be operated at the maximum frequency as determined by the second pole.

In a single stage operational amplifier, the load capacitor acts simultaneously as the compensation capacitor. The first pole position is given by

$$p_1 = \frac{g_m}{C_L} \tag{3.18}$$

This pole is again to be places so that sufficient phase margin remains. The second pole location is given by the cascode transistor transconductance and the nodal capacitance associated with it. This capacitance comprises of the gate capacitance of the transistor together with the parasitic drain/source and interconnect capacitances. Thus the second pole is at least comparable to that of the two stage operational amplifier. Therefore from

this point of view of analyzing the two configurations by means of first order equations, there is almost no difference at all between the maximum achievable bandwidths.

3.3.5. Slew Rate Considerations

One disadvantage of the classical two stage CMOS operational amplifier configuration for application in data acquisition circuit is the slew-rate limitation imposed by the finite current in the first stage and the compensation capacitor.

$$S.R. = 2\frac{I_d}{C_c} \tag{3.19}$$

where g_m is the input device transconductance, I_d is the device bias current and $\omega 1$ is the unity-gain frequency of the amplifier. The single stage folded cascode operational amplifier has the same problem except that the maximum slew rate is determined by the load capacitance

$$S.R. = 2\frac{I_d}{C_L} \tag{3.20}$$

In reality, the two stage configuration has another source of slew rate limitation in the CMOS technology. This is because in the previous discussion, it is assumed that the output stage has a low enough output impedance such that driving a (capacitive) load would not be a problem. This assumption may be valid for a bipolar output stage which employs emitter followers with extremely low output impedance. The CMOS source follower, however, has a output impedance which is about 2-3 orders of magnitude higher than its bipolar counterpart. It is therefore definitely necessary to take this into account when designing amplifiers for driving large capacitive loads.

3.3.6. Methods For Improving The Slew Rate

For a two stage operational amplifier, the slew rate is linked to its unity gain frequency $\omega 1$ by the relation

$$S.R. = 2\frac{I_d}{g_m}\omega 1 \tag{3.21}$$

If it is assumed that ω_1 does not change, the maximum slew rate depends on the ratio

$$\frac{I_d}{g_m} \tag{3.22}$$

which for a MOS transistor is proportional to

$$\sqrt{I_d / k}$$

or

$$V_g - V_t$$

The above expression leads to the conclusion that increasing the first stage current will lead to a higher slew rate. However the $V_g - V_t$ cannot be increased beyond certain bound because other performance parameters such as the common mode range will be affected.

Alternatively, a class AB input stage can be used which does not limit the output current even with a large input voltage. An example of such a stage is shown in Fig. 3.12.[19]. This is a CMOS realization of a known bipolar input circuit. In this circuit, the input n-channel and p-channel transistors are connected at their sources and the bias point is set by diode connected p-channel transistors and current sources. The current available for charging the compensation capacitor is therefore not limited by a tail

current source as in the case of the conventional differential pair input stage. This advantage is obtained at the cost of higher offset voltage, increased power consumption and higher input noise. Interestingly enough the operation of this stage corresponds to that of a class AB amplifier.

The same principles can be applied to increase the slew rate of a single stage operational amplifier. Notice that the folded cascode configuration cannot be used as the upper current sources ($I_1 I_2$ in Fig. 3.9) would limit the current drive from the top. Instead the current can be mirrored to the output as shown in Fig. 3.13. The current mirrors at the bottom are necessary only for the fully differential implementation. For a single ended implementation, one of the signal path can be reflected by means of a current mirror to supply the complementary signal at the output.

3.4. Design of Operational Amplifier for the Converter

Our previous discussions were centered around the general question of how to circumvent certain difficulties found typically in MOS operational amplifier designs. Based upon this knowledge, we want to come to the decision of choosing a particular operational amplifier design for the purpose of building an algorithmic A/D converter. Before doing so, we shall briefly reiterate the boundary conditions for the design. In order to achieve better power supply rejection and cancellation of charge injection effect from MOS transistor switches, a fully differential operational amplifier design is employed. The operational amplifier must have a large or unlimited slew rate as the signal inside the analog loop can vary from the minimum to the maximum value within one clock cycle. Furthermore high gain and large bandwidth are required to facilitate complete charge transfer and faster conversion rate.

Under these constraints, the fully differential class AB single stage configuration [20] seems to be the logical choice, mainly because it can provide high gain and fast settling for large signals. For a class A MOS operational amplifier in general, gain and slew rate must be compromised. This is because the gain increases as the quiescent current in the gain stage decreases but the slew rate decreases. On the other hand, the class AB stage draws a very small current in its "equilibrium" position, but is capable of delivering a large current with a large overdrive. The class AB stage thus possesses the advantage in that it provides for variable gain and therefore slew rate for different input differential voltages. In analyzing the settling time for switched capacitor filters, it has been indicated that the slewing period of the voltage waveform can be comparable or larger than the time period spent in the small signal mode for a class A operational amplifier. By using the class AB configuration, the settling time can be shortened as much as 4 times. The single stage configuration is chosen because it is particularly suitable for implementing the class AB operation. In addition, it simplifies the compensation scheme especially when the load consists of a small capacitor as in the case of the ratio-independent algorithmic A/D converter.

3.4.1. Circuit Description

The operational amplifier consists of two parts, the forward amplifier and the common mode feedback circuit. The forward amplifier is shown in Fig. 3.14. It consists of the input transistor pairs M_1 to M_4 which are cross-coupled at their sources. The input differential voltage V_{in+} and V_{in-} are fed to the gates of the n-channel transistors M_1 and M_2 directly and to the gates of the p-channel transistors M_3 and M_4 through the source followers M_5 and M_6. Transistors M_7 and M_8 serve to bias the p-channel transistors M_3 and M_4. The transconductance of the first stage is

$$g_{mdiff} = g_{mn}\, g'_{mp} / g_{mn} + g'_{mp} \qquad (3.23)$$

where

g_{mn} : transconductance of the input n-channel transistor g'_{mp} : reduced transconductance of the input p-channel transistor.

Here the transconductance of the p-channel transistor has been reduced by the transfer characteristics fo the source follower

$$g'_{mp} = g_{mp} \frac{g_{mn\,follower}}{g_{mn\,follower} + g_{mnb\,follower} + g_0}$$

where

g_{mp} : transconductance of the p-channel transistor

$g_{mn\,follower}$: transconductance of the follower transistor

$g_{mnb\,follower}$: backgate transconductance of the follower transistor

g_0 : conductance of the follower current source.

This current is then reflected to the output through the four current mirrors.

The output stage consists of double cascode transistors. At a bias current level of $40\mu a$ and $\frac{W}{L}$ of about 100, the impedance achieved at the output node is about $200M\ \Omega$ Now the additional n-channel devices with their gates attached to ground serves to minimize the effect of impact ionization substrate on the impedance level at the output node. The cascode-transistors have uniform gate lengths close to $5\mu m$. Quantitatively, the output node impedance is found to be

$$r_{out} = r_{nout} \,//\, r_{pout} \tag{3.24}$$

where r_{nout} and r_{pout} are the output impedances of the n and the p-channel current sources respectively. The output impedance of the n-channel current source is given by

$$r_{nout} = (g_{mncas\,1} r_{oncas\,1} r_{oncas\,2}) // r_{oncs} \tag{3.25}$$

where

g_{mncasN} : transconductance of the cascode transistor N

r_{oncasN} : output impedance of the cascode transistor N

r_{oncs} : output impedance of the current source

The backgate transconductance of these transistors has been neglected. Thus the overall gain can be given by

$$A = g_{mdiff} r_{out} \tag{3.26}$$

In this design, the input stage has been modified into a form such that equal currents flow in the n- and the p-channel transistors in contrast to the class AB stage that was introduced earlier. Since the operational amplifier is used predominantly in a inverted mode, no design measure has been taken to extend its input common mode voltage range.

3.4.2. Common Mode Feedback Stage Design

3.4.2.1. A Conventional Design

Fig. 3.15 shows a common mode feedback circuit using two source followers as voltage divider for deriving the common mode feedback signal[21]. This circuit works mainly by assuming that the MOS transistors behave as linear devices which is only roughly true when we linearize the V-I characteristics of the devices around their operating point and when they have a small $\frac{W}{L}$ ratio and are run at relatively large current levels. Normally because the MOS transistors are approximately square law devices, one common problem of this common mode feedback stage design is that even though the differential signal may be perfectly symmetrical, the voltage at the common mode feedback point, labeled P_{cm}, varies.

3.4.2.2. Balanced Common Mode Feedback Circuit Design

In the design that is used with the forward amplifier in this A/D converter, we use two differential stages to cancel the inherent nonlinearity in the transfer characteristics. A simplified schematic of the circuit is shown in Fig. 3.16. It consists of two differential pairs with the drains of the middle transistors M_2 and M_3 connected together. The gates of these transistors are connected to the output of the forward operational amplifier. The gates of the other two transistors M_1 and M_4 are connected to ground to give the common mode voltage a reference. In this diagram the drain currents of M_1 and M_4 are not used and bled to ground. But they can be summed to provide the complementary common mode signal if needed. If we consider this circuit a V-I converter, then the sum of the currents I_{d2} and I_{d3} can be constant only if the two input differential signal is equal and opposite in sign. Or in other words

$$V_{out+} = -V_{out-} \tag{3.27}$$

This can be explained with the V-I transfer characteristics of the individual differential

stage as follows. The transfer characteristics of a differential pair is shown in Fig. 3.17. Using a first order model for the MOS transistors and neglecting the body bias effect, the curve can be given approximately by the equation

$$\Delta I_d = k\,\Delta V_i \sqrt{(2I_{ss}/k) - (\Delta V_i)^2} \tag{3.28}$$

where

$$k = \mu_n C_{ox} \frac{W}{2L}$$

is a constant relating to the device. In the presence of other nonidealities such as body bias effect etc, the shape of the transfer curve will change, but it will still be symmetrical with respect to the origin if the devices are assumed to matched perfectly. This is the basis for the operation of the balanced CMFB stage. Referring back to Fig. 3.16, the two plots on the bottom indicate the V-I characteristics of each of the differential pairs. Since the pairs have the same ($\frac{W}{L}$) ratios, these characteristics are the same. Now a positive voltage change on the gate of M_2 for example would have shifted the operating point of the left differential pair to a point P_+ on the V-I curve, yielding the current δI_+ correspondingly. The purpose of the common mode feedback circuit is to guarantee that the common mode signal be equal to ground. In other words, the common mode feedback signal should be 0 if the differential signals are equal but opposite in sign. Let us investigate what is the condition for this to be true given that one signal is already at V_+. The condition that this is true is synonymous with having the combined current equal to its quiescent value or the current I_{d3} must be complementary to I_{d2}. This is true if the right differential pair were to find itself at a operating point on the curve at equal but opposite distance from the origin due to symmetry. Therefore the circuit ensures a first order cancellation of the nonlinearity.

In the practical realization of the above circuit, the drain currents in the pairs $M_1 M_4$ and $M_2 M_3$ are summed and redistributed by means of cascode transistors M_5—M_8. The redistributed currents are then fed to the lower n-channel current mirrors in the forward amplifier to achieve the negative feedback action. The transistors have to be operated at a high $V_g - V_t$ in order to guarantee that they not be driven into the current limit region.

3.4.3. Complete Operational Amplifier

Fig. 3.18 shows the schematic of the complete operational amplifier. The common mode feedback signal is fed to the lower n-channel current mirrors. The bias voltages bias1, bias2, bias3 and bias4 are generated by a seperate bias generator circuit. For the supply voltage of +/-5V, the simulated dc gain of the amplifier with $40\mu a$ bias current in each branch is 106dB. A Bode plot of its open loop frequency response is given in Fig. 3.19 where it can be seen that there is a phase margin of about 65 degree with a compensation load capacitor of 4 pF. The sizes of the transistors are given in table 3.1. Use of $5\mu m$ gate length is a compromise between the lowering of output impedance and improvement of the transistor frequency response with decreasing gate length.

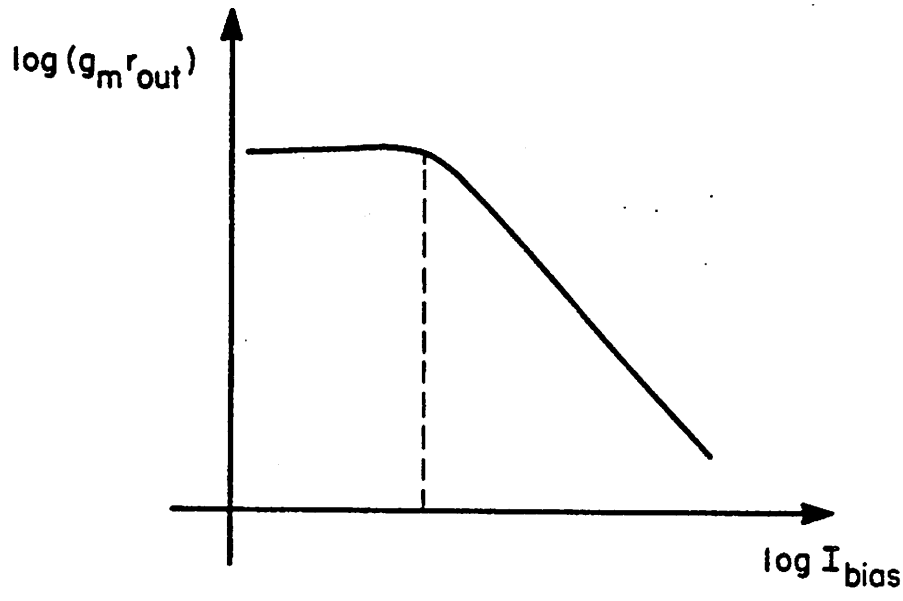


Figure 3.1

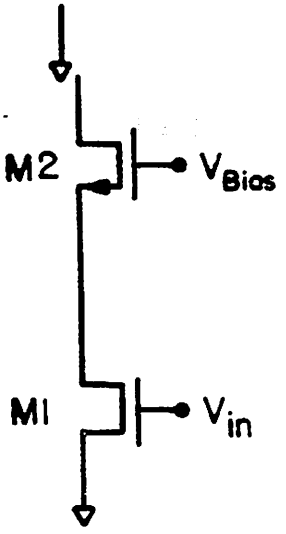


Figure 3.2

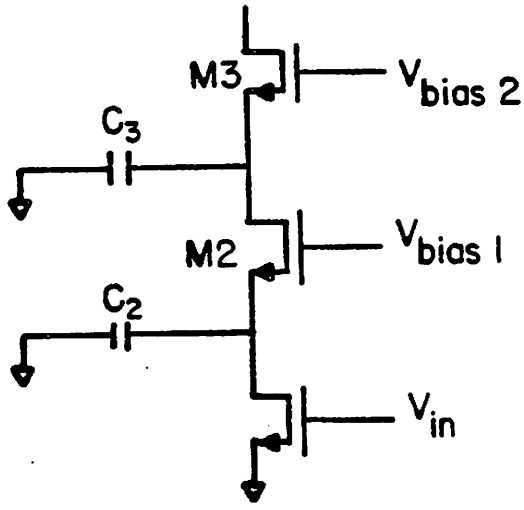


Figure 3.3

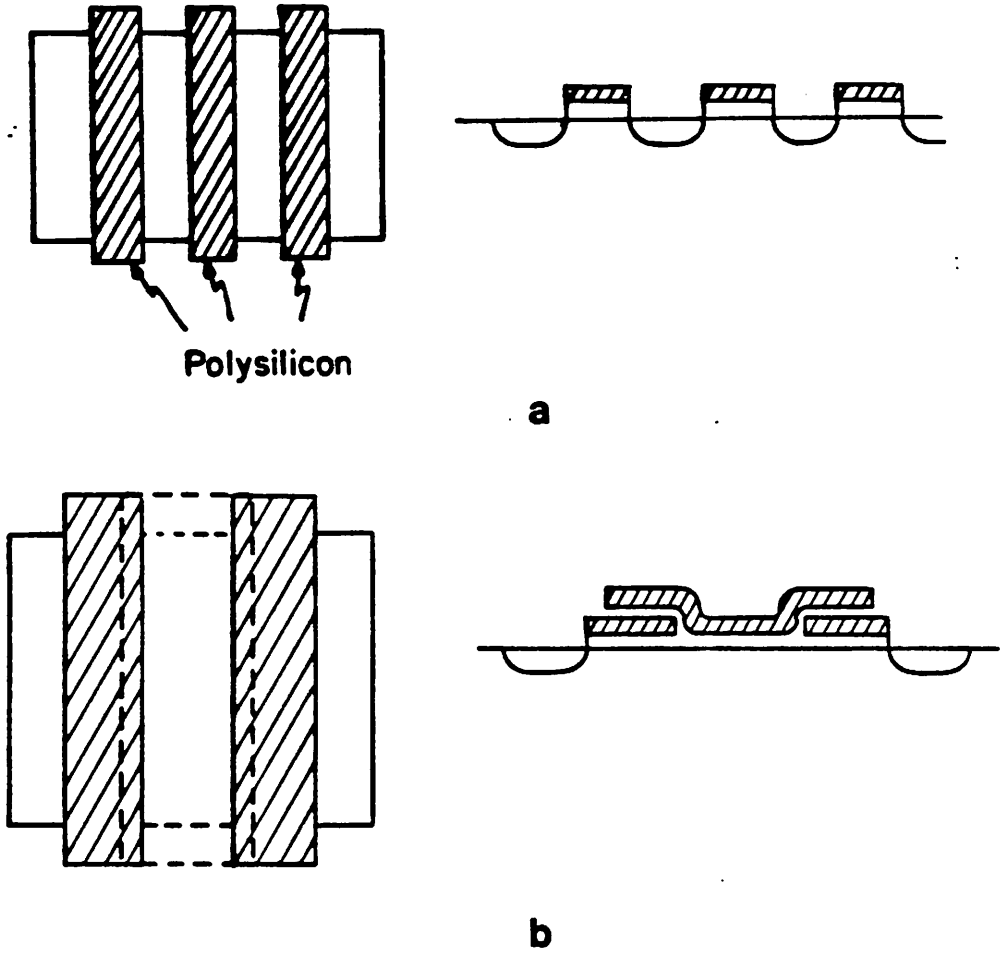


Figure 3.4

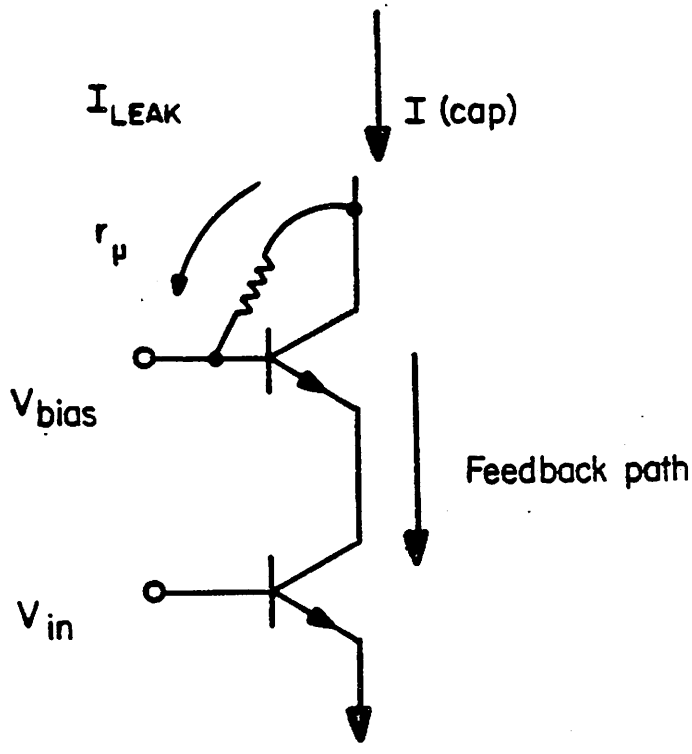
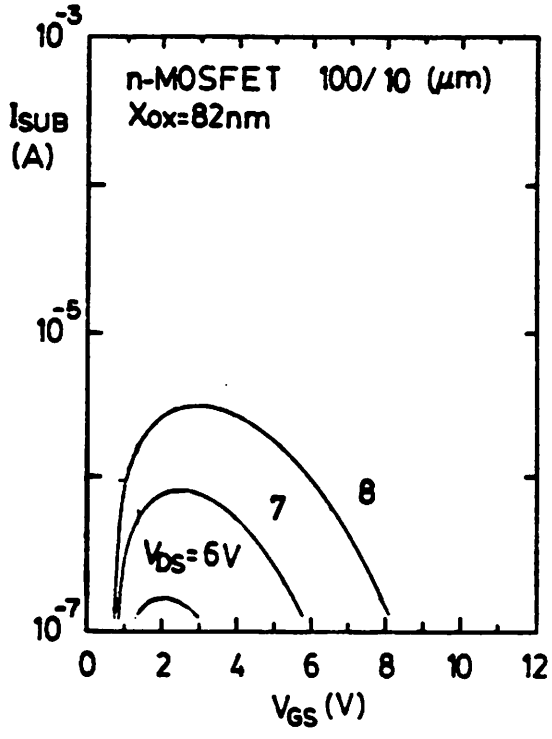


Figure 3.5

a


n-MOSFET 100/10 (μm)
Xox=82nm
ISUB (A)
10^-3
10^-5
10^-7
8
7
VDS=6V
0 2 4 6 8 10 12
VGS (V)

b

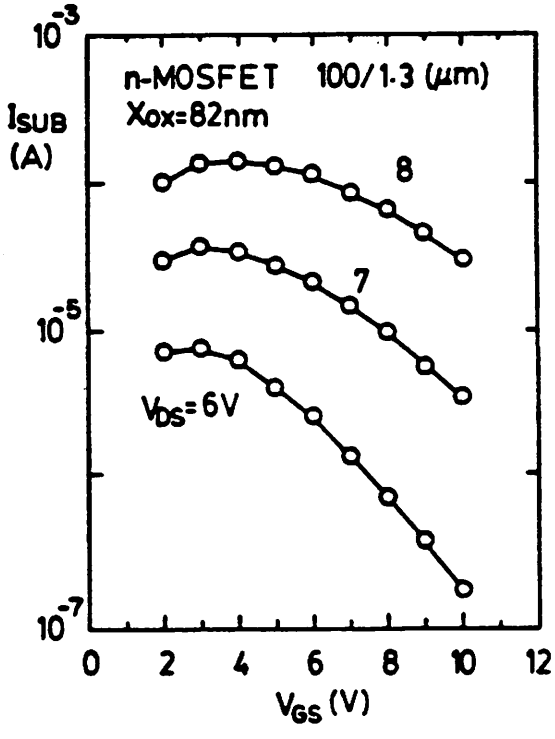

n-MOSFET 100/1.3 (μm)
Xox=82nm
ISUB (A)
10^-3
10^-5
10^-7
8
7
VDS=6V
0 2 4 6 8 10 12
VGS (V)

Figure 3.6

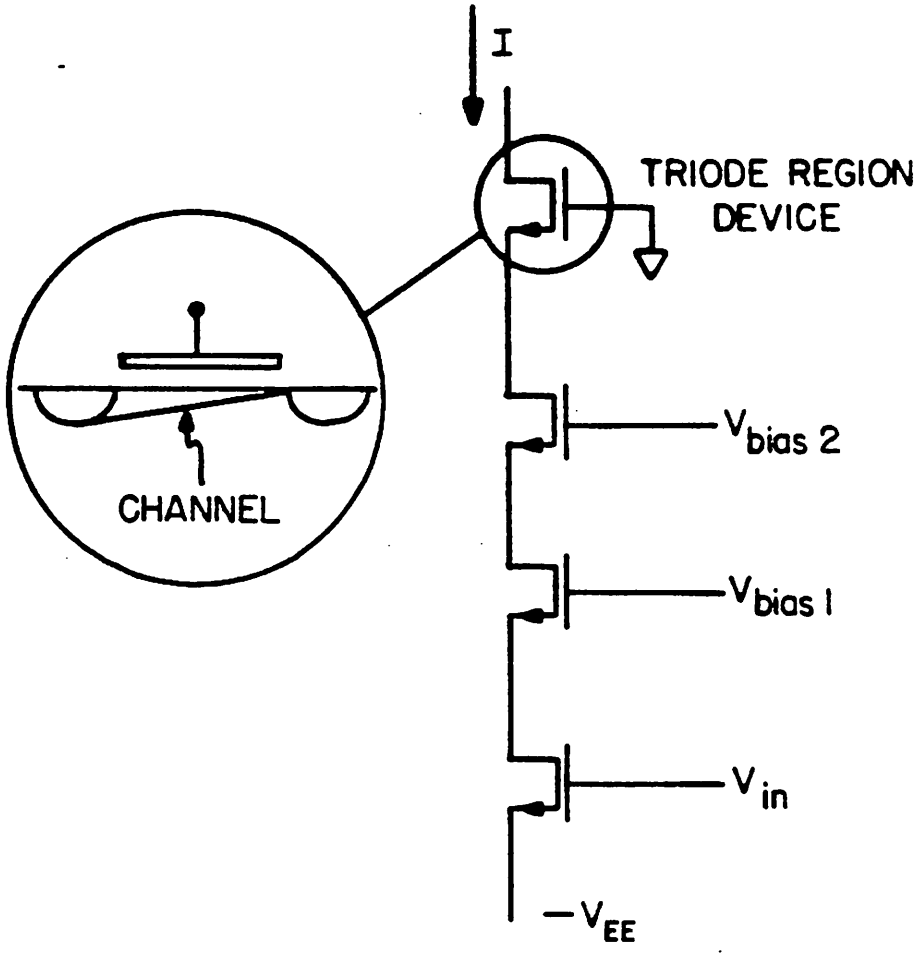


Figure 3.7

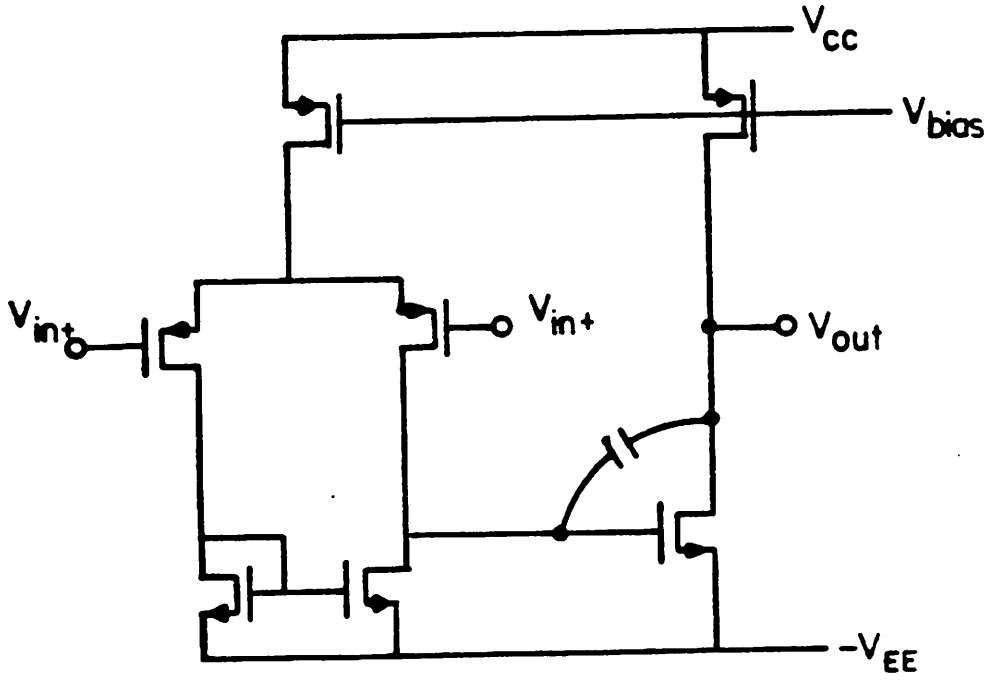


Figure 3.8

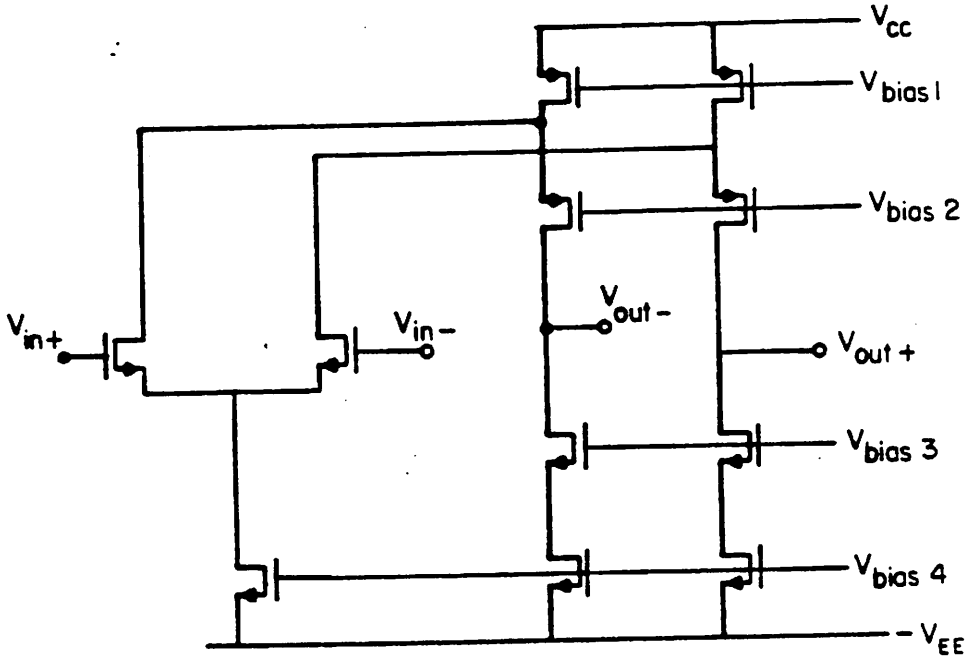


Figure 3.9

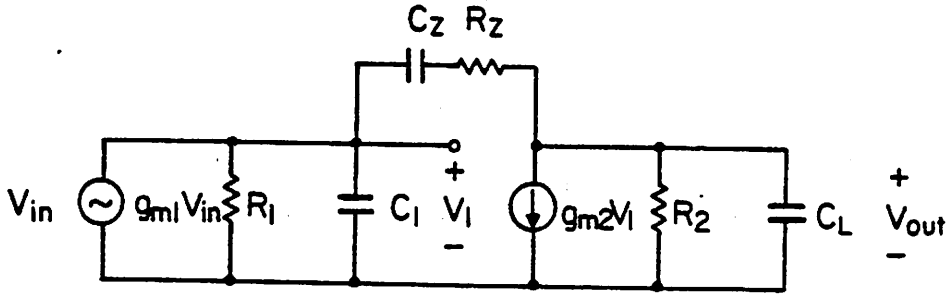


Figure 3.10

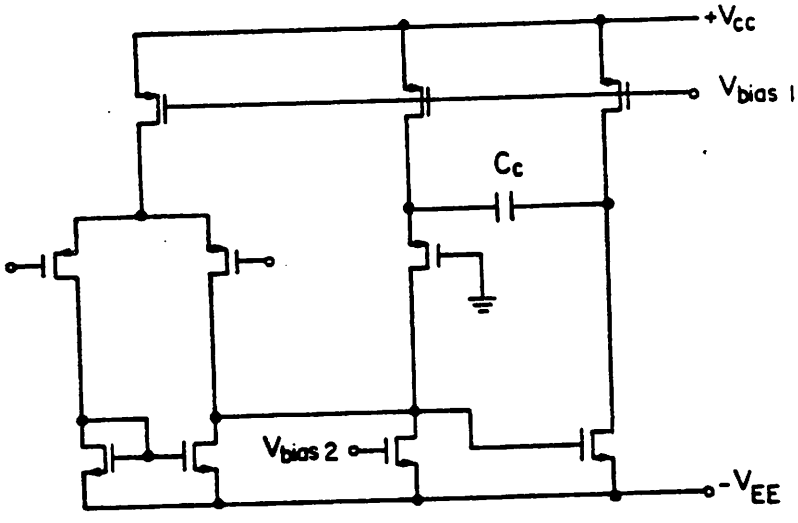


Figure 3.11

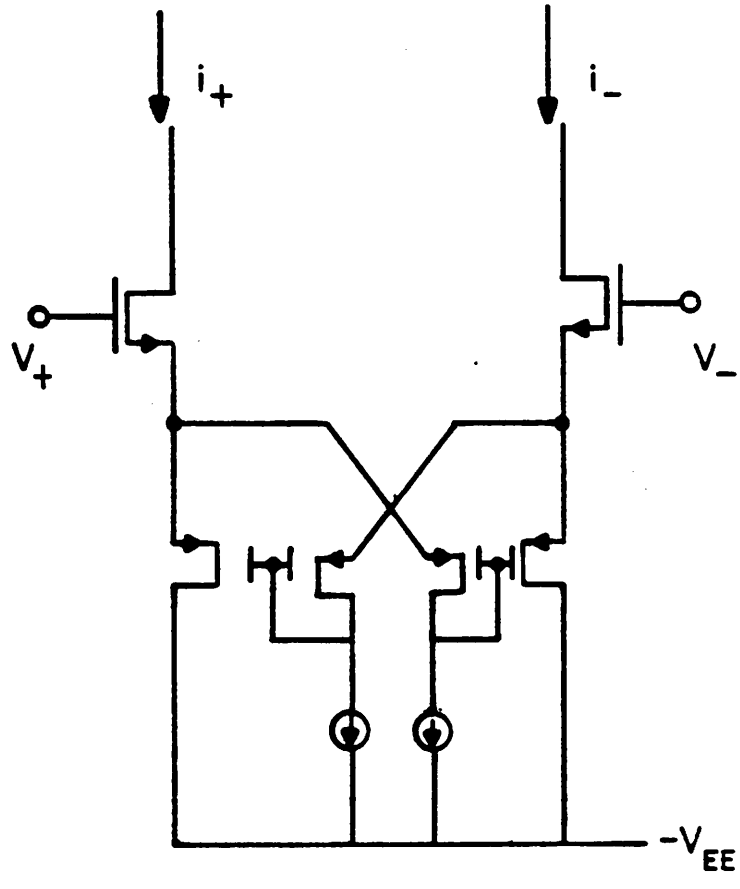


Figure 3.12

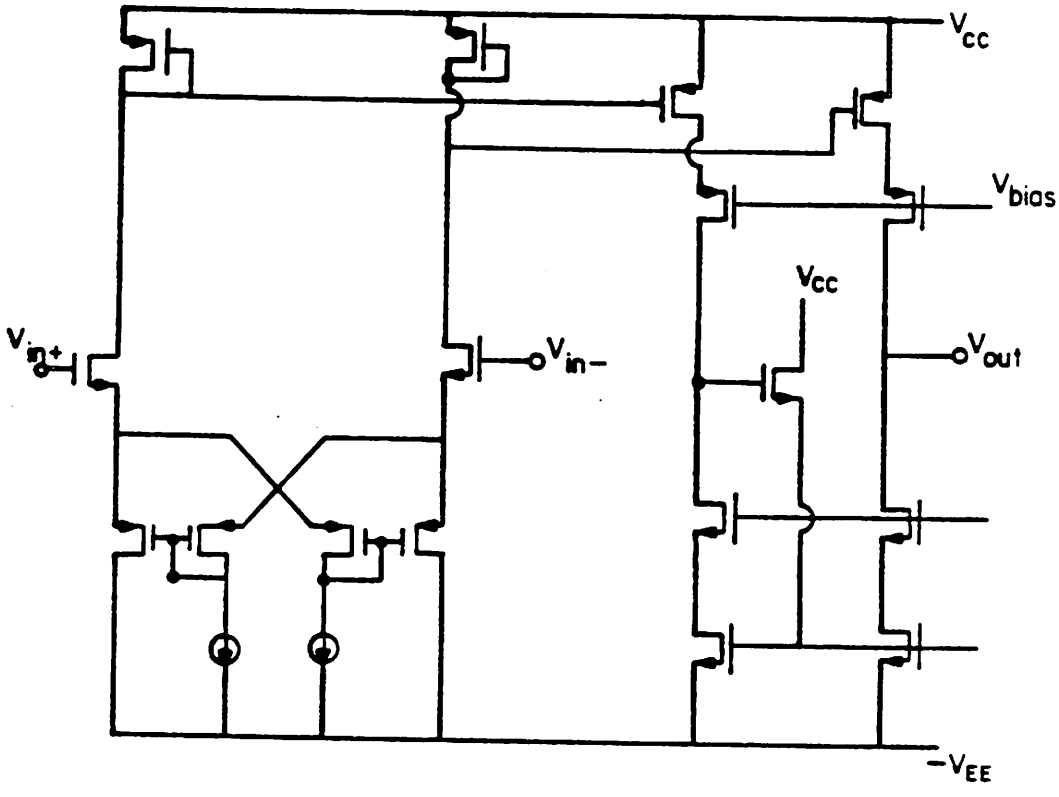


Figure 3.13

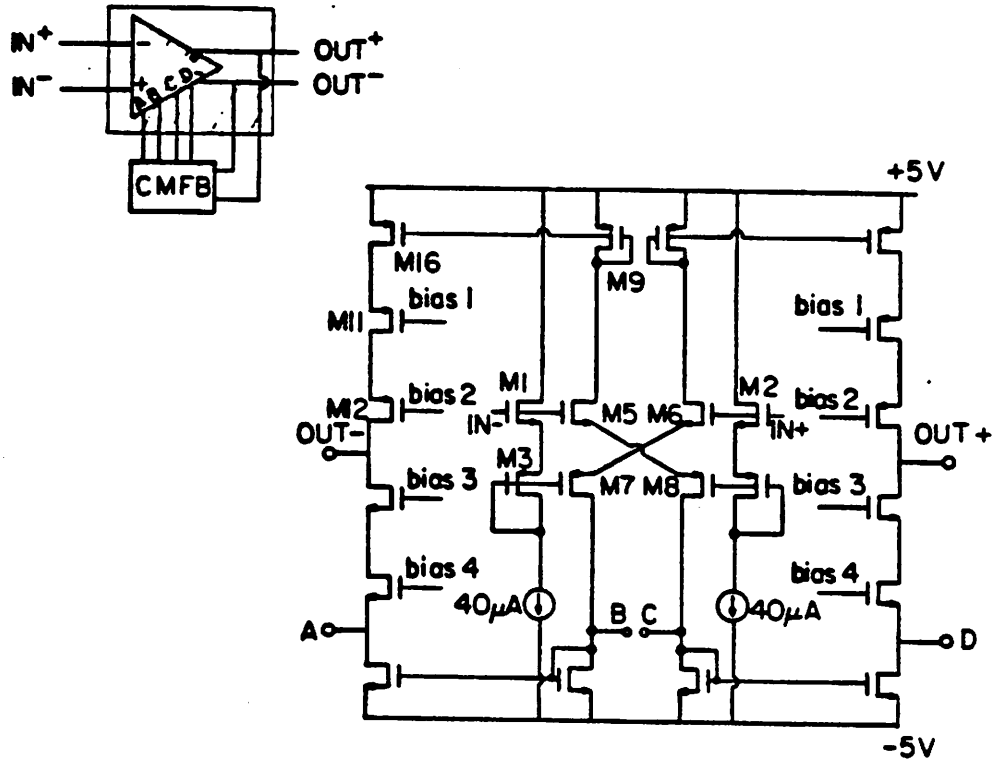

IN+
IN-
OUT+
OUT-
CMFB
+5V
M16
M9
bias 1
M11
bias 1
M1
M2
M2
bias 2
IN-
M5 M6
IN+
bias 2
OUT-
OUT+
M3
bias 3
M7 M8
bias 3
bias 4
40μA
B C
40μA
bias 4
A
D
-5V

Figure 3.14

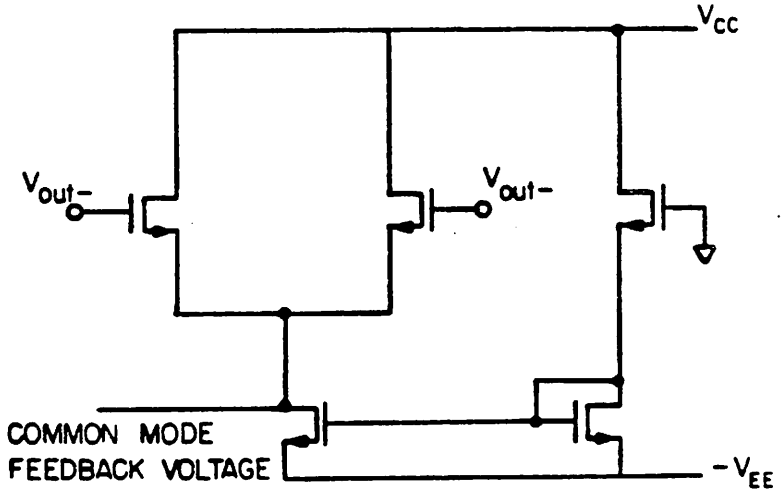


Figure 3.15

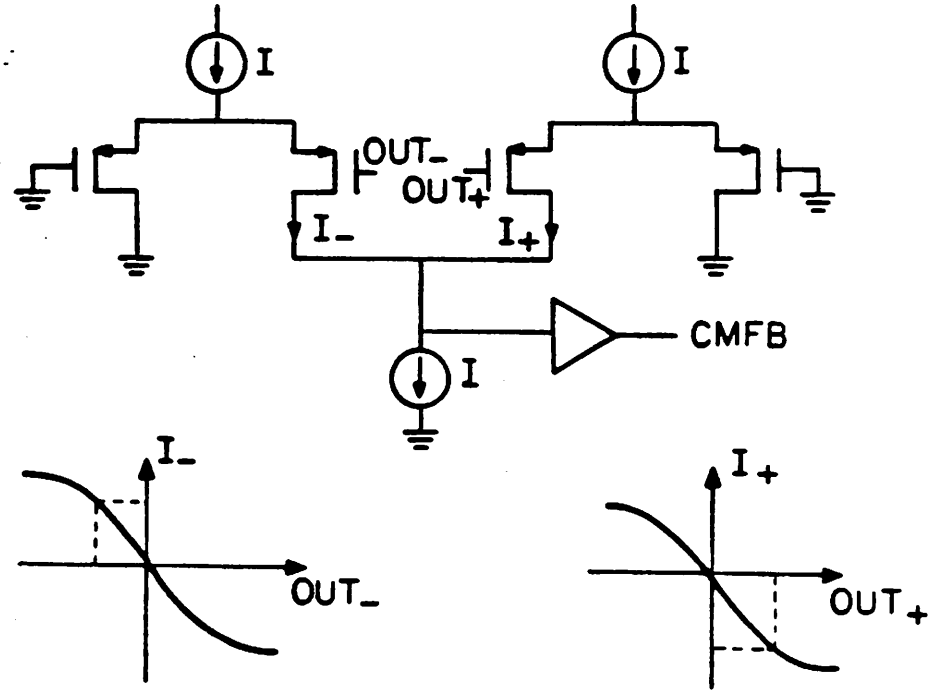
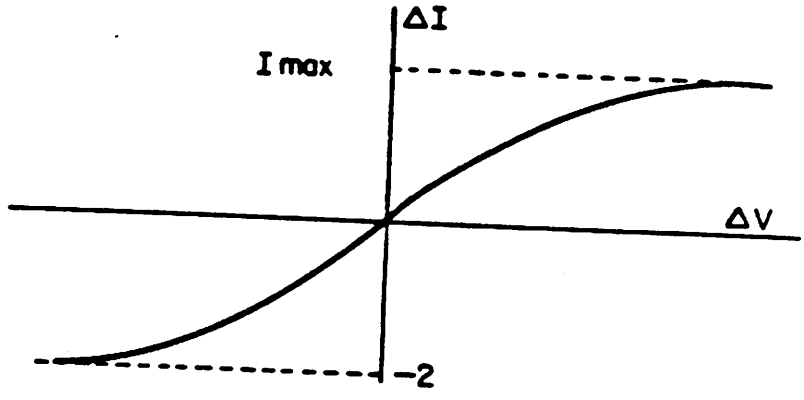


Figure 3.16



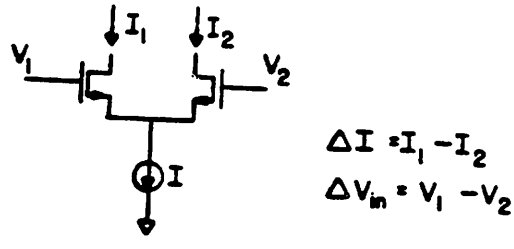


Figure 3.17

SCHEMATIC OF OP AMP

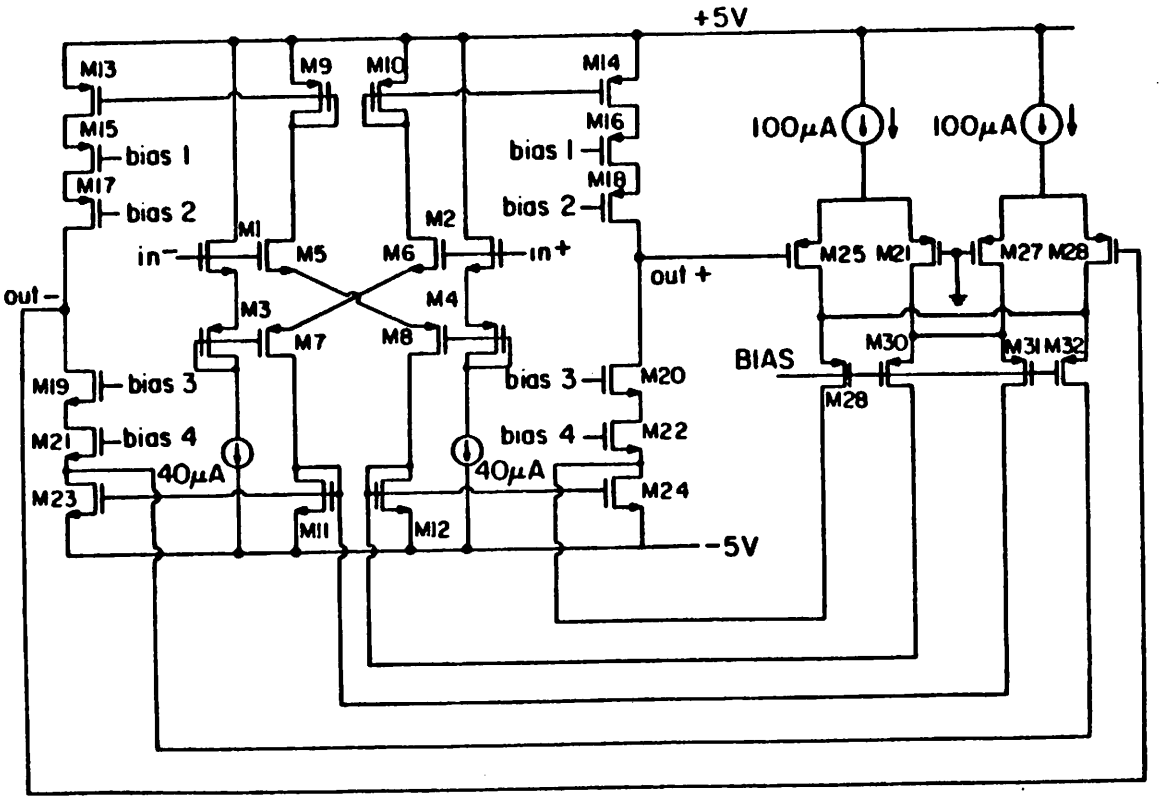


Figure 3.18

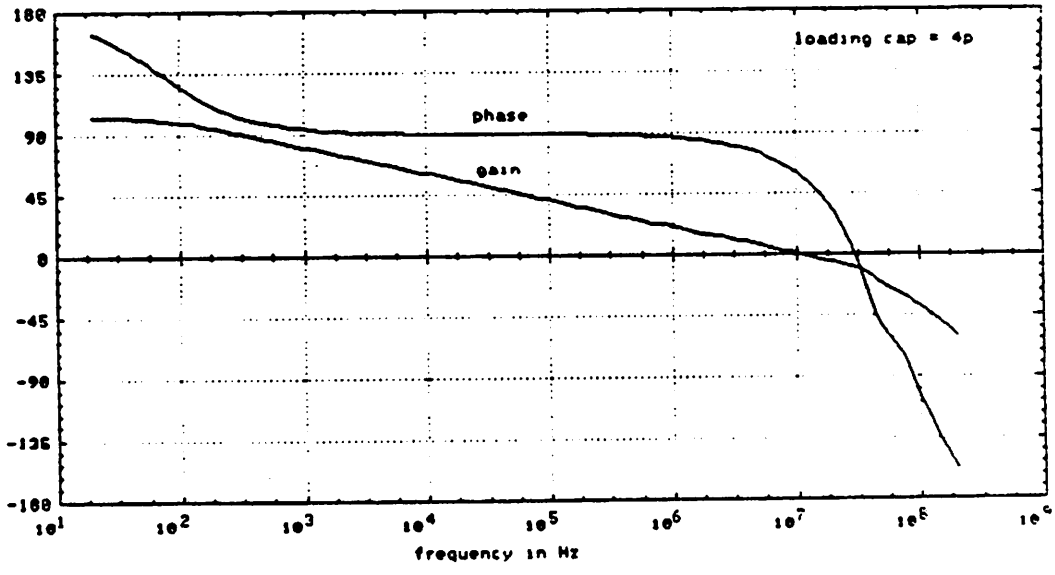

180
135
90
45
0
-45
-90
-135
-180
loading cap = 4p
phase
gain
10^1
10^2
10^3
10^4
10^5
10^6
10^7
10^8
10^9
frequency in Hz

Figure 3.19

	W (μm)	L (μm)		W (μm)	L (μm)
M1	100	5	M17	120	5
M2	100	5	M18	120	5
M3	100	6	M19	80	5
M4	100	6	M20	80	5
M5	100	5	M21	80	5
M6	100	5	M22	80	5
M7	100	6	M23	80	5
M8	100	6	M24	80	5
M9	120	5	M25	10	20
M10	120	5	M26	10	20
M11	80	5	M27	10	20
M12	80	5	M28	10	20
M13	120	5	M29	30	5
M14	120	5	M30	30	5
M15	120	5	M31	30	5
M16	120	5	M32	30	5

Table 3.1

CHAPTER 4

Organization and Operation of the Algorithmic A/D Converter

4.1. Introduction

In this chapter, the organization of the analog building blocks to form the ratio-independent algorithmic analog to digital converter is discussed. The organization of this converter differs from that of a conventional algorithmic A/D converter in that it uses many switches to put it into different configurations. The ratio-independent algorithm is carried out at the cost of extra clock cycles. It represents a case where accuracy is traded against time by employing special hardware features. In fact the inaccuracy in the matching of the capacitors is corrected during every cycle of the conversion. As such it can also be called a self-correction method of analog to digital conversion.

Following this, the external logic circuitry which generates the control sequence is described. The control of the converter involves the simultaneous control of 23 clock lines, some of which have common switching during a part of the repetitive control timing scheme. One particular feature of the control logic is that for certain signal there is a fixed delay. In a single chip implementation the control can be integrated as a ROM with a single bit address feedback. It is estimated with a $5\mu m$ CMOS technology, the digital control circuitry should occupy less than 30% of the total analog circuit area.

Finally some factors which limit the accuracy of the algorithmic analog to digital converter are discussed. Offset and noise, both the kT/C as well as the 1/f noise, can immensely affect the overall accuracy of the system. Solution to canceling the 1/f noise

consists in applying a path-reversing algorithm to the A/D conversion process, utilizing the full advantage of the differential implementation of this particular converter. The result obtained can be compared to the double correlated sampling technique used in CCD transversal filters. The kT/C noise however is fundamentally a wideband noise and can be reduced only by increasing the size of the capacitors.

4.2. Description of the Converter Circuit

A circuit schematic of the algorithmic A/D converter is shown in Fig. 4.1. It consists of 3 operational amplifiers, 8 capacitors and 35 switching transistors. They are arranged to form the analog loop and the comparator. The transistors are all of the same size and have a uniform drawn channel length of $5\mu m$ and a width of $12\mu m$. The origin intentions in applying a larger transistor width are both for lower charging time for the capacitor and better matching between the two neighboring signal paths, with the latter reason perhaps more relevant than the former. Six of the capacitors and two of the operational amplifiers are committed to perform the sample/hold function and the ratio-independent multiply-by-two function. Their operations have been described in detail earlier in chapter 3. The other two capacitors and the operational amplifier are used as an AC coupled comparator.

4.3. Operation of the Converter

To see how an analog to digital conversion can be accomplished with these building blocks, we trace through the individual steps of a conversion using the restoring algorithm as an example. All the analog circuit operation occurs within unit time steps. The start of each step is synchronized to the edge of a clock and takes place in half the clock

period. The clock rate is therefore determined by the slowest process that can take place in a conversion, which is usually the operational amplifier settling time.

At the beginning of the conversion, the input signal is selected through the multiplexing switches M_1 and M_2 and is sampled on the sample/hold amplifier. This is shown in Fig. 4.2. For the initialization of the sample/hold amplifier, switches M_3 and M_4 are turned on, thus putting the operational amplifier into a unity gain feedback configuration. During this time, the multiply-by-two amplifier can be initialized. This constitutes the first half clock cycle.

In the second half clock cycle, the sample and hold operation is completed by connecting the capacitors C_1 and C_2 to the output of operational amplifier 1 with M_5 and M_6. The X2 operational amplifier is put into the unity gain feedback configuration and the sampling capacitors C_5 and C_6 are then connected so as to take in the first unit of signal charge for the multiplication. This is accomplished by turning on M_{13} and M_{14}. The capacitors for intermediate storage C_7 and C_8 are connected to ground to store the offset voltage of the operational amplifier. For convenience, the reference subtraction capacitors C_3 and C_4 can be initialized to the position such that reference subtraction or addition can be done immediately during the charge transfer phase. Usually this is possible during a conversion cycle since the sign of the loop signal from the previous conversion cycle is known. However for the first cycle, the sign of the input signal can only be tested by establishing a path through transistor pairs M_1M_2–$M_{32}M_{33}$–$M_{26}M_{27}$. If this is found to be problem for the control, the sign of the loop signal can alternatively be obtained at the end of the first charge transfer cycle where only the magnitude but not the sign of the loop signal is distorted by the capacitor ratio.

After this charge sampling, the signal would then be transferred to the capacitors C_7 and C_8. As mentioned before, the sequence of switching off the transistors is very

important in order to avoid signal dependent charge injection. In this case switches M_{18} and M_{19} must be turned off first. Then switch M_{15} is turned on. Connecting the two plates of C_5 and C_6 together instead of connecting them directly to ground avoids incurring a common mode signal on the input of the operational amplifier. Thus in the third half clock the charge is transfered while the sample/hold amplifier is holding the signal ready for the second sampling and transfer of charge.

In the fourth half clock cycle, the input sampling capacitors C_5 and C_6 take in another sample of the signal. At the same time the storage capacitors C_7 and C_8 must be disconnected from the operational amplifier so that the latter can be reset. For this purpose the switches M_{20} and M_{21} are provided. At the end of the third half clock cycle, these switches are first turned off to disconnect the storage capacitors from the inputs of the operational amplifier. The reason why these transistors instead of the switches M_{24} and M_{25} are use to separate the capacitors is again based on charge injection argument. We can imagine that after a charge transfer the output of the operational amplifier and similarly the switches M_{24} and M_{25} are at different potentials. Thus if these switches are turned off there would be a signal dependent charge injection component introduced into the capacitors C_7 and C_8. This occurs through the current path $M_{20}M_{18}$ because M_{18} is to be turned on immediately after M_{24} is turned off.

In the fifth half clock cycle, C_5 and C_6 are connected directly to the output of the operational amplifier in much the same way as it is done in the sample/hold amplifier. Simultaneously the storage capacitors are connected back to the input of the operational amplifier with their other plates grounded. The negative feedback of the operational amplifier forces the charges which had resided on C_7C_8 to redistribute themselves onto the input sampling capacitors, thereby completing the multiply-by-two process. Now depending on the sign of the signal, the reference capacitors C_3 C_4 would have been set

up for the addition or subtraction of the reference voltage. This is done by either connecting these capacitor to the reference voltages or ground when the operational amplifier is reset. Subsequently by switching them to the other voltages, reference addition or subtraction could be accomplished. Fig. 4.7a-b shows the operation of subtracting the reference. This reference addition or subtraction operation is usually incorporated into the fifth half clock cycle. A comparator decision is also made at the end of this cycle for the restoring algorithm. Then in the sixth half clock cycle the signal is restored or kept depending on the outcome of the comparator decision. This is achieved by reversing the switching sequence or doing nothing at all. During this time, the sample/hold amplifier is reset and is taking in the next signal for feeding back into the loop. Thus this process would go on until all the digital information has been extracted.

The comparator is set up to compare a differential signal to zero. Before the start of a comparison, the ac-coupled comparator is be reset to store the offset voltage. This is done in the following manner (Fig. 4.8a-b). First the operational amplifier is nulled and the offset voltage stored on the capacitors which act to frequency compensate the closed loop operational amplifier. Following this the operational amplifier is operated open-looped and drives a regenerative latch which has a small input capacitance. Minimal capacitance loading on the output node of the operational amplifier maximizes its bandwidth and therefore the response time.

4.4. Control Logic for the Ratio-Independent Algorithmic A/D Converter

4.4.1. Description

Ideally the converter traverses through three "macrostates". They are characterized in Fig. 4.9. The first state which is denoted S_0 is the idle state. From here the converter

receives its **START** signal to start the conversion. Once started it moves to the second state S_1 in which it samples the input signal. After this it goes to the third state, in which the conversion is continued by recycling the residual signal through the loop and performing the A/D conversion. The condition for exiting this state is given by a counter which counts the number of bits derived. With the signal **END CYCLE** the converter reverts back to state S_0 where it waits for the next **START** signal. Depending on the particular algorithm used for the conversion, each macrostate may contain a sequence which takes a finite number of clock cycles to complete.

4.4.2. Logic Implementation

The control logic can of course be realized using random logic but it is best realized with a combination of counters and a ROM as shown in Fig. 4.10.

The control sequencer consists of an address counter, a cycle counter, a multiplexer and a ROM. Its architecture sequencer is very similar to that of a microprogram controller. The address counter is advanced by the clock **CLK** and it can be preset to the **NEW ADDRESS** by means of a **LOAD** signal which causes the **NEW ADDRESS** to be loaded with the next incoming clock. The address can be 4 bits wide and is sufficient to decode the control sequence for a restoring or non-restoring A/D conversion algorithm. The **ROM** decodes this address and output the control data to a latch which latches this data with **CLK**. The data path has a width of 26 bits. Of these 26 bits, 19 are used for the control of the control of the A/D converter, 4 for the **NEW ADDRESS** field, 1 for clocking the cycle counter and the remaining 2 for selecting the multiplexer. The cycle counter is advanced by the control line **CYCLK**. It is used to count the number of times the analog signal had passed through the converter loop. Both the counters can be reset by the **CLR** signal. The multiplexer selects one of the four signals to the output as the

LOAD signal. These signals are

1) END

2) COMPARATOR

3) JUMP

4) NO-JUMP

The first two signals, when selected, force a conditional jump to the **NEW ADDRESS. END** is the carry output of the counter and it is true when the cycle counter has finished the /12 count. Simultaneously it inhibits the clock input to the address counter and freezes the output of the ROM. **COMPARATOR** is the comparator output. The next two signals **JUMP** or **NO-JUMP** force the address counter to jump to the new address or to continue on to the next address unconditionally.

Thus we see that the input clock to the address counter would be inhibited once the desired number of bits have been reached. The sequencer or A/D converter can be restarted by clearing the counters using the **CLR** input. The output control lines from the latch can selectively delayed by inserting inverters. Since only the transistor pairs $M_3M_4M_{18}M_{19}$ and $M_{20}M_{21}$ have to be turned off *prior* to the others. All *other* lines must be delayed with respect to these three control lines. Table 4.1 shows an example of encoding a restoring A/D algorithm. Table 4.2 shows the correspondence between the transistors and the control lines. It can be seen that the ratio-independent algorithmic A/D converter can be controlled with a small set of control codes. With a $5\mu m$ CMOS technology, it is estimated that all this logic can be implemented in less than $800 mil^2$ of area.

4.5. Noise in the Ratio-Independent Algorithmic A/D Converter

In previous chapters we have discussed the effects of the loop offset and loop gain on the accuracy of the algorithmic converter. The basic underlying assumption is that the analog loop is noise free. In this section the noise property of the converter is discussed briefly.

4.5.1. Sources of Noise in Algorithmic A/D Converter

The most important sources of noise in the algorithmic A/D converter are the noise of the operational amplifier and the noise from the MOS switches. They are identical to those existing in the switched capacitor filter and therefore the analysis of the noise performance is similar. In the following discussions we are first concerned with the thermal noise in the MOS switches which has become widely known as the $\frac{kT}{C}$ noise. It represents a fundamental limit to the performance of any switched capacitor circuit and therefore gives a upper bound to the resolution of the algorithmic A/D converter. Next we discuss the $\frac{1}{f}$ or Flicker noise in MOS integrated circuits. A Double Correlated Sampling Technique has evolved recently that allows a frequency dependent cancellation of the $\frac{1}{f}$ noise. A variant of this technique is studied here and it is applicable to the differential channel algorithmic A/D converter. In all the subsequent discussions we use the term noise to designate both the expected noise variance or the actual noise voltage. However it will become obvious out of the context as to exactly each term we are referring to.

4.5.1.1. kT/C Noise in MOS Circuit

When a capacitor is connected to a voltage source through a MOS transistor, the wide band noise which stems from the channel resistance of the transistor is sampled onto the capacitor. Fig. 4.11a depicts the circuit with V_r representing the thermal noise voltage source in the transistor. The equivalent circuit is shown in Fig. 4.11b. The band-limiting effect of the equivalent RC network gives the transfer function

$$H(s)=\frac{1}{(1+sR_{ch}C)}. \tag{4.1}$$

If we assume the channel noise of the MOS transistor to be white then the variance of the noise at the capacitor will be

$$\int_0^\infty 4kTR_{ch}\,|H(j\omega)|^2 d\omega = kT/C \tag{4.2}$$

This noise is independent of the magnitude of the on resistance of the MOS transistor but is only dependent on the absolute temperature T and the capacitance C. It becomes therefore widely referred to the $\frac{kT}{C}$ noise.

4.5.1.2. Operational Amplifier Noise

The MOS operational amplifier has two dominant noise sources. They are the wideband thermal noise and the Flicker noise (1/f). Fig. 4.12 shows the typical noise spectrum of an MOS operational amplifier. The magnitude of the 1/f noise decreases rapidly with increasing frequency until the wideband thermal noise becomes dominant at the corner frequency f_{corner}. This frequency is dependent on the process and may lie between 1kHz to 1MHz. The Flicker noise is highly correlated at low frequencies. This

offers the opportunity of using statistical techniques to cancel this noise.

For a single stage MOS transconductance operational amplifier, the equivalent input noise resistance for the wideband thermal noise is approximately equal to $\frac{2}{3g_m}$ where g_m is the effective transconductance [22]. Henceforth we shall use this model to analyze the noise property of the operational amplifier.

4.5.2. Noise Performance of the Multiply-by-Two and Sample/Hold Amplifier

Since the multiply-by-two amplifier is the basic building of the algorithmic A/D converter, we will examine some of its noise properties. From this analysis we can deduce a first order magnitude of the noise performance of the converter.

The basic circuit used in our analysis is shown in Fig 4.13. This circuit configuration differs from that of a bottom-plate switched capacitor integrator in that during the first sampling phase there is an offset storage operation and that during the second sampling phase the integrating capacitor is disconnected from the operational amplifier. Thus these differences have to be taken into account when analyzing the noise performance of this circuit. In calculating the noise of a sampled data circuit, the following facts are important:

1) In calculating the noise on capacitors, we assume that the noise powers from the same noise source during different sampling periods sum up together. This is a simplifying assumption which may not be necessarily true since the noise from the same source during different times may actually be correlated and not statistically independent.

2) With negative feedback, the noise voltage source of the operational amplifier can be moved unchanged outside of the feedback loop. Also since the bandwidth of the operational amplifier is higher than the f_{corner} of the 1/f noise, we shall talk of an equivalent input thermal noise of the operational amplifier which is essentially white with the 1/f noise averaged out in the frequency band.

3) We shall assume that the operational amplifiers used in the calculations are single pole circuits, *i.e.* the gain-frequency function can be described as

$$A(\omega)=\frac{A_0}{1+\frac{\omega}{\omega_0}} \tag{4.3}$$

where ω_0 is the pole frequency. Also the output of the operational amplifier behaves like an ideal voltage source for frequencies beyond its unity gain frequency.

During the first sampling phase (Fig. 4.14a), the operational amplifier is connected in its unity gain feedback configuration. The equivalent circuit for noise consideration is shown in Fig. 4.14b. The thermal noise of the transistors are represented by the voltage sources v_{nmi}^2. The operational amplifier noise is summarized in the noise source v_{nopamp}^2.

For low frequencies, the input impedance looking into the input of the unity gain operational amplifier is very low and can be neglected. However at higher frequencies, this input impedance increases due to the finite gain bandwidth of the operational amplifier. The implication of this is that at low frequencies, the noise voltages from transistors M_4 and M_5 do not contribute to noise in capacitor C_s and *vice versa* between C_f and M_1. At frequencies near to the unity gain frequency of the operational amplifier, the input impedance of the negative feedback amplifier increases and

eventually is equal to a parallel combination of the input capacitance of the operational amplifier and the channel resistance of M_3 in series with the output impedance of the operational amplifier. In this case the voltage transfer function from v_{nm1}^2 to C_f is not zero. Although analytical solution can be obtained for these and intermediate frequencies, a simplifying assumption is to say that there are only two frequency regions. The first corresponds to the low frequency case where the input is almost equal to a grounding point. The second case corresponds to the very high frequency case where the input has impedance much higher than the capacitances C_s or C_f . This is approximately valid since the input capacitance of the operational amplifier is always lower than the external capacitances. In this case, noise voltages v_{nm1}^2 generates the same voltage on C_s and C_f . Because these noise voltages are in phase, during the subsequent charge transfer phase they cancel each other and produce no effect on the voltage on C_f .

In addition to these, we see that the sampled noise voltages on the capacitors C_s and C_f have one component in common. It consists of the noise of the operational amplifier plus the associated nulling switch M_3. Although this noise is sampled onto the capacitors, it does not affect the value of the output voltage during the charge transfer phase (Fig. 4.15a) because it cause correlated noise voltages on the two equal capacitors. Under this consideration, the sampled noise on the two capacitors are

for C_s : kT/C_s

for C_f : kT/C_f

During the charge transfer phase the circuit is connected as shown in Fig. 4.15a. Noise analysis in this case is more involved since all the noise generators contribute noise to the integrating capacitor C_f . For the noise source v_{nm2}^2, it has been shown that[22] the noise on C_f from v_{nm2}^2 is

$$v_{fm2}^2 = \frac{kT}{C_s}\frac{(1+A_0\omega_0 R_{ch2}C_s}{2+A_0\omega_0 R_{ch2}C_s)} \tag{4.4}$$

Thus depending on the value of $A_0\omega_0 R_{ch2}C_s$ this expression can have different values. For the noise source $v_{nm4}^2 + v_{nm6}^2$, it can be shown that for frequency up to a fraction of the unity gain frequency of the operational amplifier it is attenuated by the gain of the operational amplifier and therefore does not contribute significantly to the total noise. Beyond the unity gain frequency, the transfer function from the noise source to the capacitor is

$$H_{vf}(s) = \frac{1}{(1+\frac{C_f}{C_s}) + s(R_{ch2}+R_{ch4}+R_{ch6})} \tag{4.5}$$

and so the noise is

$$v_{fm4m6}^2 = \int_{\omega_0}^{\infty} 4kT(R_{ch4}+R_{ch6})|H^{2v}{}_f(\omega)|\,d\omega \tag{4.6}$$

Finally there is the noise of the operational amplifier. Suppose R_{teq} is the equivalent thermal resistance of the operational amplifier then this noise is

$$v_{fopamp}^2 = \frac{kT}{C_s}\frac{R_{teq}}{R_{ch2}} \tag{4.7}$$

which is assuming that the bandwidth of the input RC network formed by M_2 and C_s is dominant because the product of $A_0\omega_0$ is extremely large.

During the second charge sampling phase (Fig. 4.16a), the situation is similar to that of the first charge sampling phase except that the capacitor C_f is not connected. The component of kT/C noise on C_f therefore vanishes. Similarly, the final circuit

configuration is identical to that of the first charge transfer phase. The total noise contribution v_{nx2}^2 can be written as the sum of all the noise contributions.

In a similar way the noise contribution of the sample/hold amplifier can be calculated to be $v_{ns/h}^2$. The total input equivalent noise of the converter is equal to

$$v_{ntotal}^2 = 4 (v_{nx2}^2 + v_{ns/h}^2) \tag{4.8}$$

4.5.3. Noise Cancellation Techniques

Unfortunately, the kT/C noise in the capacitors and wideband noise of the operational amplifier cannot be canceled with any known method because they are random in nature. kT/C noise can be reduced by increasing the size of the capacitor. The operational amplifier noise is reduced by using special low noise designs. However since the Flicker or 1/f noise is correlated at low frequencies, we can make use of this property to obtain noise cancellation.

The method consists of reversing the signal path once after the determination of the most significant bit. This is easily done by using the path reversing transistors M_{34} and M_{35} shown in the schematic of the converter. The effect of doing this is as follows. Suppose that the 1/f noise were simply an offset voltage in the loop. Let its magnitude be given by v_{nf} . We consider the effect of passing it through the analog loop. After the first cycle in the loop, this voltage becomes $2v_{nf}$. Since the path is reversed after the first cycle, this voltage would reverse its sign and becomes $-2v_{nf}$. In subsequently cycles, successive v_{nf} is added to this voltage and their result multiplied by 2 and passed back into the loop. Thus the result of the loop voltage v_{iloop} appears as follows

2nd cycle: $v_{2loop} = 2(-2v_{nf} + v_{nf}) = -2v_{nf}$

3rd cycle: $v_{3loop} = 2(-2v_{nf} + v_{nf}) = -2v_{nf}$

nth cycle: $v_{nloop} = 2(-2v_{nf} + v_{nf}) = -2v_{nf}$

Therefore the total loop error voltage due to v_{nf} after n cycle is only $-2v_{nf}$. If we compare this to the result of passing v_{nf} without reversal of path through the analog loop, it can be easily deduced that the result for the latter would be $2^n v_{nf}$ which increases beyond bound as the number of cycle n is increased. If this voltage is referred back to the input. The first case yields an equivalent input voltage of $-\frac{v_{nf}}{2^n - 1}$. The latter yields an equivalent input voltage of $-v_{nf}$.

Applying this principle to the cancellation of 1/f noise, it can be deduced that provided the sampling frequency is high enough, the 1/f noise can be canceled effectively. Fig. 4.17 shows the equivalent z-domain diagram of this cancellation technique. It shows five cycles in the analog loop. Fig. 4.19 shows the frequency response of the converter towards noise based on the structure in Fig. 4.18 for 12 cycles in the analog loop. The equivalent gain is 2^{12}. The sampling frequency has been arbitrarily set at 250kHz. The gain at low frequency is very small and rises to the value of 2^{12} or 72dB as the frequency approaches the sampling frequency. The dip in gain as the frequency approaches the fundamental and harmonics of the sampling frequency is due to harmonic effect.

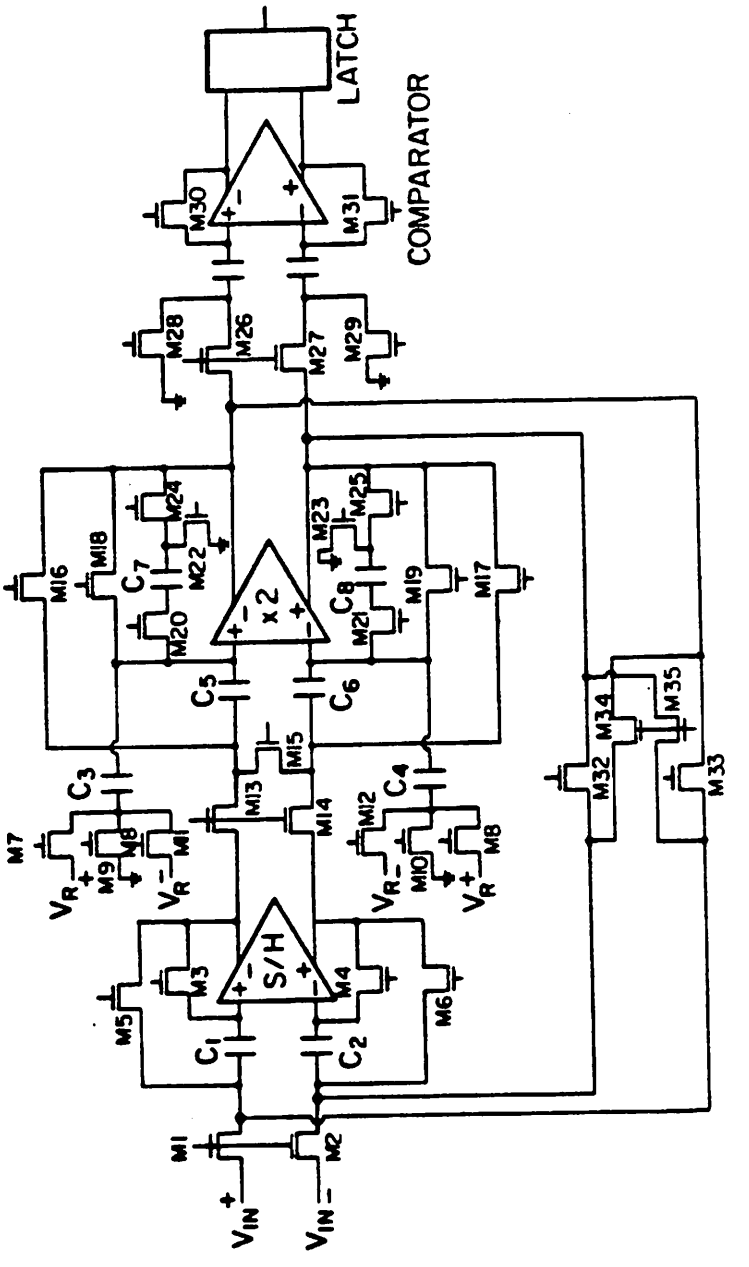

LATCH
COMPARATOR
M30
M31
M28
M26
M27
M29
M24
M23
M25
M16
M18
C7
M22
M20
x 2
C8
M21
M19
M17
C5
C6
M15
C3
M13
M14
C4
M12
M7
M9
M8
M1
M10
M8
S/H
M3
M4
M5
M6
C1
C2
M1
M2
V_R
V_{IN}
M32
M34
M35
M33

Figure 4.1

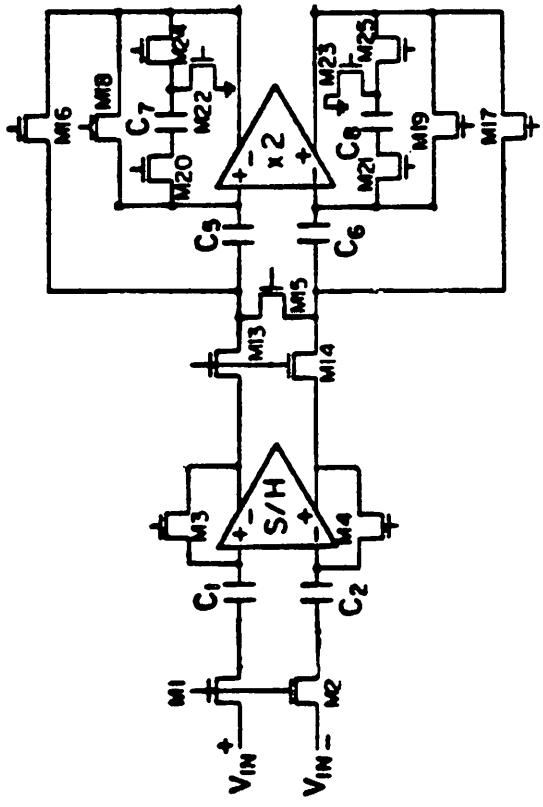

M16
M18
C7
M24
M22
M20
x 2
M23
M25
C8
M21
M19
M17
C5
C6
M15
M13
M14
S/H
M3
M4
C1
C2
M1
M2
+
VIN
VIN -

Figure 4.2

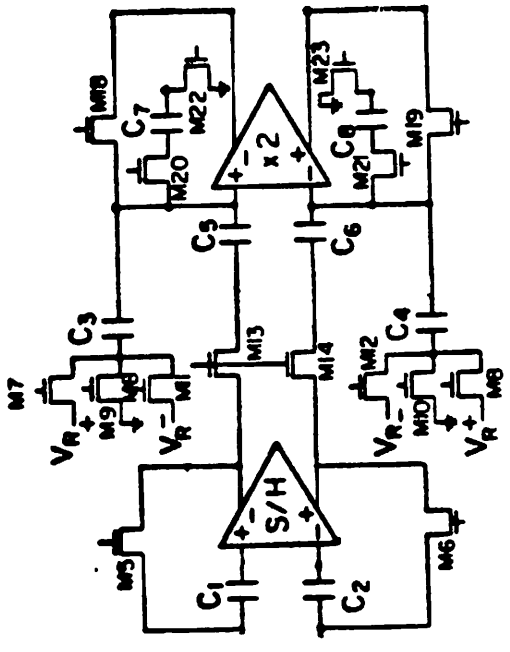

M18
C7
M22
M20
x2
M23
C8
M19
M21
C5
C6
C3
C4
M13
M14
M2
M7
M9
M8
M1
M10
M8
S/H
M5
M6
C1
C2

Figure 4.3

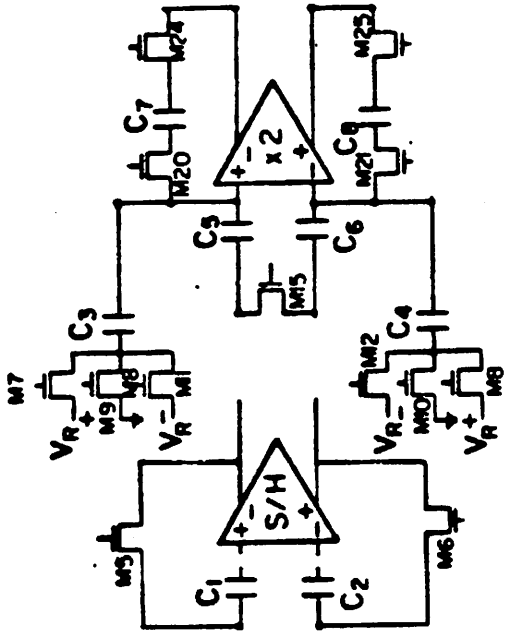

M24
M23
C7
C8
x2
M20
M21
C5
C6
M15
C3
C4
M12
M7
M9
M8
M11
M10
M8
VR+
VR-
VR-
VR+
S/H
M5
M6
C1
C2

Figure 4.4

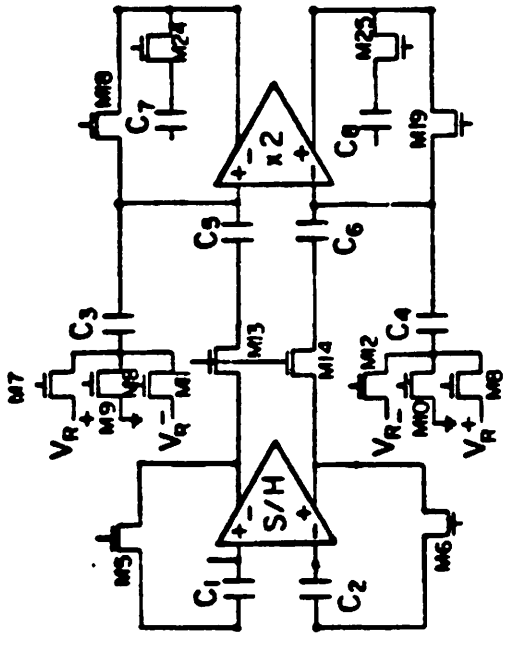

M24
M25
M18
M19
C7
C8
x2
C5
C6
C3
C4
M13
M14
M12
M17
M9
M1
M10
M8
VR-
VR+
S/H
M3
M6
C1
C2

Figure 4.5

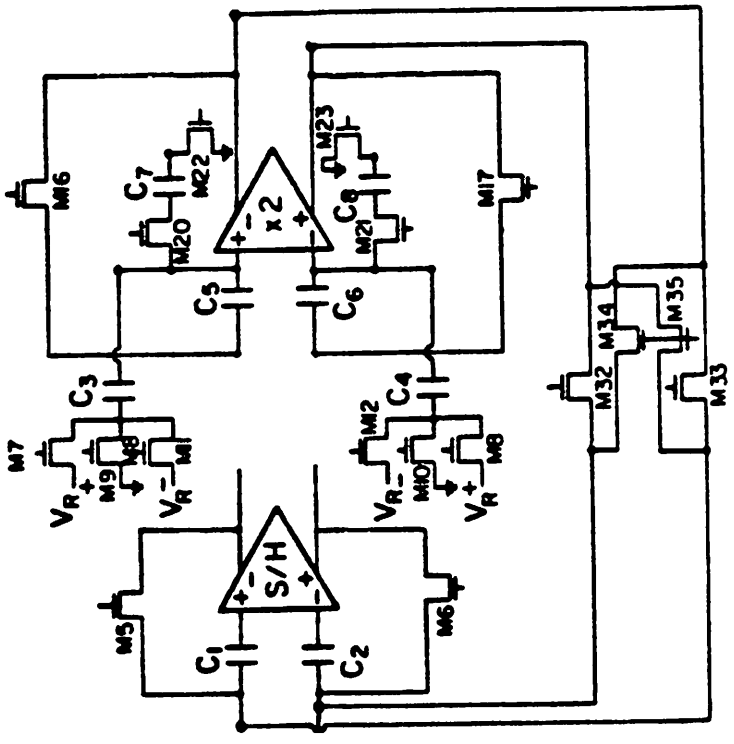

M16
M17
M23
M22
C7
M20
x2
C8
M21
C5
C6
C3
C4
M34
M35
M32
M33
M7
M12
M9
M8
M1
M10
M8
V_R^+
V_R^-
V_R^-
V_R^+
S/H
M5
M6
C1
C2

Figure 4.6

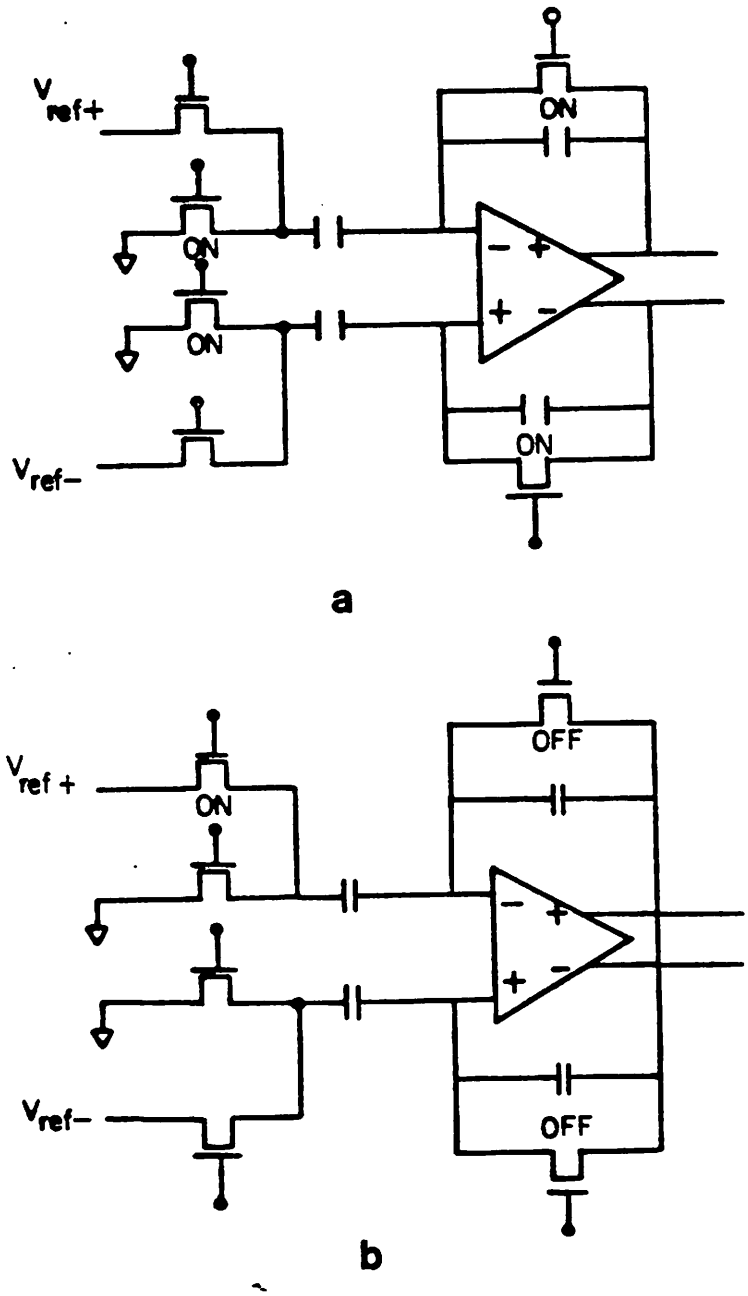


Figure 4.7

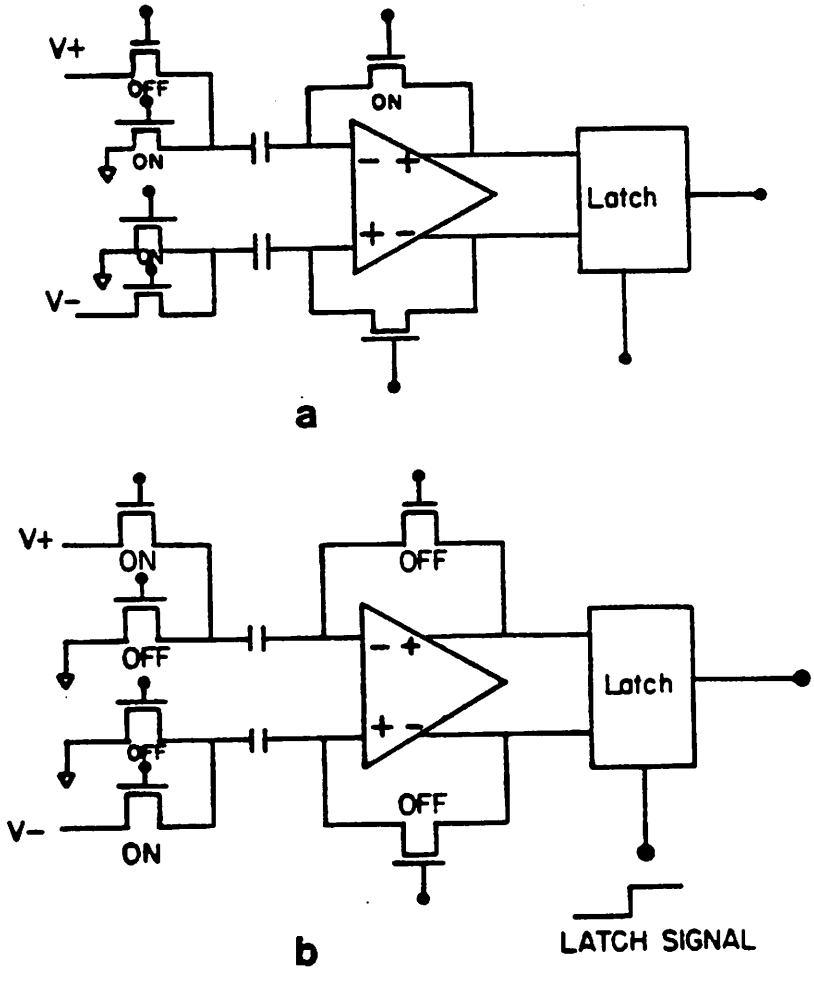


Figure 4.8

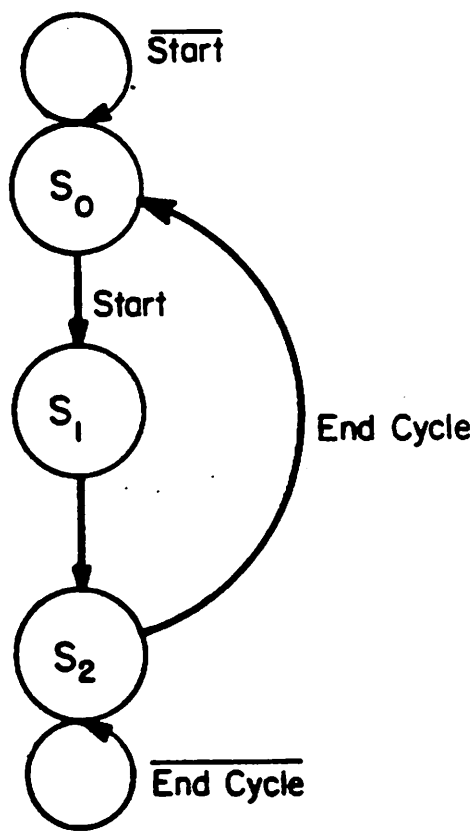


Figure 4.9

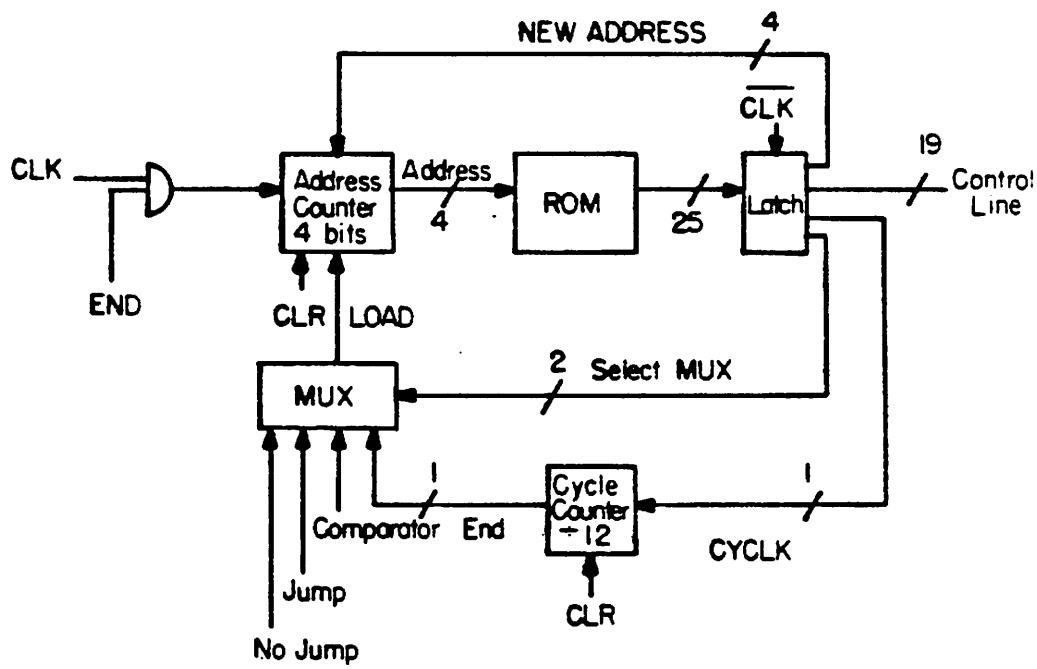
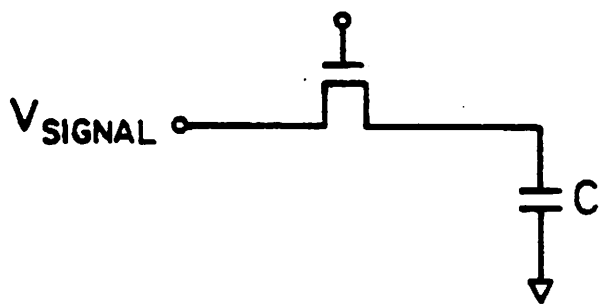

NEW ADDRESS
4
CLK
19
CLK
Address Counter 4 bits
Address
4
ROM
25
Latch
Control Line
END
CLR
LOAD
2
Select MUX
MUX
1
Cycle Counter ÷12
1
Comparator
End
CYCLK
Jump
CLR
No Jump

Figure 4.10



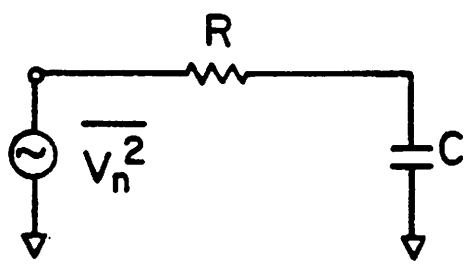


Figure 4.11

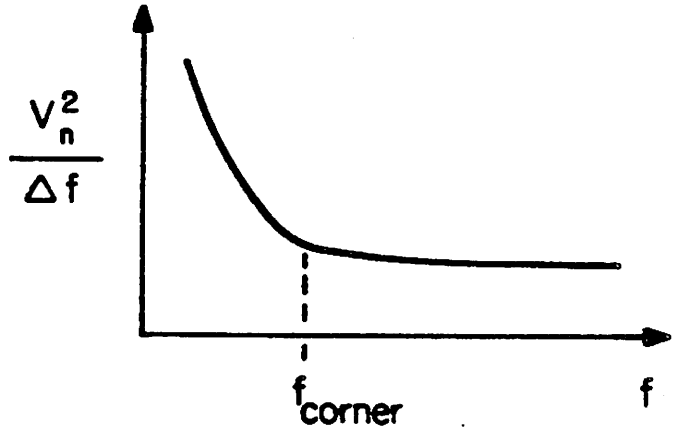


Figure 4.12

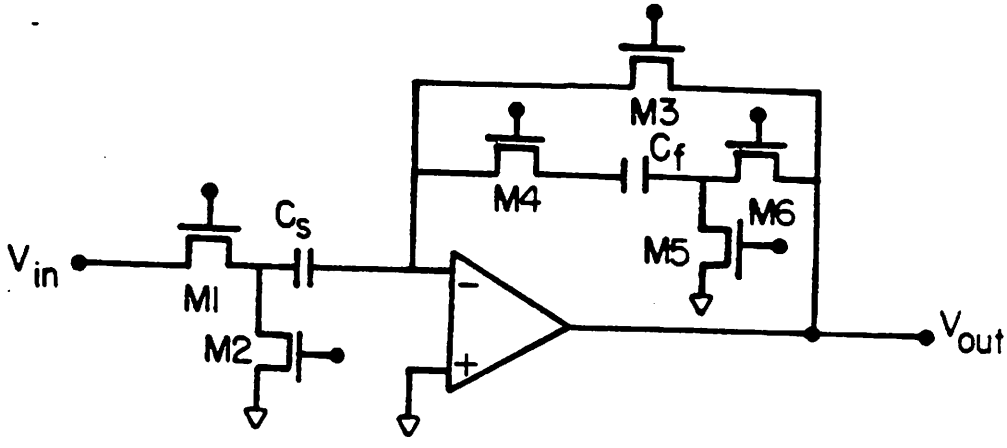
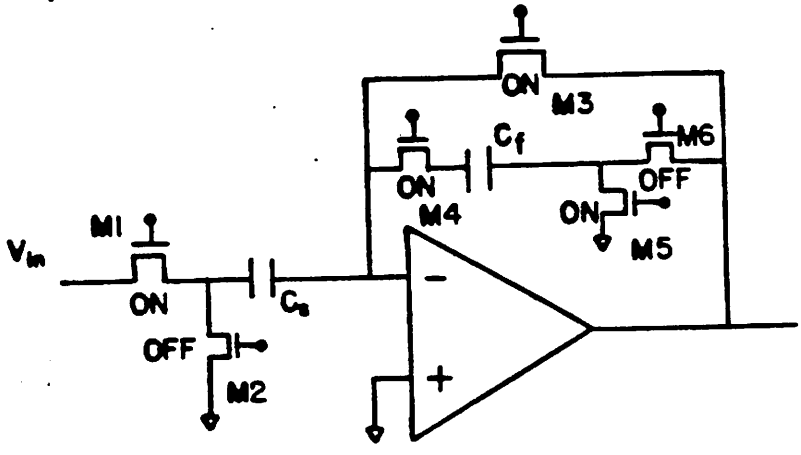
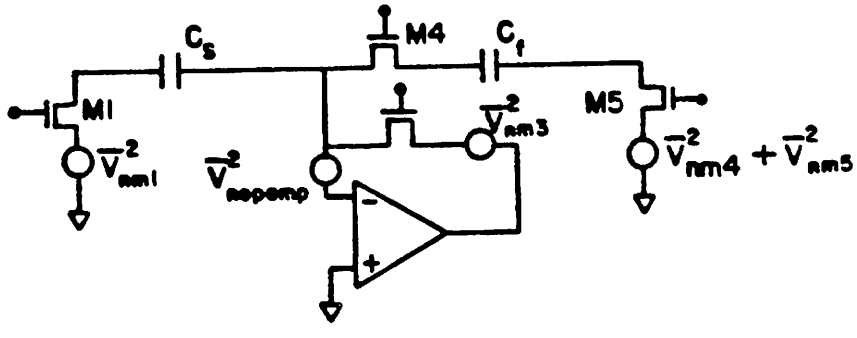


Figure 4.13

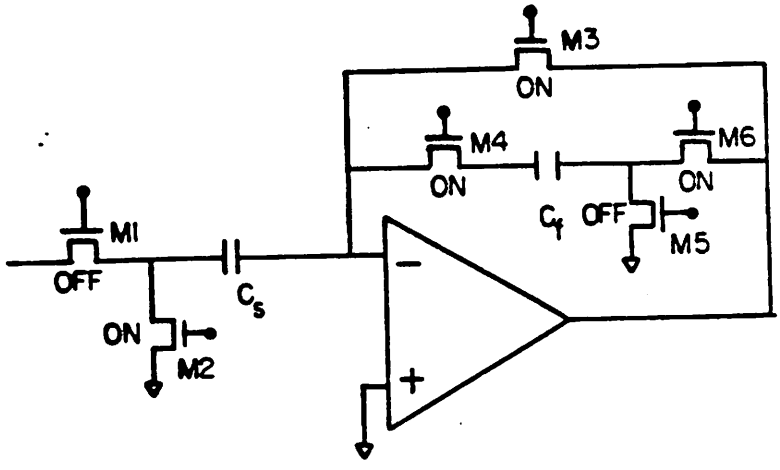


a

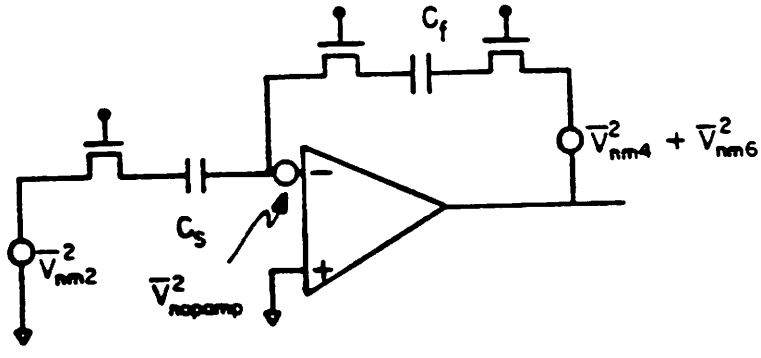


b

Figure 4.14

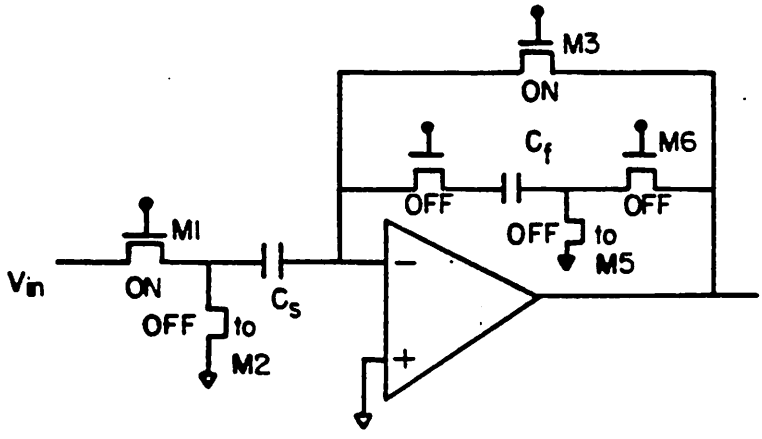


a

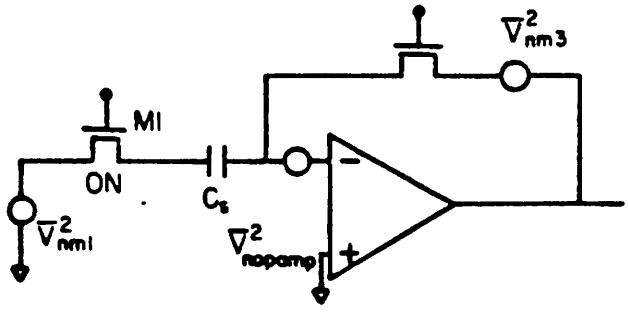


b

Figure 4.15



a



b

Figure 4.16

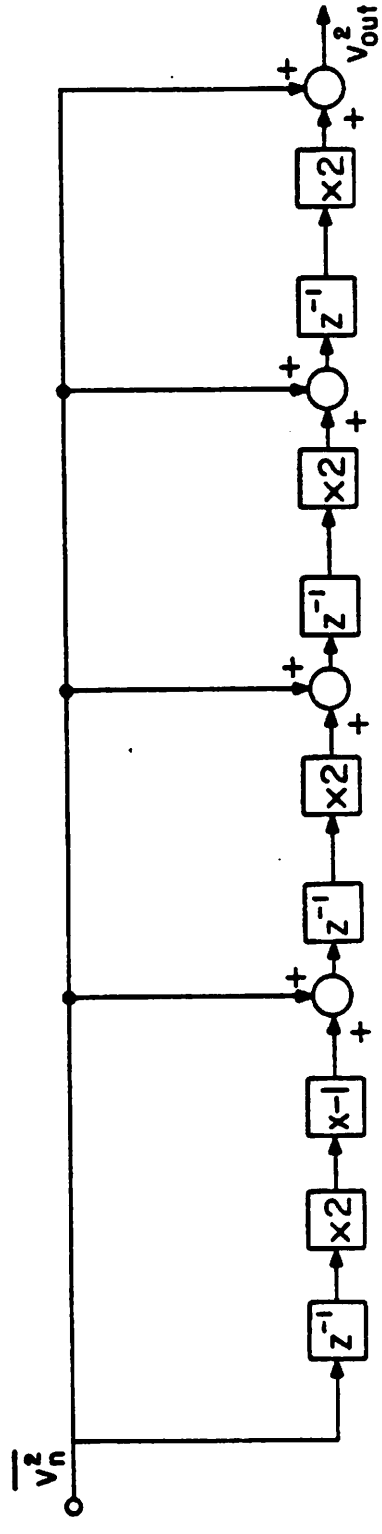


Figure 4.17

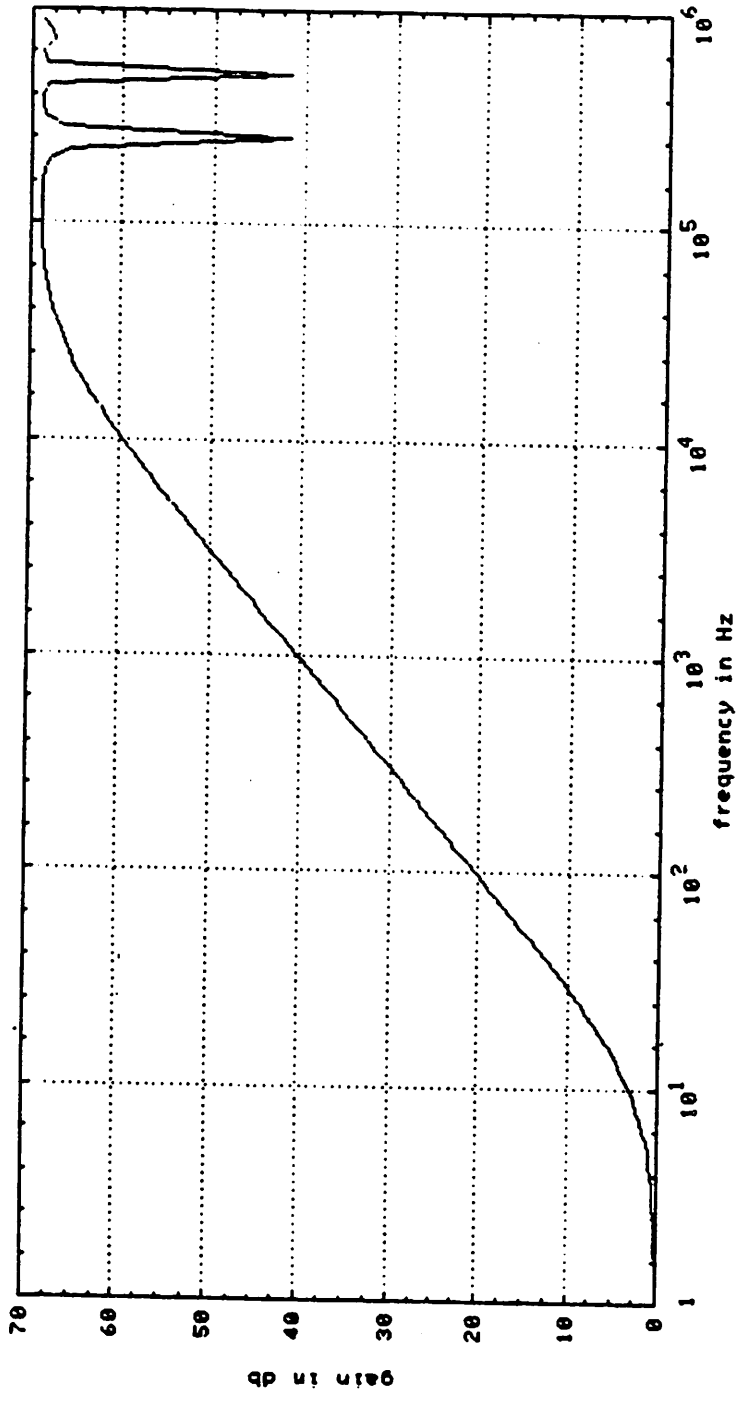

gain in db
70
60
50
40
30
20
10
0
1
10^1
10^2
10^3
10^4
10^5
10^6
frequency in Hz

Figure 4.18

address	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}
C_1	1	0	0	0	0	0	0	0	0	0	0
C_2	1	0	0	0	0	1	0	0	0	0	0
C_3	0	1	1	1	1	1	1	1	1	1	1
C_4	0	0	0	1	0	0	0	0	0	1	0
C_5	0	1	1	0	1	1	0	1	1	0	0
C_6	0	0	0	0	0	0	1	0	0	0	1
C_7	0	1	0	1	0	0	1	0	0	0	0
C_8	0	0	1	0	0	0	0	0	0	0	0
C_9	0	0	0	0	1	1	0	1	1	1	1
C_{10}	0	1	0	1	0	0	1	0	0	0	0
C_{11}	0	1	1	0	1	1	0	1	1	1	1
C_{12}	0	1	0	0	1	1	0	1	1	1	1
C_{13}	0	0	1	0	0	0	0	0	0	0	0
C_{14}	0	1	0	1	0	0	1	0	0	0	0
C_{15}	0	0	1	0	1	0	0	1	1	0	1
C_{16}	0	1	0	1	0	0	1	0	0	0	0
C_{17}	0	0	0	0	1	1	0	1	1	1	1

Table 4.1

transistor pair	
M_1M_2	C_1
M_3M_4	C_2
M_5M_6	C_3
M_7M_8	C_4
M_9M_{10}	C_5
$M_{11}M_{12}$	C_6
$M_{13}M_{14}$	C_7
M_{15}	C_8
$M_{16}M_{17}$	C_9
$M_{18}M_{19}$	C_{10}
$M_{20}M_{21}$	C_{11}
$M_{22}M_{23}$	C_{12}
$M_{24}M_{25}$	C_{13}
$M_{26}M_{27}$	C_{14}
$M_{28}M_{29}$	C_{15}
$M_{30}M_{31}$	C_{16}
$M_{32}M_{33}$	C_{17}
$M_{34}M_{35}$	C_{18}

Table 4.2

CHAPTER 5

Experimental Results

5.1. Introduction

In this chapter we discuss some of the practical techniques for realizing the algorithmic analog to digital converter. The CMOS process that was used to fabricate the experimental test chip is first described briefly. It has important implication since its feature sizes and available levels of interconnect determine the layout of the converter. Following this, the layout of the experimental chip is described. Finally some experimental results of the operational amplifier as well as the A/D converter are given.

5.2. Description of the CMOS Process

The CMOS process that is used to fabricated the ratio-independent algorithmic analog to digital converter is originally developed at Berkeley.[23] A description of the detailed processing sequence is included in the appendix. Some of the salient features of the process is as follows. The gate oxide thickness is about 45nm and the capacitor oxide thickness about 53nm. The junction depths for the n-channel and the p-channel drain/source junctions are about 0.5 and 0.7 μm. Basically it is a single polysilicon/single metal process with the capacitor formed with a extra bottom plate implantation. Thus the capacitor bottom plate(including the source/drain region of the transistors) together with the polysilicon and the metal give $1\frac{1}{2}$ levels of interconnect.

However because the crossover between the polysilicon and the diffusion has a large interelectrode capacitance, it can introduce a large amount of signal feedthrough and therefore the diffusion is not generally very useful for interconnecting signal lines if it has to pass under another polysilicon line. The specific resistance of the n^+ source/drain diffusion is close to 18Ω/square. For the p^+ region it is slightly higher than 25Ω/square. The polysilicon has a value higher than 45Ω/square of resistance. The minimum feature size for the gate and the contact holes is $4\mu m$.

5.3. Description of the Experimental Chip

As a test vehicle for the technique of ratio-independent algorithmic analog to digital conversion, an experimental integrated circuit A/D converter was designed and fabricated with the above mentioned CMOS silicon gate technology. The die photograph of the chip is shown in Fig. 5.1. In the middle of the chip the three operational amplifiers can be identified. Because they are fully differential operational amplifiers, their layout is perfectly symmetrical except for the biasing paths. In order to extend the frequencies of the non-dominant cascode poles in the opamp's, the polysilicon gates of the cascode transitors have been put as close to each other as possible;in this case $4\mu m$. In addition there are no overlap between the input lines and the output lines as this would represent a feedback capacitance which cannot be "switched around" as it is required in the ratio-independent switching algorithm. As a rule, in this type of converter, care should be taken to avoid running long input lines close to output lines because it introduces unnecessary interelectrode capacitance between the two. The p-channel transistors of the operational amplifiers are placed in two wells visible in the photograph. The wells have a diffusion contact around them to minimize latchup susceptibility.

The capacitors are formed between a bottom plate of heavily phosphorous doped diffusion layer and a polysilicon top plate. The SiO_2 dielectric is grown with the same thermal cycle that produces the gate oxide. But because it is formed from heavily doped silicon, the breakdown field strength is generally smaller, about 6-8 MV/cm, than that of the gate oxide, which can reached 10-12 MV/cm. For the same reason the capacitor oxide is thicker than the gate oxide. Its nominal thickness is about 63 nm compared to 50 nm for the gate oxide. The oxide integrity is generally good, exhibiting minute leakage until the breakdown field strength is reached. The capacitors are designed with the uniform value of about 4 pF. This value was chosen because it was felt that it represents a reasonable compromise between the value of the compensation capacitors and the kT/C noise. Also it was not known what accuracy the converter will eventually reached with the various algorithms available to cancel the offset and the 1/f noise. The capacitors are placed at a distance from the operational amplifiers so that the substrate displacement current generated during fast switching of the capacitors would not disturb the amplifiers. No substrate contacts have been placed around the capacitors.

The signal bus and the power bus are separated and run parallelly to each other. The signal bus lies on top of the chip as shown in Fig. 5.1. It contains the input signal lines, the reference signal lines, the ground line and the output lines of the operational amplifiers. Immediately below the signal bus are the switching transistors which are connected to the outside by means of extended polysilicon gates lines. These polysilicon gates lines run underneath the signal lines which are metal and are connected together at the other side by means of metal lines. This measure ensures equalized run times of the control signal to the gates of the two transistors situated on the differential signal paths. The transistors are connected to various points on the signal lines by means of n^+ diffusion strips. Again it can be seen that some diffusion strips are intentionally drawn out to equalize the impedances on the differential channels. In fact the overall symmetry

of the chip is evident. This is necessary from the standpoint of charge injection cancelation. The power bus contains the +5V, -5V, ground and the biasing voltages for the operational amplifier. They are connected to the biasing circuitry located on the lower right hand corner of the chip.

5.4. Control Circuit

The control circuit consists of a microprogrammable sequencer together with the supportive electronics. The sequencer is interfaced to a 8085 CPU through the S100 bus system and can be accessed as a series of port locations.[24] The control code can be loaded in the sequencer and debugged using a set of programs written in BASIC and assembly language.

The control circuit consists of 2 parts : the controller board which is physically located inside the S100 microcomputer and the translator board which houses the experimental chip and is electrically connected to the controller board by means of a 60 pin twisted cable. The translator circuit performs a 5V to a +5V/-5V voltage level translation as the signal coming out of the controller is TTL compatible. The controller is a microprogrammable sequencer similar in architecture to the 2910 bipolar bit slice microprocessor controller except it supports only the conditional and unconditional jumping using a full address feedback. The controller has a word width of 36 bits and a depth of 1k. In our experience, most of the control sequences can be encoded in half this depth.

The microprogrammable memory is composed of fast 2147 static RAM with a cycle time of 70 ns. The rest of the logic is realized with 74 series LS-TTL integrated circuits. The sequencer can send out control signal at the maximum rate of 7 MHz. Notice that there is a one clock delay between the actual output of the signal at the translator board

and the output at the controller because the signal is latched at the translator board with the clock. The fundamental clock period used for controlling the A/D converter is composed of several clock periods of the controller. With this method clock delays can be easily realized.

5.5. Test Setups

The operational amplifier is tested for differential gain using the circuit shown in Fig. 5.3b. It is a modification of a standard test circuit for operational amplifiers. The LF356 are connected in a non-inverting unity gain mode and buffer the output of the operational amplifier under test. The differential signals are then fed to a differential to single ended converter with unity gain. From there it is connected to a open loop 741, then to a resistive divider which is connected to the input of the opamp again. The transfer characteristics of the device under test can be traced by putting a periodic signal source on the input of the 741 opamp.

For the testing of the A/D converter, the experimental chip is mounted on the translator board. The reference voltages are supplied externally through batteries. The control codes are loaded into the microprogrammable sequencer which can be initiated to run by the S100 microcomputer. The data coming out of the converter is first processed before it is stored or sent to another computer. A code density test is employed to check for possible missing codes and the differential non-linearity in the converter. For the integral linearity, a DC test is used.

5.5.1. Code Density Test

In the code density test, a sine wave with extreme spectral purity is applied to the A/D converter under test. The A/D converter is made to sample this signal asynchronously with respect to the input sine wave frequency. The amplitude of this sine wave should be slightly larger than the full scale input of the converter so that theoretically all codes appear. Now if a large number of samples are taken, a histogram of relative occurrences of the codes can be constructed from which essential informations about the behaviour of the converter can be derived. The setup that accomplished this is shown in Fig. 5.2. The histogram of the codes is constructed using a dedicated LSI-11 microcomputer and then the information is forwarded to a VAX computer for further processing. More details about the test can be found in reference[7].

In contrast to the classical test which uses a D/A converter to characteristize the A/D converter, the code density test is a true dynamic test with the A/D converter operating at the sampling frequency of interest. In addition any random noise from the system is averaged out.

5.6. Experimental Results and Discussions

5.6.1. Operational Amplifier

Fig. 5.3a shows the transfer characteristics of a typical sample of the double cascode single stage operaional amplifier. Out of nine samples that were measured, the minimum gain obtained was larger than 40,000 or 92dB. The maximum gain obtained was close to 120,000 or 102dB. This large variation in the gain value can be traced back to the gain sensitivities on biasing current and gate lengths of the cascode transistors. The output conductance of the transistor is very sensitive to the gate length and therefore the $g_m r_{out}$ product of a transistor detoriates rapidly as the channel length drops below a certain

value which lies around 4-5μm in the process used. A correlation was found between the samples found with lower gain and shorter gate lengths of the transistors due to overetching of polysilicon.

With the on chip biasing circuitry, an output swing of +/-3.5V is obtained at a +/-5V voltage supply. The settling time to 13 bit level is estimated from results of the A/D converter to be about 1.5 μs. This time is about 70% longer than the time predicted by a SPICE simulation. One possible explanation of this discrepancy could be that all the parasitic capacitances are underestimated in the simulation. The experimentally verified performances of the operational amplifier are summarized in table 5.1.

5.6.2. A/D Converter

Fig. 5.4 shows a plot of the differential nonlinearity error versus code for a typical unit operating as a 12 bit converter at a reference voltage of 4V. The converter is sampling in randomly a test sine wave of 1.13kHz at a rate of 8kHz. The test was carried out at room temperature and a nominal supply voltage of +/-5V. The plot shows that there are no missing codes at 12 bit. There are occurrences of differential nonlinearity errors close to 0.9LSB at the major carry points. The plot is unsymmetrical around the ground point which corresponds to the code 2048. This asymmetry is probably due to the uncancelled portion of the charge injection offset. Calculations show that in order to maintain the differential nonlinearity error below the 1 LSB level, the residual charge injection offset in the loop must be kept below 1/2 LSB or approximately 0.5mV.

To confirm the previous results, another experiment was carried out. The A/D converter was operated as a cylic multiply-by-two amplifier. The control code was written so that after a input signal is sampled in, it is multiplied by 2^{12} times so that the voltage waveform can be observed at the output of the operational amplifier. When operated

slow enough (with $10\mu s$ clock cycles), the differential output voltages can be buffered and observed with an oscilloscope. Fig. 5.5 shows an example of the voltage waveform observed at the output of the multiply-by-two amplifier with an input equal to 0.15V. The double voltage peaks represent the transfers of charge onto the storage capacitor and then back onto the sampling capacitor. This voltage saturates after the 6th cycle. If the input voltage is 0, the resultant voltage waveform is shown in Fig. 5.6. The resultant voltage is only visible as an envelope of noise voltage. The shape of the envelope corresponds to the power series 2^n . The envelope is skewed towards the positive side due to the residual offset voltage in the loop. Application of an input voltage equal to 0.9mV restores the symmetry of the voltage envelope. Since the input equivalent offset voltage of the whole multiplication is equal to two times the loop offset voltage, the loop voltage is about 0.45mV or 1/2 LSB of the conversion.

An rough estimate of the loop noise can also be made. The amplitude of the noise voltage that is amplified 2^{12} can be read from the oscillogram. This noise voltage is found to be about 0.5mV or 1/2 LSB. This is also found to be the noise level obtained at the output of the A/D converter with a low noise DC input.

The integral linearity of the converter is tested by means of a dc test whose setup is shown in Fig. 5.7. A very stable programmable voltage supply is used to generate the voltage input to the converter. The converter is tested at its major carries. The linearity error depends on the conversion rate. With low conversion rate it is smaller. At 2kHz to 4kHz the worst case error was about 1.5 LSB. At 8kHz it increases to about 2.3 LSB. Above 8 kHz the error increases rapidly. This correlates with the appearances of missing codes at the major carry points. No satisfactory explanations for this effect has been found. There are two plausible reasons. The first is that the comparator might not be fast enough for small signals. The second is that the common mode settling time of the opera-

tional amplifier, which is approximately a factor of 10 slower than that of the differential signal, might be affecting the conversion.

Although no temperature measurement of the converter was made, the nature of the conversion is such that it should not be sensitive to temperature. Observation of the linearity error plots at different ambient temperatures confirms this result.

5.7. Suggestions for Improvement

Throughout the testing of the A/D converter it was found that the level of switching noise coming from the controller was considerable. Because the control signal were supplied to the converter through a 64 pin ceramic chip carrier, the coupling between neighboring channels was large. In fact the largest capacitance between two neighboring pins was found to be close to 10pF. This large interelectrode coupling was responsible for a part of the switching noise found in the converter. A second source of the switching noise was the coupling through the chip substrate. This coupling is caused by the use of diffusion strips for the control signal as well as the large bonding pads which have a substantial capacitance to the substrate. Thus the switching noise can be reduced by including the control logic on chip and running this logic synchronously to the switching of the operational amplifier.

The linearity of the converter is still limited by the residual loop offset voltage and noise. Better methods for eliminating this offset voltage and reducing the noise are required. For the present converter, the following techniques might be appropriate

1) Larger transistors should be used to obtain better matching between neighboring channels. At the same time, it serves to reduce the 1/f noise of the transistors. P-channel transistors, though having lower channel mobility, should be used instead

of the n-channel transistor as switches for short channel processes. They have inherently better noise properties and higher breakdown voltages.

2) Complementary switches should be used to balance the impedance. Impedance mismatch between the two neighboring channels is one major reason for large offset voltages.

3) Larger capacitors should be used for the sampling and integrating capacitors to reduce the kT/C noise. At the same time, use of p-channel transistors at the input of the operational amplifier should be beneficial.

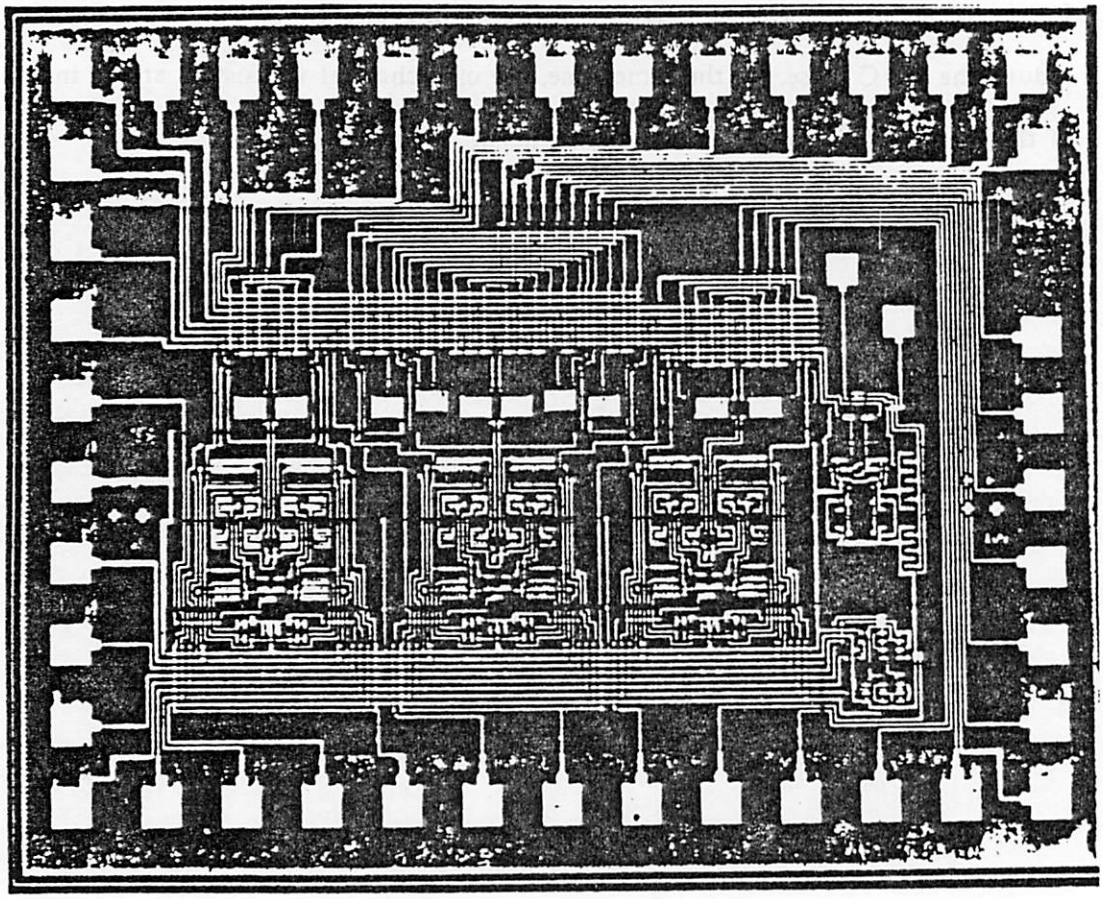


Figure 5.1

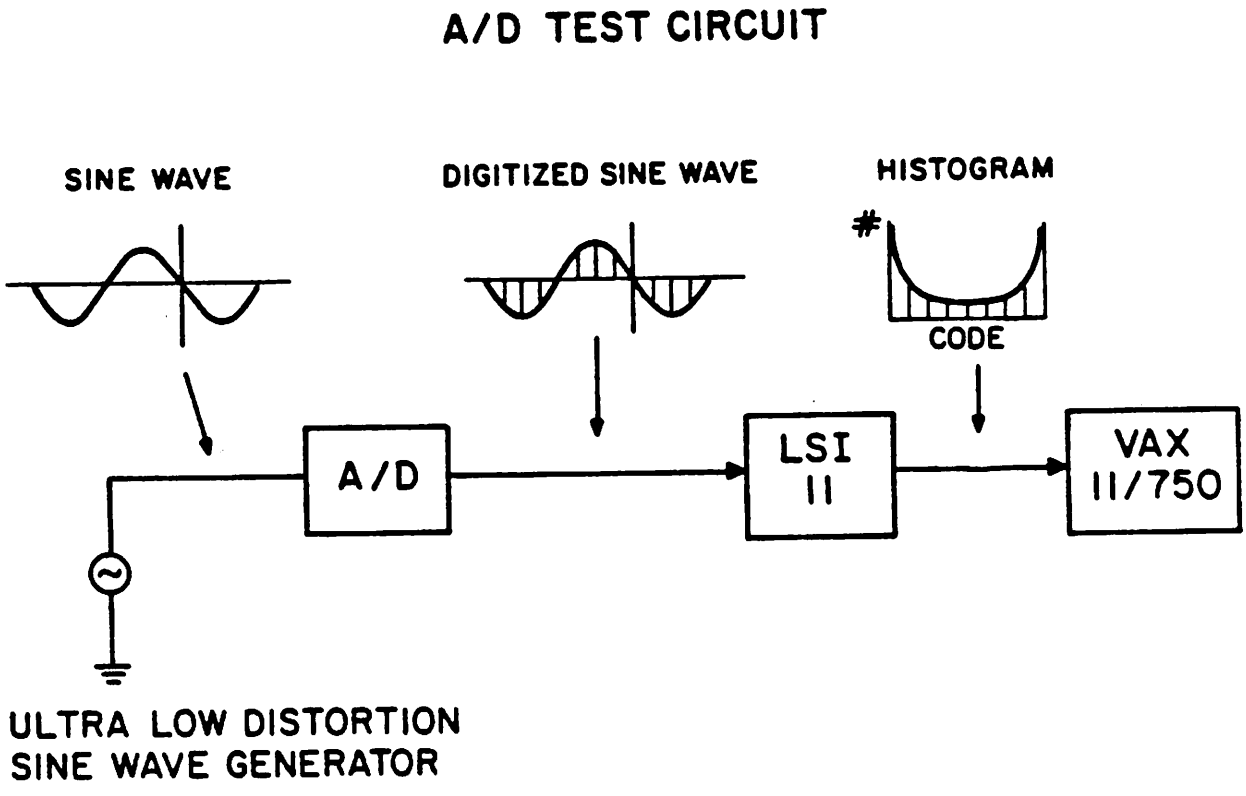
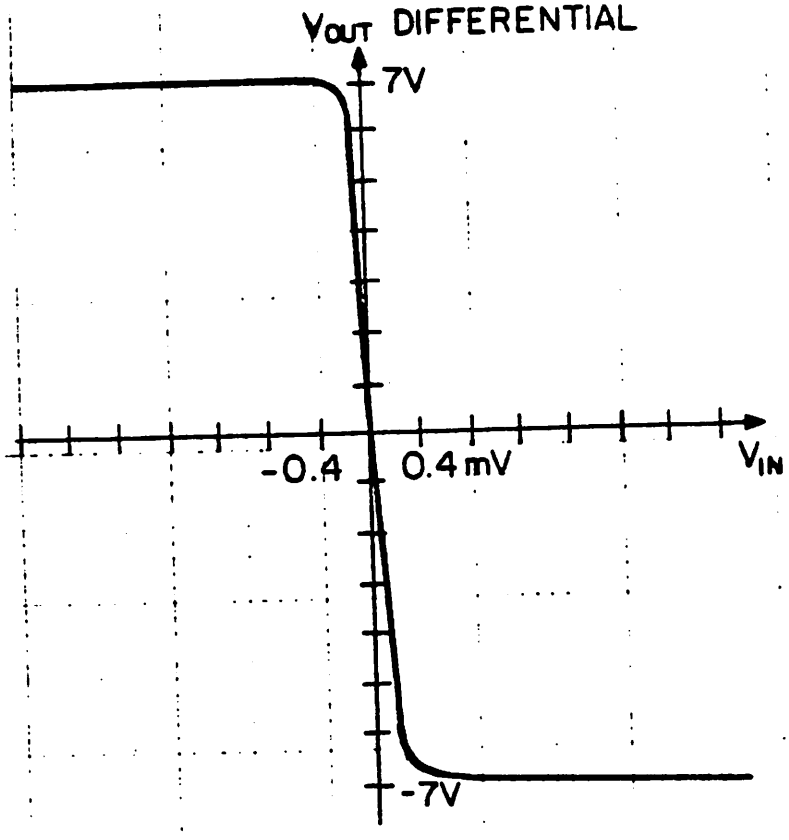
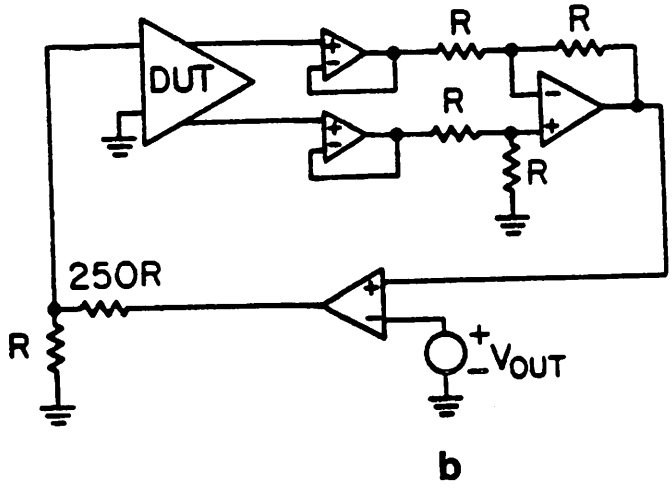

A/D TEST CIRCUIT
SINE WAVE
DIGITIZED SINE WAVE
HISTOGRAM
#
CODE
A/D
LSI
11
VAX
11/750
ULTRA LOW DISTORTION
SINE WAVE GENERATOR

Figure 5.2



a



b

Figure 5.3

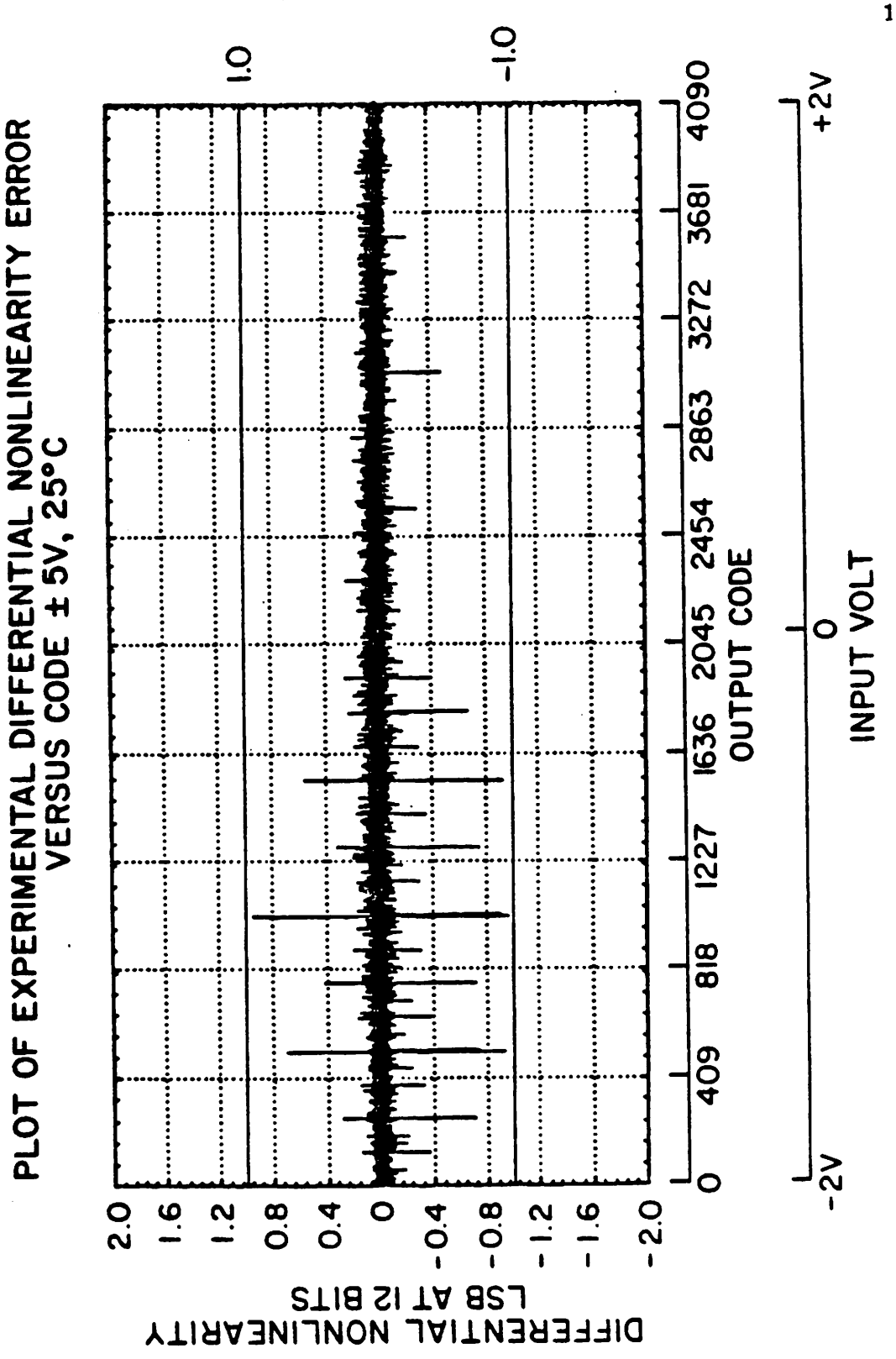

PLOT OF EXPERIMENTAL DIFFERENTIAL NONLINEARITY ERROR
VERSUS CODE ± 5V, 25°C
DIFFERENTIAL NONLINEARITY
LSB AT 12 BITS
2.0
1.6
1.2
0.8
0.4
0
-0.4
-0.8
-1.2
-1.6
-2.0
1.0
-1.0
0
409
818
1227
1636
2045
2454
2863
3272
3681
4090
OUTPUT CODE
-2V
0
+2V
INPUT VOLT

Figure 5.4

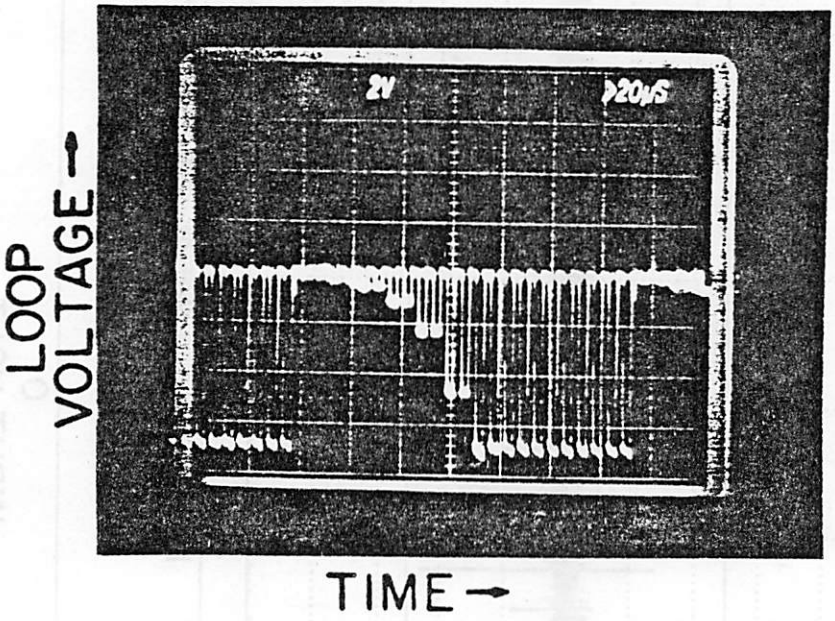


Figure 5.5

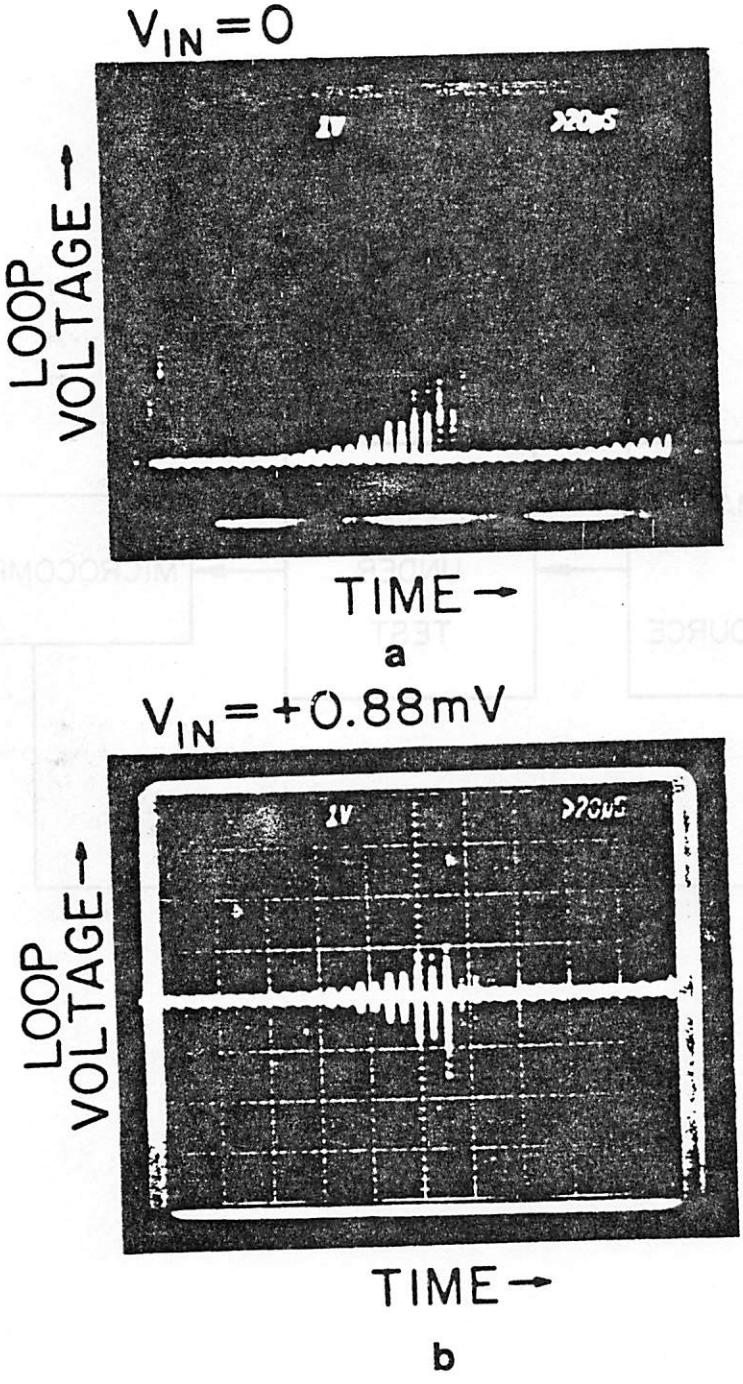


Figure 5.6

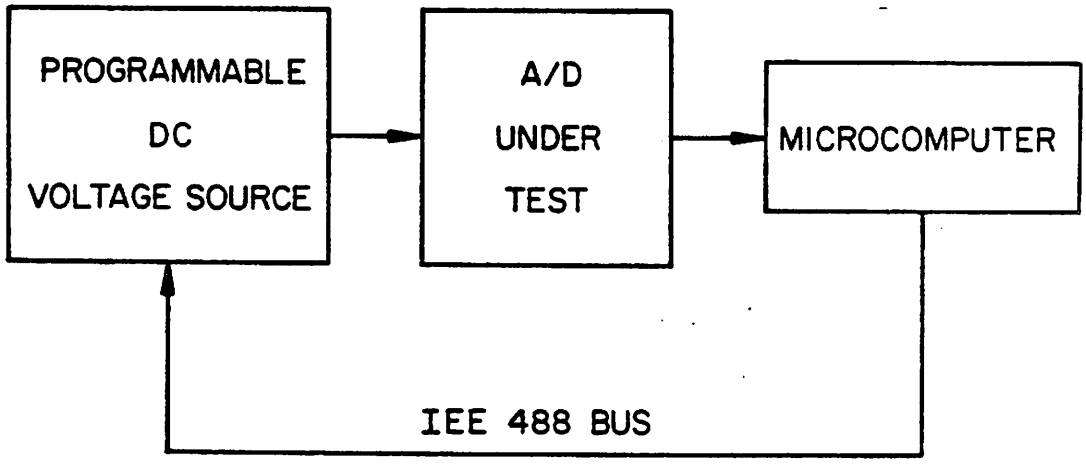

PROGRAMMABLE
DC
VOLTAGE SOURCE
A/D
UNDER
TEST
MICROCOMPUTER
IEE 488 BUS

Figure 5.7

OPERATION AMPLIFIER PERFORMANCE

(±5 V, 25°C)

GAIN (±6V DIFFERENTIAL OUTPUT)	>40,000
OUTPUT VOLTAGE SWING	±3.5V
SETTLING TIME FOR 4V	<1.5μs
DIFFERENTIAL OUTPUT AND 8pF LOAD	
POWER CONSUMPTION	4mW
AREA	400mil²

Table 5.1

SUMMARY OF PERFORMANCE OPERATION AS A 12 BIT A/D CONVERTER. ±5V 25 C	
DIFFERENTIAL NON - LINEARITY :	
4kHz SAMPLING RATE	0.019% (0.8 LSB)
8kHz SAMPLING RATE	0.022% (0.9 LSB)
INTEGRAL NON - LINEARITY :	
4kHz SAMPLING RATE	0.034% (1.5 LSB)
8kHz SAMPLING RATE	0.081% (3.2 LSB)
POWER CONSUMPTION	17mW
AREA	2400mil²
SUPPLY VOLTAGE RANGE	±4.5V - 6.5V
TECHNOLOGY	5μ POLYSILICON GATE CMOS

Table 5.2

CHAPTER 6

Conclusion

The algorithmic analog to digital conversion technique has been described analyzed. The conversion is done by multiplying the signal by a factor of two, compare it to the reference signal and subtracted off the reference signal from the result if necessary. This technique is capable of performing analog to digital conversion with very little analog hardware and provides the sample and hold function. In addition, it possesses the so called floating capability where a signal can be multiplied with the powers of 2 before the actual analog to digital conversion begins. The linearity of the converter, however, is sensitive to the loop gain error and loop offset error of the analog amplifier loop.

Conventionally, gain setting in MOS gain blocks have been made by matching ratioed capacitors. Therefore in realizing the algorithmic analog to digital converter, the accuracy of the conversion has been limited by the matching of the capacitors. A new method of obtaining integral gain values independent of capacitor values has been devised. Essentially it uses an MOS integrator to integrate and store the input voltage. By exchanging the sampling and integrating capacitors in the last stage of the process, ratio-independent multiplication of the input signal would then be achieved. In this ratio-independent multiplication scheme, charge injection from MOS transistors poses the major problem from the standpoint of accuracy. By employing a special switching scheme and a differential circuit configuration, the charge injection effect can be minimized.

The various design aspects of the CMOS operational amplifier used in the algorithmic analog to digital converter are discussed. In the present design, a fully

differential class A/B configuration has been adopted. Large gain is obtained by boosting up the output node impedance through double cascode transistor current sources. Improved common mode behavior is accomplished with a balanced common mode feedback circuit that employs two identical differential pairs to cancel their inherent nonlinearity.

The limitation factors to the conversion accuracy are identified to be the loop offset error and the various noise sources in the converter. These noise sources include the 1/f noise of the operational amplifier, the kT/C noise of the sampled data capacitors and the switching noise of the control circuitry. Methods to minimize the contribution of each of these error or noise source are discussed. In the design of the layout of the A/D converter for an CMOS process, special precaution can be taken to minimize the impedance imbalance and charge injection difference on the two channels of the fully differential circuit. This leads to reduced loop offset voltages. Noise reduction can be obtained by switching the path of the differential signal during the analog to digital conversion process. The switching noise poses a more serious problem which can practically limit the resolution of the converter to less than 14 bits.

An experimental test chip has been built to verify the ratio-independent analog to digital conversion technique. A $5\mu m$ polysilicon gate CMOS technology is used. The experimental circuit demonstrates resolution to 12 bits and a integral linearity of 10 bits at a sampling rate of 8kHz. The chip area of the analog circuit is $2400 mil^2$. By including the control logic on chip as well as using better matched transistors to reduce the loop offset voltage, higher resolution and better linearity are achievable.

Appendix I. Berkeley CMOS Process

Process Flow

1. INITIAL WAFER PREPARATION

1.1. Piranha Clean

Clean for 10 mins using glass or teflon beakers and holders. Do not use plastic beakers. Piranha will attack most plastics and leave deposits on the wafers. Never clean wafers and tweezers together.

1.2. Oxide Dip and Water Break Test

Dip the wafer in 10:1 solution of HF until the wafer becomes hydrophobic. HF solution will completely clear off the surface if the wafer has been properly cleaned.

2. INITIAL OXIDATION

Grow 3000 Å of oxide in the initial oxidation furnace. TCA clean the furnace prior to oxidation.

TEMP=1100			
Push	Ox	5.0	5 min
Ox	Ox	5.0	240 min
Anneal	N_2	5.0	10 min
Pull	N_2	5.0	5 min

3. N-WELL DEFINITION

3.1. Standard Photolithography

HMDS	3 min
N_2 purge	5 min
Spin Resist	AZ1350J 6000 rpm/30 sec
Soft Bake	90 °C/15 min

Pattern	Cannon 5.6
Develop	MicroPosit Developer:DI =1:1 /60sec

3.2. Oxide Etch

BHF 2 min

prepare the BHF solution at least 24 hours prior to use to get reproducible etch rate.

3.3. Photoresist Removal

Acetone	2 min
Methanol	1 min
DI rinse	1 min
Piranha	5 min

3.4. N-Well Implantation

Phosphorus 100 keV 8 ° 1.5×10^{12}

3.5. N-Well Drive-In

Buried layer furnace - TCA clean prior to use

Piranha clean 5 min

Temp = 1100 °C

Push	N_2	15 cm	3 min
Oxidation	O_2	15 cm	280 min

Ramp temp to 1150 °C

Drive-in	$N_2 : O_2$	5.0 cm : 15 cm	720 min
Anneal	N_2	10 cm	20 min
Pull	N_2	10 cm	3 min

The drive-in is performed in 10% O_2 atmosphere

4. Buffer Oxidation

4.1. Oxide Etch Back

HF : DI = 1 : 5

Etch until wafer becomes hydrophobic (~ 8 min)

4.2. Oxidation

Piranha clean 3 min

N-drive furnace - TCA clean prior to use

target : 550 Å

Temp = 1000 °C

Push	N_2	10 cm	3 min
Oxidation	O_2	11 cm	50 min
Anneal	N_2	10 cm	10 min
Pull	N_2	10 cm	3 min

5. Nitride Deposition

Load 2 wafers at a time, facing outward. Better uniformity can be obtained by loading the wafer facing inside, which takes a considerably longer time and gives more particular contamination. Use 4 dummy slats (half wafers) to smooth out the gas flow.

To prevent contamination, nitride should be deposited immediately after the oxidation.

Target : 1500 Å
NH_3 600 mT
SiH_4 100 mT

Determine the deposition time by a dummy run, deposition rate varies significantly at times.

6. Active Area Definition

6.1. Standard Photolithography

6.2. Nitride Etch

Descum is necessary to remove photo resist residue. Use barrel reactor. 5 mins. of 100W O_2 plasma at 65 °C is enough.

Etch Plasma-Therm SF_6O_2 60°C 100 W

Determine the etch rate by a dummy run. Approximately 10 min is needed.

7. P-Field Definition

7.1. Standard Photoresist Removal

7.2. Standard Photolithography

7.3. P-Field Implantation

Boron 100 keV 8° 1.5×10^{13}

8. Backside Implantation

8.1. Standard Photoresist Removal

8.2. Spin on Protective Photoresist

Prebake	IR or 115°C Oven	
HMDS	3 min 5 min Purge	
AZ-1350J	5000 rpm	30 sec
Hard bake	115 °C	10 min

8.3. Backside Oxide Etch

Etch in BHF until wafer becomes hydrophobic.

8.4. Implantation

BF_2 200keV 8° 2×10^{15}

9. LOCOS

9.1. Standard Photoresist Removal

9.2. Oxidation

P-Drive Furnace. Target : 8500 Å

Temp = 950 °C

Push	N_2	5 cm	3 min
Oxidation	Wet O_2	2 cm	255 min
Anneal	N_2	10 cm	20 min
Pull	N_2	10 cm	3 min

Set the bubbler heater to 110 V. 2 refills of the bubbler are needed at this setting. Switch the gas to nitrogen during refills and subtract this time from the oxidation time.

10. Nitride Removal

Oxide Dip: HF : DI 1 : 10 40 sec

This HF dip removes thin oxide grown over nitride during LOCOS.

Phosphoric Acid 155 °C 30 min

Use reflux to maintain the temperature at 155 °C.

11. Threshold Implantation

Boron 50 keV 8 ° 7×10^{11}

Compensate for noise

12. Capacitor Area Definition

12.1. Standard Photolithography

12.2. Pre-implantation Bake

N_2 1 Torr60 W 30 min

12.3. Capacitor Implantation

Phosphorus 80 keV 8 ° 2.5×10^{15}

Allow minimum of 20 mins per wafer to avoid damage to the photoresist

13. Gate Oxidation

13.1. Photoresist Removal

Plasma Ash O_2 1 Torr30 min

Piranha 5 min

13.2. Buffer Oxide Removal

HF : DI 1 : 10 80 sec

13.3. Oxidation

N Drive-in Furnace : TCA Clean prior to use

Target : 500 Å

Temp = 1000 °

Push	N_2	10 cm	5 min
Oxidation	O_2	11 cm	45 min
Anneal	N_2	10 cm	10 min
Pull	N_2	10 cm	5 min

14. Polysilicon Deposition/Doping

14.1. Piranha Clean

Clean for 10 mins. Use teflon tweezer only. Metal tweezer will create a nucleation center to cause a milky polysilicon.

14.2. Polysilicon Deposition

Polysilicon should be deposited right after the oxidation to prevent contamination. Place wafers with the active side facing each other. Dummy slats are not necessary.

Target : 3500 Å

SiH_4 600 mT

Determine the deposition rate by a dummy run.

14.3. Polysilicon Doping

N-predep Furnace

Temp = 950 °C

Push	N_2	4 cm	5 min
Oxidation	$O_2 : N_2$	2.5 cm : 5 cm	10 min
Doping	$O_2 : N_2 : POCl_3$	2.5 cm : 5 cm : 3 cm	30 min
Anneal	N_2	4 cm	5 min
pull	N_2	4 cm	5 min

Turn on the source cooler 30 mins. prior to doping.

15. NMOS Definition

15.1. Deglaze

HF : DI 1 : 10 10 sec

The phosphorus rich glass should be removed on which photoresist will not adhere.

15.2. Standard Photolithography

15.3. Standard Plasma Etch

Descum - Etch in Plasma Term

Polysilicon etches much faster than nitride. The etch rate should be determined by a dummy run.

15.4. Standard Photoresist Removal

15.5. NMOS S/D Implantation

As 180 keV 8 ° 3×10^{15}

16. PMOS Definition

16.1. Piranha Clean

16.2. Standard Photolithography

16.3. Standard Plasma Etch

Due to the implantation damage, polysilicon etches faster at this time. Be careful not to over-etch.

16.4. Pre-implantation Bake

N_2 1 Torr 60 W 45 min

16.5. PMOS S/D Implantation

Boron 60 keV 8° 2×10^{15}

Allow at least 20 mins per wafer.

16.6. Photoresist Removal

Plasma Asher O_2 1 Torr 120 W 40 min

Piranha Clean 5 min

16.7. Polysilicon Re-oxidation

P-drive Furnace

Temp = 1000 °C

Push	N_2	10 cm	5 min
Oxidation	O_2	11 cm	30 min
Anneal	N_2	10 cm	10 min
Pull	N_2	10 cm	5 min

17. CVD Passivation

17.1. Piranha Clean

17.2. CVD Glass Deposition

3500 Å Undoped CVD Glass
6500 Å 7 % PSG

Due to high reflow temperature, a sandwich layer of undoped CVD glass is needed to prevent counter doping of p+ regions.

17.3. Reflow

N Drive-in Furnace

Temp = 1050 °C

Push	N_2	5 cm	5 min

Reflow	N_2	5 cm	20 min
Push	N_2	5 cm	5 min

18. Contact Definition

18.1. Standard Photolithography

Increase the light dose to 5.9

18.2. Contact Hole Etch

Etch in BHF

Bake 130 °C 10 min

Repeat etch-bake cycles as many times as required to prevent contact cut blooming. Recommended etching times are 15 sec., 30 sec., 45 sec., 60 sec., 90 sec. and then repeat 120 sec. etch until all the contact hole monitors are clear of oxide. Then over etch for 1 min to assure a complete etch.

18.3. Standard Photoresist Removal

19. Metalization

19.1. Oxide Dip HF : DI 1 : 10 5 sec

This oxide dip removes a thin oxide grown over the vials during the piranha clean. Be careful not to over etch the CVD glass. PSG etches extremely fast.

19.2. Pre-metalization Bake

Throughly dry the wafer under IR lamp or 110 °C oven.

19.3. Aluminum Sputtering

Target : 0.75 μm

20. Metal Run Definition

20.1. Standard Photolithography

Due to a high reflectivity of aluminum, the exposure should be reduced to 5.5.

20.2. Aluminum Etch

Aluminum Etchant Type A 40 °C

Approximately 90 sec. is needed. Allow 10 sec. over etch after all the aluminum is gone over the unmasked area.

21. Backside Metalization

21.1. Protective Photoresist Spin-on

IR Bake 20 min

AZ-1350J 5000 rpm 30 sec

Hard Bake 120 °C 20 min

21.2. Backside Polysilicon Etch

Polysilicon on the backside of the wafer is not removed if Plasma-Therm is used for the polysilicon etch. It should be removed before the backside metalization.

Standard Plasma Etch

21.3. Backside Oxide Etch

Etch in BHF until the wafer becomes hydrophobic.

21.4. Aluminum Sputtering

Target : 1 μm

21.5. Photoresist Removal

Acetone 3 min - Methanol Rinse - DI Rinse

22. Sintering

Sintering Furnace

Temp = 375 °C

Forming Gas 15 cm 20 min

REFERENCES

[1] Schmidt, Electronic Analog Digital Conversion, Van Nostrand, 1970

[2] Tom Hornak, A High Precision Component Tolerant ADC, Digest of Technical Papers, 1975 ISSCC, New York, NY, February 1975.

[3] R. H. McCharles, V. A. Saletore, W. C. Black, Jr. and D. A. Hodges, An Algorithmic Analog-to-Digital Converter, Digest of Technical Papers, 1977 ISSCC, Philadelphia, PA. February 1977.

[4] R. W. Webb, F. R. Cooper and R. W. Randlett, A 12b A/D Converter, Digest of Technical Papers, 1981 ISSCC, New York, NY, February 1981.

[7] R. H.McCharles, *ibid*

[6] R. G. Meyer, Notes for EECS 142, U.C. Berkeley, 1983.

[7] J. Dornberg, Full Speed Testing of A/D Converters, submitted for publication in JSSC.

[8] J. L. McCreary and P. R. Gray, All MOS Charge Redistribution Analog-to-Digital Conversion Technique, IEEE Journal of Solid-State Circuits, vol. SC-10, pp. 371-379, Dec. 1975.

[9] R. Van de Plaasche, Dynamic Element Matching for High Accuracy D/A Converters, Digest of Technical Papers, 1976 ISSCC, New York, NY, February 1976.

[10] H. S. Lee, D. A. Hodges and P. R. Gray, Self-calibrating CMOS A/D Converter, *submitted for publication in JSSC.*

[11] C. C. Lee, A New Switched-capacitor Realization for Cyclic Analog-to-Digital Converter, Digest of Technical Papers, 1983 International Symposium on Circuit and

Systems, Newport Beach, California, May 1977.

[12] Bing J. Shu, Switch-Induced Error Voltage on a Switched Capacitor, Master Thesis, University of California Berkeley, June 1983.

[13] L. Bienstman and H. J. De Man, An Eight Channel 8 Bit Microprocessor Compatible NMOS D/A Converter with Programmable Scaling, IEEE Journal of Solid-State Circuits, vol. SC-15, pp. 1051-1059, Dec. 1980.

[14] K. C. Hsieh, P. R. Gray, D. Senderowicz and D. G. Messerschmitt, A Low Noise Chopper Stabilized Differential Switched-Capacitor Filtering Technique, IEEE Journal of Solid-State Circuits, vol. SC-16, pp.708-716, Dec. 1981.

[15] Shinji masuda, *private communication.*

[16] R. R. Troutman, Low-level Avalanche Multiplication in IGFET's, IEEE Transaaction on Electron Devices, vol. ED-23, pp. 419-425, April 1976.

[17] P. R. Gray, Notes for EECS 241, U.C. Berkeley, 1982.

[18] B. K. Ahuja, An Improved Frequency Compensation Technique for CMOS Operational Amplifiers, IEEE Journal of Solid-State Circuits, vol. SC-18, pp. 629-633, Dec. 1983.

[19] J. A. Connelly, Analog Integrated Circuits, Wiley, New York, 1974.

[20] W. Black, Thesis, U.C Berkeley, 1979.

[21] Ron Kanishiro, Thesis, U.C. Berkeley, 1984.

[22] K. C. Hsieh, Thesis, U.C. Berkeley, 1982.

[23] R. Kanishiro, Tat Choi, Berkeley CMOS Process, unpublished, 1983.

[24] M. Chin, Master Thesis, U.C. Berkeley, 1984.