

Copyright © 1980, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

**HIGH SPEED CMOS A/D
CONVERSION TECHNIQUES**

HIGH SPEED CMOS A/D CONVERSION TECHNIQUES

by

William C. Black, Jr.

William C. Black, Jr.

MEMO Memorandum No. UCB/ERL M80/54

1 November 1980

COVER

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

HIGH SPEED CMOS A/D CONVERSION TECHNIQUES

by

William C. Black, Jr.

Memorandum No. UCB/ERL M80/54

November 1980

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

TABLE OF CONTENTS

CHAPTER 1 - INTRODUCTION	1
CHAPTER 2 - HIGH SPEED A/D CONVERTER APPLICATION AREAS	3
2.1 Digital Time-Base Correction	3
2.2 Digital Video Transmission and Storage	4
2.3 Real Time Image Processing	6
2.4 Digital Output Monolithic Imager	7
2.5 Wide Dynamic Range Digital Audio	7
CHAPTER 3 - CHARACTERIZATION AND DESIGN OF HIGH SPEED MONOLITHIC A/D CONVERTERS	9
3.1 Characterization of High Speed A/D Converters	9
3.1.1 Quasi-Static Characteristics	9
3.1.2 Dynamic Characteristics	15
3.2 Flash Converter Technique	23
3.2.1 Comparator Input Bias Currents	26
3.2.2 Resistor Matching	26
3.2.3 Comparator Intrinsic Offset	34
3.2.4 Dynamic Ladder Tap Loading	36
3.2.5 Clock Skew Between Comparators	38
3.2.6 Comparator Phase Delay	40
3.3 Successive Approximation Technique	41
3.3.1 Weighted-Capacitor Methods	44
3.3.2 Capacitor Matching Requirement	50
3.3.3 Speed Limitations	57
3.4 Pipeline Converters	62
3.4.1 Divided Reference Pipeline	64
3.4.2 Multiplied Remainder Pipeline	69
3.4.3 Pipeline Performance Limitations	71

3.5	Hybrid Converters	72
3.5.1	Two-Step Flash Converter	73
3.5.2	Flash-Pipeline Converter	74
CHAPTER 4 - PRECISION ELEMENTS		79
4.1	Resistor Matching Considerations	79
4.2	Characteristics of Planar Diffused Resistor Structures	82
4.3	Characteristics of Diffused Resistors Due to Sidewalls	87
4.4	Compensation of Voltage Dependence in a Diffused Monolithic Resistor	97
4.5	Temperature Coefficient of Resistors	101
4.6	Matching Consideration in Precision Capacitors	103
4.7	Capacitance of Nonplanar Structures	107
4.8	Voltage Coefficient of MOS Capacitors	113
4.9	Temperature Coefficient of MOS Capacitors	117
CHAPTER 5 - ANALOG CIRCUIT BUILDING BLOCKS		120
5.1	Source Followers	121
5.1.1	Frequency Response and Gain	121
5.1.2	Transfer Characteristic Nonlinearity	123
5.1.3	Elimination of Intrinsic Offset	125
5.2	Single Gain Stage Linear Amplifiers	127
5.2.1	Class-A Amplifiers	127
5.2.2	Class-B Amplifiers	133
5.2.3	Methods for Increasing Bandwidth	136
5.3	Two Gain-Stage Linear Amplifiers	140
5.3.1	Compensation Techniques	142
5.3.1A	Simple Pole-Splitting	142
5.3.1B	Pole-Splitting with Feedforward Blocking	144
5.3.1C	R-C Compensation Method	145

5.4	Integrators	149
5.4.1	Dyanmic Characteristics	150
5.4.2	Power Supply Rejection	154
5.4.2A	General Case	154
5.4.2B	Effect of Offset Cancellation Cycle	157
5.5	Output Stages	158
5.6	Voltage Comparators	164
5.6.1	Single Ended Comparators	166
5.6.2	Differential Comparators	169
5.6.3	Bandwidth Considerations	171
5.6.3A	Reset Cycle	171
5.6.3B	Compare Cycle	175
5.7	Noise in Analog Circuits	176
CHAPTER 6 - TIME INTERLEAVED CONVERTER ARRAYS		180
6.1	Comparison with Other Methods	180
6.2	Error Analysis	187
6.2.1	Noise and Distortion Error Power	187
6.2.2	Spectral Distribution of Noise and Distortion	197
6.3	A Monolithic Time-Interleaved Converter Array	203
6.3.1	Chip Description	203
6.3.2	Experimental Results	230
6.3.3	Anticipated Characteristics of Implementations in an Advanced Process	239
6.4	Application to a Digital Signal Processor	240
CHAPTER 7 - CONCLUSIONS		244
REFERENCES		
APPENDIX A		
APPENDIX B		

HIGH SPEED CMOS A/D CONVERSION TECHNIQUES

PhD

William C. Black, Jr.

Dept. of Electrical
Engineering and
Computer Sciences

Chairman of Committee

ABSTRACT

Several high speed analog to digital (A/D) conversion techniques are examined for potential application in large-scale integrated (LSI) complementary-symmetry metal-oxide-semiconductor (CMOS) converters. These techniques, both new and old, are considered with regard to fundamental limitations and ease of practical CMOS realization.

One of the new techniques examined, is demonstrated in a high speed 7-bit A/D converter chip. In this chip, an array of A/D converters with interleaved sampling times are used as if they were effectively a single converter operating at a much higher effective sampling rate. This array method allows a reduction in the required circuit die size when compared with other high speed techniques (in the same fabrication technology) and may be used to achieve higher sampling rates than those obtainable with any single converter. Using a 10 μm metal-gate CMOS integrated circuit process, the chip described here achieves a conversion rate of 2.5 MHz with full linearity and displays a signal to noise ratio of 39 dB for a 100 kHz input signal. Conversion at reduced linearity is achieved up to 4 MHz.

CHAPTER 1

INTRODUCTION

The advantages of digital signal processing and storage techniques could be exploited in many new application areas if high-speed integrated circuit (IC) analog-to-digital (A/D) converters were available at a sufficiently low cost. As present day high-speed converters use techniques which require large IC die sizes and/or fairly exotic fabrication processes, these circuits have remained too expensive for many applications. Furthermore, these A/D techniques will be exceedingly difficult to integrate along with a VLSI digital signal processor, because of the large A/D die size and/or process requirements.

The objective of this research has been to study alternative high-speed A/D conversion techniques amenable to realization in MOS LSI. In this dissertation, several techniques will be described which may allow a substantial reduction in high-speed analog interfacing cost, and hence allow a number of new application areas to become economically realizable. In the second chapter, several digital application areas are described which are of this type. While this list is not intended to be exhaustive, it demonstrates the broad range of application areas that may be impacted by developments in inexpensive high-speed A/D converters. The third chapter covers a number of specific A/D conversion methods, both old and new, and discusses the inevitable tradeoffs between speed, IC chip area and power that exist for each converter type. Chapter 4 deals with the fundamental characteristics and limitations of precision elements in integrated circuit processes, while in Chapter 5, a set of basic analog circuit building blocks are described. The sixth chapter

describes the analysis, design and experimental performance of a new single-chip A/D converter, which can provide substantial improvements in effective conversion, speed, and a significant reduction in die area compared to alternative designs for similar specifications. Conclusions are presented in Chapter 7.

CHAPTER 2

HIGH-SPEED A/D CONVERTER APPLICATION AREAS

Perhaps the largest single application of high-speed monolithic converters would be in the area of digital video. Digital techniques are presently employed in most television studios, but are not yet economical in home television receivers. A number of new techniques could be applied in both areas if less expensive high-speed converters were available. Similarly, digital techniques would have profound effects in low-cost imaging systems and in high fidelity audio equipment if the required hardware could be sufficiently integrated. A few of the application areas that would be impacted by the development of inexpensive high-speed A/D converters are presented below.

2.1 Digital Time-Base Correction

A common problem in television studios is that of synchronizing broadcast sources. As different sources may be dubbed together or asynchronously switched, they must be synchronized to avoid annoying "page flipping." This problem is particularly severe when mobile stations or satellites are employed, as the signals may be variably phased and doppler shifted in addition to being asynchronous. Many systems that were used in the late sixties and early seventies required that each broadcast source use a rubidium atomic-standard, which would not drift more than a few microseconds per day. The broadcast signal was then electrically delayed [1] by an adjustable amount which would allow synchronization. Other methods employed conventional phase-locking techniques [2,3] but required that a communication line exist between

the studio and the broadcast source (typically a phone line). More recently, digital techniques have been used to achieve synchronization within a line (corrects for magnetic tape jitter, [4]), or within an entire frame [5,6]. This latter technique is the most generally useful, as gross changes in phase or doppler shift will not result in synchronization loss.

As line synchronizers employ relatively little memory (typically a few video lines of approximately 512 bytes each), the price of the unit is dominated by the cost of the A/D and D/A converters. Thus, if an A/D and D/A subsystem could be made with a sufficiently small die area, a monolithic line synchronizer should be relatively easy to implement. This device would have possible use in nearly all analog video recorders, if it could be made at a low enough cost.

Although frame synchronizers employ a fairly large digital memory (typically about 300 Kbytes), the overall cost is still affected by the complexity of the analog interfacing. While a reduction in frame synchronizer cost would have some impact on television studios, extensions of the technique into home systems with multiple broadcast sources (multiple channels on a split screen for instance), appear possible.

2.2 Digital Video Transmission and Storage

Digital transmission and storage of video information is somewhat involved, due to the relatively high bit rates required. A number of advantages, however, could be obtained from the use of this technique.

A studio which employed a digital image transmission system between various types of equipment, is more easily adapted for use with different video standards than equipment designed for operation with one standard

only. As Europe uses two incompatible standards, PAL and SECAM, which are also incompatible with the American NTSC standard, commonality of equipment is presently next to impossible. A studio which was entirely digital could be adapted to any analog standard just prior to transmission [7]. In addition, this technique would allow simple interfacing to digital image processors and storage devices, and reduces the hardware requirements of digital frame synchronizers and dubbing equipment as the encoding has already been performed.

Digital video transmission via satellite is under active consideration, as in certain cases it will result in a reduction in the required satellite transmission power [8]. Further, this technique may allow a reduction in receiver sensitivity requirements and hence potentially lower cost in direct "satellite to home" broadcast applications.

Digital video recording methods offer many advantages over analog techniques, including high picture quality, little or no video degradation after multiple program generations and no need for adjustment. This technique has been commonly used with small disk or semiconductor memories, and has recently been demonstrated with a fairly standard magnetic tape [9]. This technique would be highly desirable in home videotape players, if a method could be obtained for reducing the required bit rate. A number of techniques have been investigated, including sub-Nyquist sampling [10] and differential coding [11]. The first method takes advantage of the statistical properties of most video images to allow a reduction in the minimum required sampling rate, although comb filtering in the reconstructed waveform is required, and some types of image aberration may occur. The second method of bit rate reduction utilizes psychological perception characteristics to reduce the required number

of encoded bits. As the techniques which allow the greatest reduction in bit rate are also the most hardware intensive [12], the optimal recording economy will be dictated by a compromise between the two extremes. An integrated video codec and a simple video processor would have profound effects upon the overall cost of a digital video recorder, and quite possibly enable this technique to be applied in even low-cost units.

2.3 Real Time Image Processing

Digital image processing has typically been used in only a few application areas, because of the relatively high costs involved. It is, for instance, routinely used in the enhancement and correction of satellite photographs. This technique is rarely used in real time systems, however, as the analog converter and signal processor bandwidths are usually quite high, and as a result quite expensive. If the cost of these components could be reduced, however, real time image processing could be used in even low level application areas. One of the most discussed application of this type, is the commercial television receiver. By utilizing fairly straightforward techniques, vertical and horizontal image enhancement, low level coring¹, high level clipping and adaptive gain control are all easily accomplished [13]. Further, digital color correction and control is fairly simple to implement. All of these techniques appear to be possible in a one or two chip codec-video processor system, which could be used alone or with shared components, in for example, a digital video recorder.

¹Coring is a nonlinear process which removes both low amplitude noise and signal about a signals baseband. This function is typically used to "crispen" both horizontal and vertical details.

2.4 Digital Output Monolithic Imager

A semiconductor imager which employed an on-chip digital encoder would have considerable reduced hardware interfacing requirements when used in a digital video system. If the A/D converter could be made sufficiently small, some degree of image processing may even be economical on-chip. This would allow, for instance, sophisticated digital motion detectors to be realized in as few as one or two chips.

2.5 Wide Dynamic Range Digital Audio

Conventional audio recording methods have difficulty in achieving a consistently wide dynamic range (> 75 dB), because of their susceptibility to mechanical imperfections and electronic noise. Recently PCM encoded audio has become popular as an expensive but high performance alternative to analog recording methods [14,15]. In this technique, audio input signals are sampled at typically a 40-50 kHz rate with at least 14 bits of resolution [16]. This data is then stored on a high bandwidth recorder (typically a modified video recorder), in a format which allows considerable immunity to tape imperfections and fallout [17,18].

In adapters which allow digital audio recordings to be made on a video tape player (which can still be used for analog video recording), the codec and anti-alias output smoothing filters dominate the unit cost. Using improved monolithic conversion techniques and switched capacitor filters [19] of sufficiently wide dynamic range [20], the entire adapter function could be reduced to a 2-4 chip circuit. This should allow a considerable reduction from the present \$2000 unit price [21], and increase the popularity and use of this technique.

Digital encoding of audio would also allow the use of improved equalizers and mixers [22], which could make ever increasing use of audio-band digital signal processing chips [23].

CHAPTER 3

CHARACTERIZATION AND DESIGN OF HIGH SPEED

MONOLITHIC A/D CONVERTERS

In this chapter the characterization and design of high speed monolithic A/D converters is discussed at length. In Section 3.1 the common methods for evaluating converter performance are shown, including both quasi-static and dynamic input signal tests. Sections 3.2-3.4 discuss the flash, successive approximation and pipeline converter methods, while Section 3.5 demonstrates hybrid, or multiple converter techniques.

3.1 Characterization of High Speed A/D Converters

The evaluation of high speed converter performance (whether anticipated or measured) is an important aspect of the design and use of these devices. The tests discussed in this section cover both quasi-static and dynamic input conditions and are generally useful over a wide range of application areas.

3.1.1 Quasi-Static Characteristics

Perhaps the most obvious A/D converter test is that of DC linearity. In this test, a slow ramp is applied to the input of the A/D converter as shown in Fig. 3.1. The output of the converter is then reconstructed with a high-quality D/A and compared with the input signal via an X-Y recording mechanism. The resulting pattern for an input signal which extends over the entire dynamic range of the A/D is the converter transfer characteristic. An example of this characteristic for an ideal A/D converter is shown in Fig. 3.2a, while characteristics which result

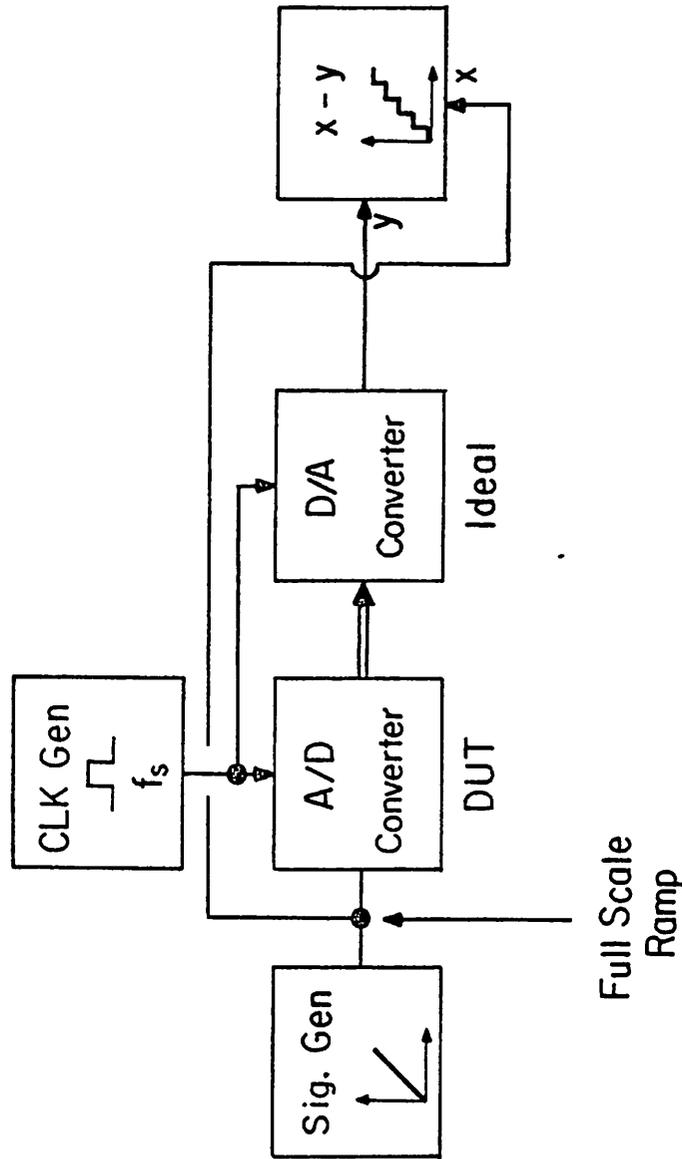


Fig. 3.1 Measurement of an A/D converter's transfer characteristic.

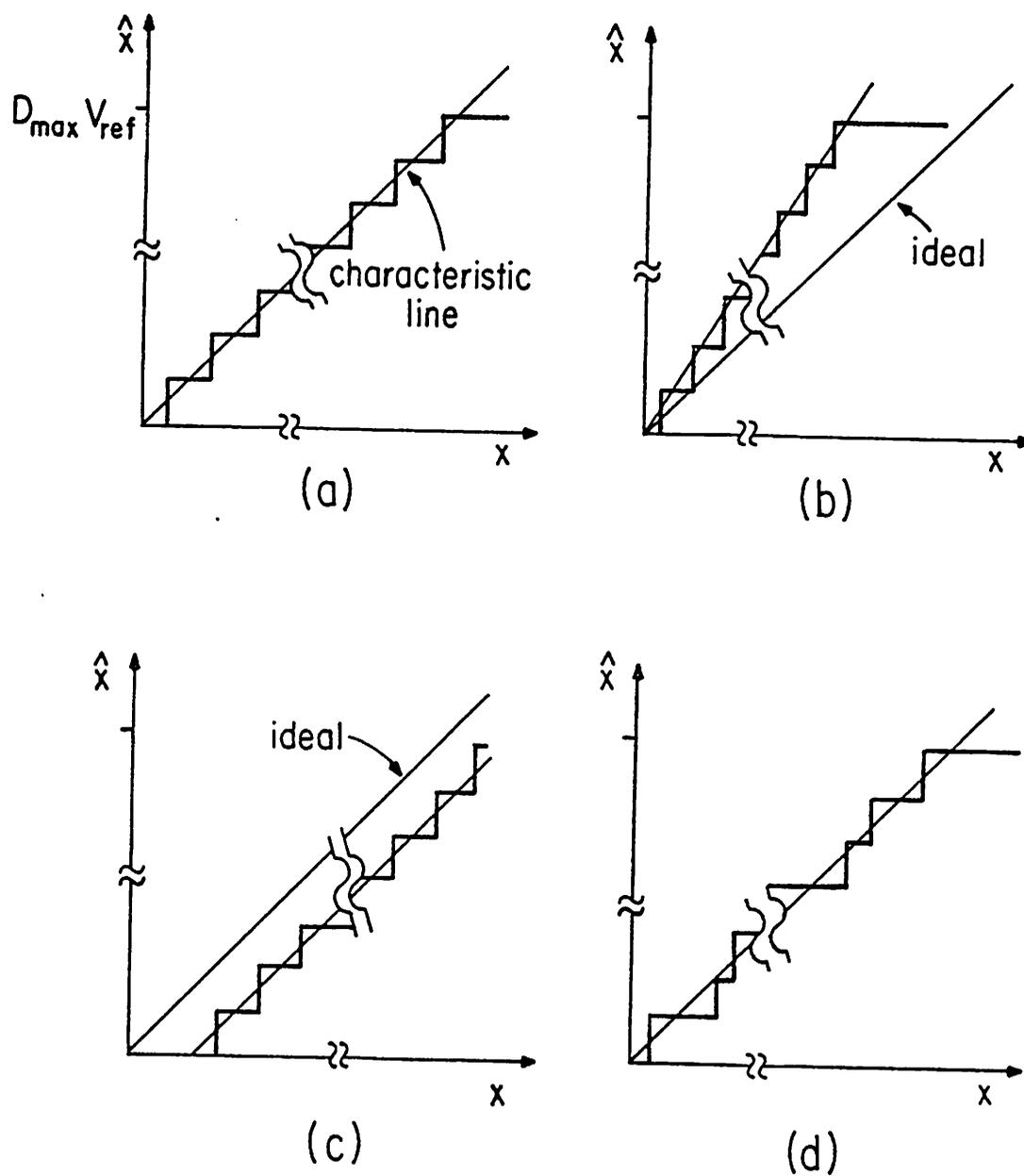


Fig. 3.2 Quantization characteristics of linear A/D converters. (a) Ideal. (b) Gain error. (c) Offset error. (d) Nonlinearity.

because of various converter nonidealities are shown in b through d. As shown in Fig. 3.2a, the "characteristic line" of a converter is defined in terms of the first and last transition points within the converters dynamic range. The slope of this line indicates the converter gain, while the intercept with the X-axis indicates converter offset.¹ As illustrated in Fig. 3.2b and Fig. 3.2c, both gain and offset are somewhat variable in many converter types and in real circuits may result in a reduction in the converters dynamic range. In most applications, however, a small constant error in either the gain or offset is not considered a performance limiting problem. A far worse nonideal effect is that demonstrated in Fig. 3.2d. In this figure a transfer characteristic is shown which exhibits nonlinear behavior. This nonlinearity at low frequencies is invariably caused by simple precision component mismatches within the A/D converter, and is usually described in terms of the integral and differential nonlinearity specifications.

The integral nonlinearity of a converter is defined to be the maximum deviation of the transfer characteristic, from the characteristic line, and is usually expressed in terms of converter LSBs. Similarly, the differential nonlinearity, is defined to be the worst case step size error relative to 1 LSB. In other words, if \hat{x} is the quantized voltage value of x , and if the positive going input voltage which is just high enough to create an A/D output code of D_n is defined as $V(n)$, then the linearity specifications are given by

$$INL \triangleq \max |x - ax - b| \Big|_{x_{\min}}^{x_{\max}} \quad (3.1)$$

¹Alternately, the characteristic line is sometimes defined with respect to the midpoints of the first and last conversion steps [24].

and

$$\text{DNL} \stackrel{\Delta}{=} \max |1 \text{ LSB} - V(n+1) + V(n)| \begin{matrix} D_{\text{max}}-1 \\ D_{\text{min}} \end{matrix} \quad (3.2)$$

where a is the converter gain and b is the offset. Note that when the DNL exceeds 1 LSB, either missing codes or nonmonotonocities may result, as shown in Fig. 3.3. These types of converter errors are often regarded as catastrophic in control systems², however in communication systems where the effects of nonlinearities are usually statistical in nature, nonmonotonocities may not result in a significant degradation in performance unless they occur often or are near the quiet channel bias point. In these systems the INL specification tends to be more important, as in many converter types it more accurately reflects the nonlinear performance over the entire input range.³

It is interesting to note that in practical test situations the foregoing analysis may usually be simplified. For most binary weighted converters the entire transfer function may be inferred by measuring the input voltage required for each major bit transition (from 0...0111 to 0...1000 for example) or by evaluating the differential nonlinearity at

²In a feedback system, nonmonotonic conversions may result in oscillatory behavior about a nonmonotonic region. In mechanical servo systems, however, a high frequency oscillatory signal is often applied as a matter of course, to maintain the system in a state of "dynamic" rather than "static" friction [25]. In situations such as these, system performance will not necessarily be degraded unless the nonmonotonic region is large.

³In all converters it is the case that if the INL < 1 LSB then the DNL < 1 LSB. The converse, however, is not necessarily true. In some converter types, most notably those employing linear strings of precision components, a grossly nonlinear "bow" like characteristic may still exhibit a DNL < 1 LSB.

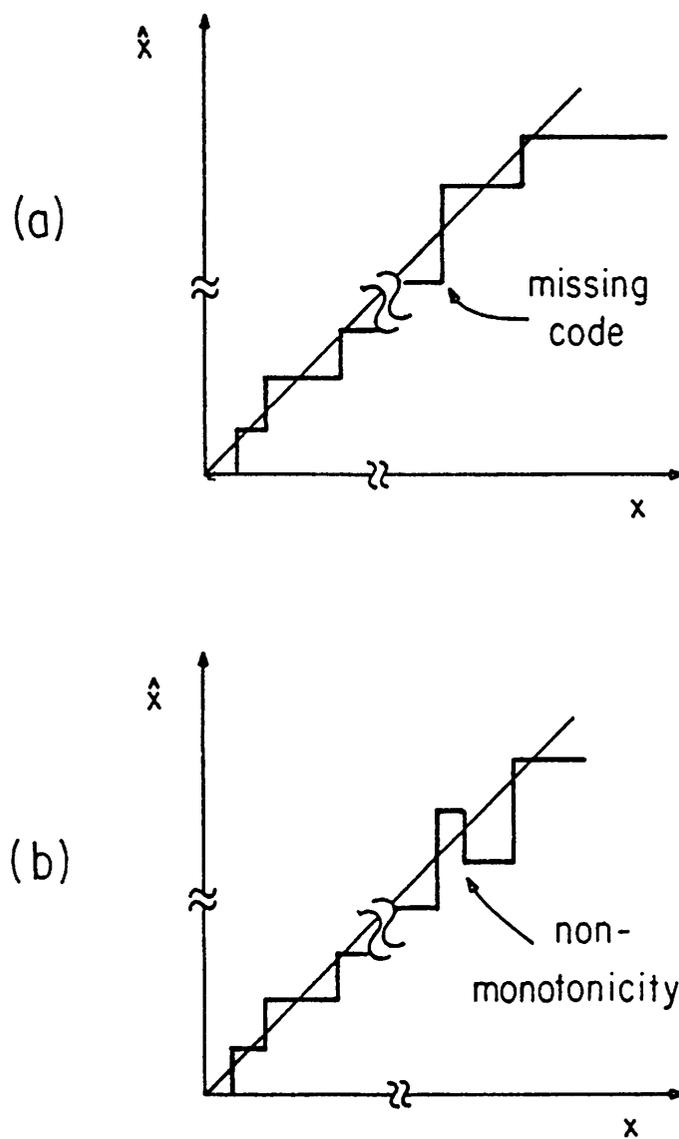


Fig. 3.3 Quantization characteristics for converters with a $DNL > 1$ LSB.
(a) Missing code. (b) Non-monotonic behavior.

an odd number of points across the entire input range [26].

3.1.2 Dynamic Characteristics

Dynamic tests of a high frequency A/D converter are generally more useful in the evaluation of a device than static tests, as they more accurately reflect the converter's characteristics during real operation. Although different applications may require fairly unique converter test procedures, several types of dynamic tests are generally useful. Four of these tests are described here, they are: single-frequency signal to noise ratio (SNR), noise power ratio (NPR), code density test and gain and phase linearity.

The first of these, the single-frequency signal to noise ratio test, is used extensively in this and subsequent chapters as a performance measure for a variety of converter types. In this test, a typically full scale sinusoidal signal is fed into the A/D to be measured, as shown in Fig. 3.4. The output of the A/D is then reconstructed with a high quality digital-to-analog converter, and bandlimited to the Nyquist rate of the sampler or below. The ratio of the measured power at the original signal frequency to all other measured power is the single frequency signal to noise ratio.⁴ The power at nonfundamental frequencies, or the "error power," is due to the A/D converter displaying both finite resolution and limited linearity. As shown in Fig. 3.5, the converter quantizing error may be seen to be

$$\epsilon(x) = \delta(x) \tag{3.3}$$

⁴This test as indicated, will not detect delay distortion terms at the fundamental frequency. For these to be included, the notch filter of Fig. 3.4 must be made phase as well as frequency selective.

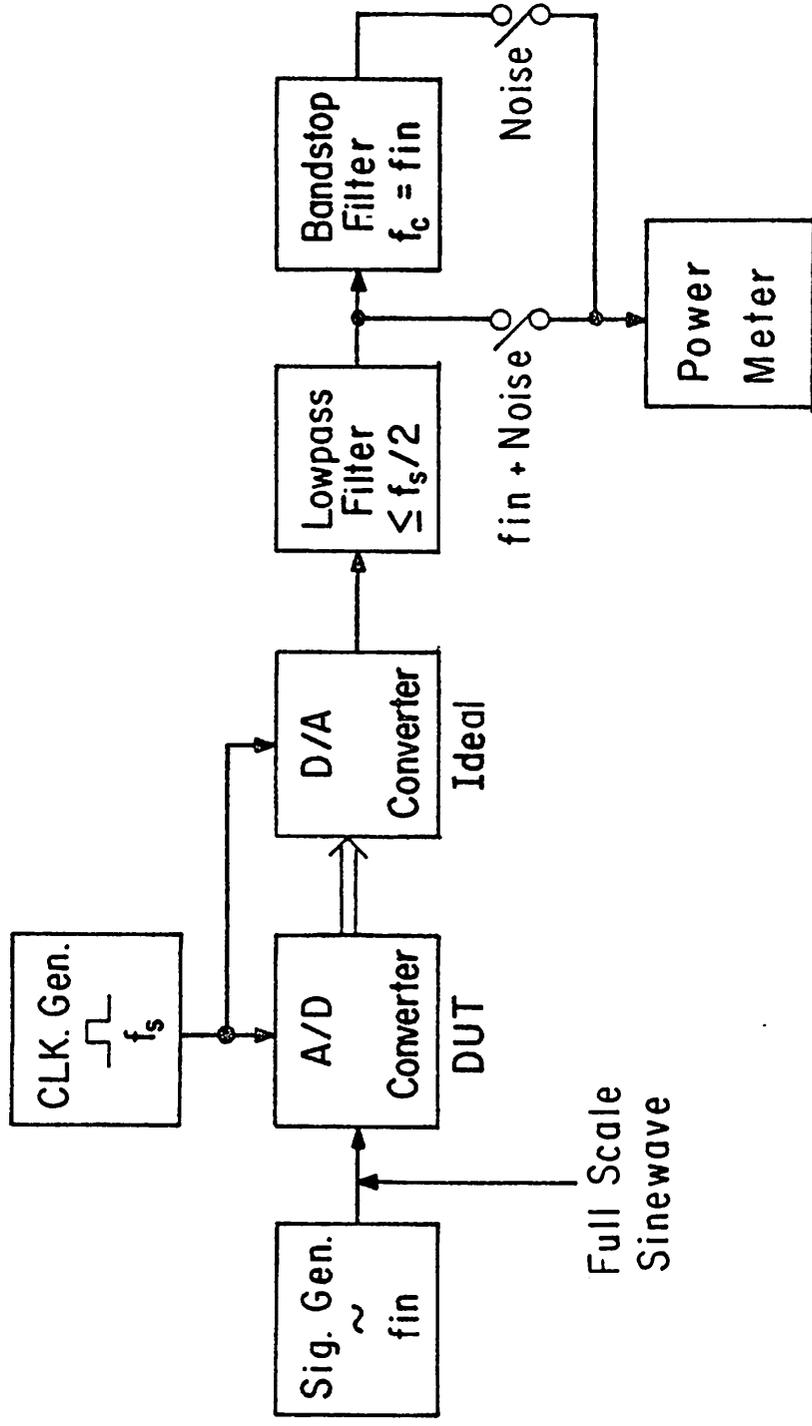


Fig. 3.4 Single-frequency signal to noise ratio test.

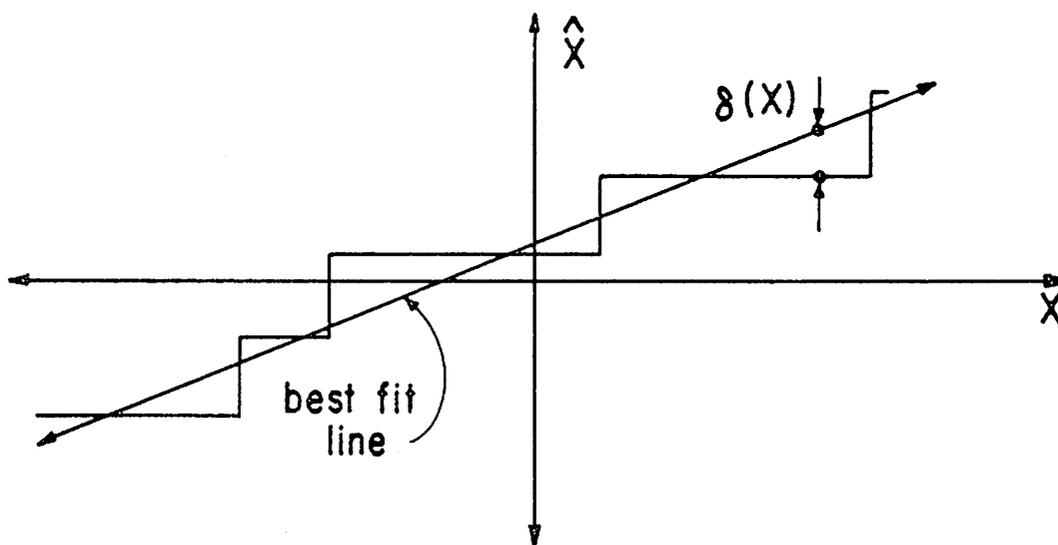


Fig. 3.5 Quantization and nonlinearity error in an A/D converter.

where

$$\delta(x) = \hat{x} - (ax+b) \quad (3.4)$$

and where a and b are error minimization terms corresponding to the best fit converter gain and offset. This results in an average error power (referred to 1 ohm) of simply

$$E_p = \langle \delta^2 \rangle \quad (3.5)$$

where $\langle \delta^2 \rangle$ indicates the mean or expectation value of δ^2 , and where a and b in Eq. (3.4) have been chosen to minimize E_p for a specified input distribution. In general, for an input signal with probability density $p(x)$,

$$\langle f(x) \rangle \triangleq \int_{-\infty}^{\infty} f(x)p(x)dx \quad (3.6)$$

$$a \triangleq \frac{\langle \hat{x}x \rangle - \langle \hat{x} \rangle \langle x \rangle}{\langle x^2 \rangle - \langle x \rangle^2} \quad (3.7)$$

and

$$b \triangleq \langle x \rangle - a \langle x \rangle \quad (3.8)$$

Note that a and b are potentially quite different from the gain (a) and offset (b) described previously, as the new values take into account all converter nonlinearities and the probability distribution of the input signal, while the old values were only a function of the converter transition points at the very ends of the transfer characteristic.

The determination of these parameters, a and b , is only necessary during computer simulations of converter performance and is not required

experimentally. This S/N ratio test is particularly amenable to computer simulation in that it allows a dynamic performance measure to be obtained without the need for frequency domain calculation. All of the necessary calculations may be done in terms of input probability functions and specific converter nonidealities, as will be outlined in a subsequent section.

Another useful dynamic converter measurement, is the noise power ratio test. In this case a bandlimited gaussian noise source is used as the input to the A/D converter, as shown in Fig. 3.6. The power from the reconstructed output in a selected frequency band is measured, and then measured again with the same frequency band notched out of the input source. The ratio of the two power measurements indicates the relative performance level of the converter, as dynamic nonlinearity, quantizing and aperture effects will tend to spill noise power into the notched-out band [27]. This tends to be a more complete test of a converter than a sinusoidal SNR measurement of a single input amplitude. As a rule, however, it is a more difficult test to perform and is computationally more difficult to simulate, as both time and frequency domain effects must be included in the simulation. This test is relatively common for evaluation of voiceband communication systems, but is rarely discussed in the context of high frequency or video bandwidth converters.

Yet another useful converter measurement is the output code density test. In this measurement, a signal with a known probability density is applied to the converter under test as shown in Fig. 3.7a. The output of the n -bit converter is then used as a memory pointer to indicate which of the 2^n memory locations is to be incremented. Assuming each location

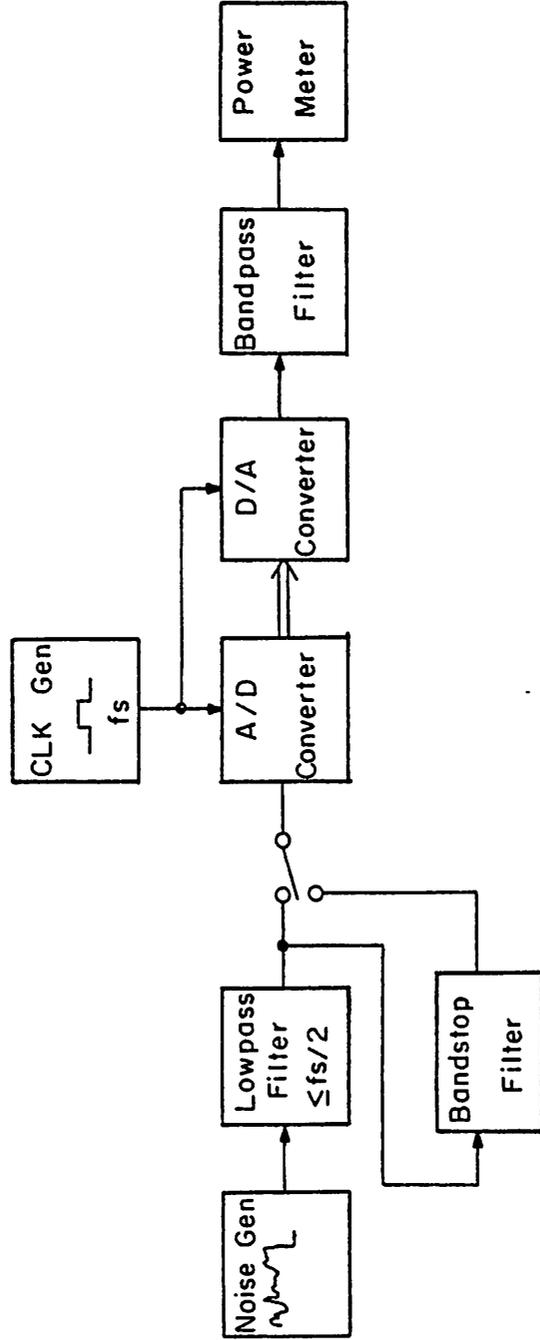


Fig. 3.6 Noise power ratio test.

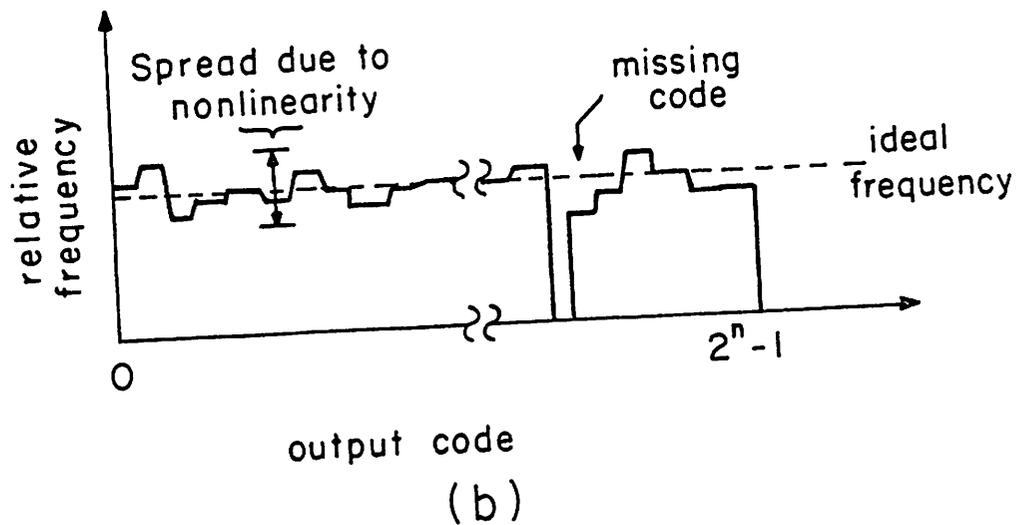
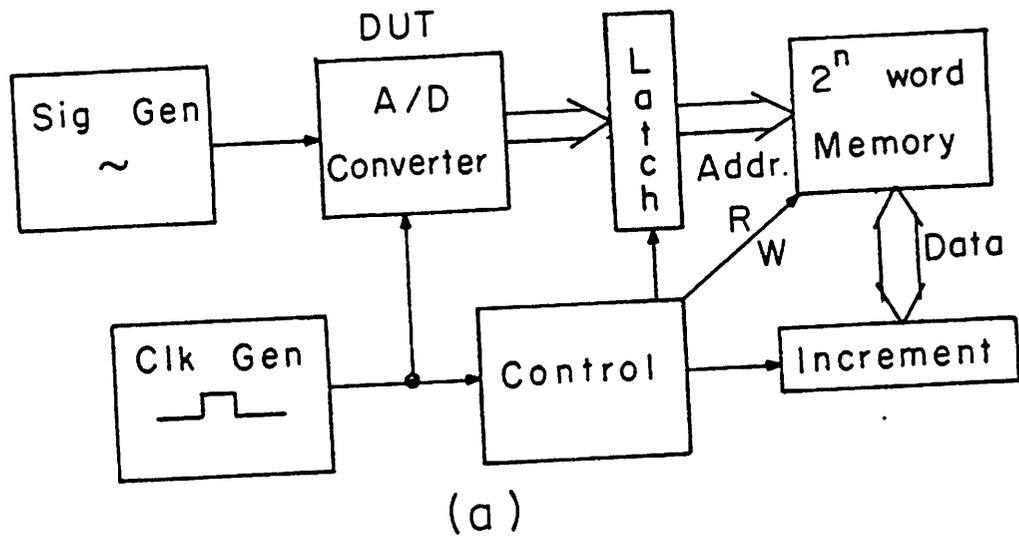


Fig. 3.7 Code density test. (a) Test configuration. (b) Output code density of a nonideal A/D converter for a full scale uniform probability input signal (such as a sawtooth).

was initially zeroed, this makes each location a counter which reflects the relative frequency of a specific output code. As converter nonlinearities will tend to distort the distribution of codes, various types of converter behavior may be deduced from the pattern of memory counts. An example of this is shown in Fig. 3.7b. This figure indicates a typical output frequency map for a nonideal A/D converter with a sawtooth input waveform. Ideally, this input signal should produce a constant frequency of occurrence. Note that nonlinearities produce a nonuniform distribution, while missing codes are indicated by the lack of any counts at all.

A problem with this technique, is that nonmonotonic behavior may not result in a variation of count distribution from that of a circuit which is operating properly. In fact, for certain input distributions, several of the A/D output bits may actually be switched without any change in apparent performance based upon this test. Further, this test can be relatively difficult to perform with high speed converters, in that the memory must be read, the contents incremented and the results subsequently stored all in one conversion cycle. In addition, this test can take a significant time to perform in high resolution converters as a large number of counts, on the average, must be obtained in each memory location. As a rule this converter test may be used to supplement the tests presented previously, but is almost never used alone.

The last test to be described, the differential gain and phase measurement is quite important in color video systems and is included here for completeness. In this test a high frequency sinusoid is impressed upon a linear ramp, which usually has a period equal to the line scan time of an NTSC (in North America) or PAL or SECAM (Europe) video signal. This composite signal is fed into the A/D under test, as

shown in Fig. 3.8, and subsequently reconstructed. The high frequency component in the output is then separated from the ramp, and compared in gain and phase to the same signal in the original composite input, as a function of position along the ramp. As the chrominance component of a color video signal is phase encoded atop the luminance signal, a variation in converter response with input amplitude may result in chromatic aberration at different luminance values. A variety of video relevant converter errors may be detected with tests of this type [28,29].

3.2 Flash Converter Technique

The most straightforward procedure for realizing a fast n -bit A/D conversion, is by performing 2^n simultaneous voltage comparisons with 2^n linearly graduated voltage sources. This general "flash" technique is shown in Fig. 3.9. As is evident from the figure, a resistor string is typically used for the generation of unique tap voltages, while a gate array (possibly in conjunction with a ROM) is used for the required 2^n to n line conversion. This method has long been used in discrete and hybrid circuits, but has been extremely expensive to implement at resolution levels above 4 or 5 bits. Recently flash circuits have been designed in monolithic bipolar form in initially 4-bit [30] and subsequently 8-bit [31] configurations. Thus far only a single MOS flash circuit has been reported [32], it being a 6-bit converter fabricated in a silicon on sapphire technology.

Depending upon the choice of technology, a number of potential problem areas must be considered in the flash converter design. Several of these including some not-so-obvious ones, are discussed below.

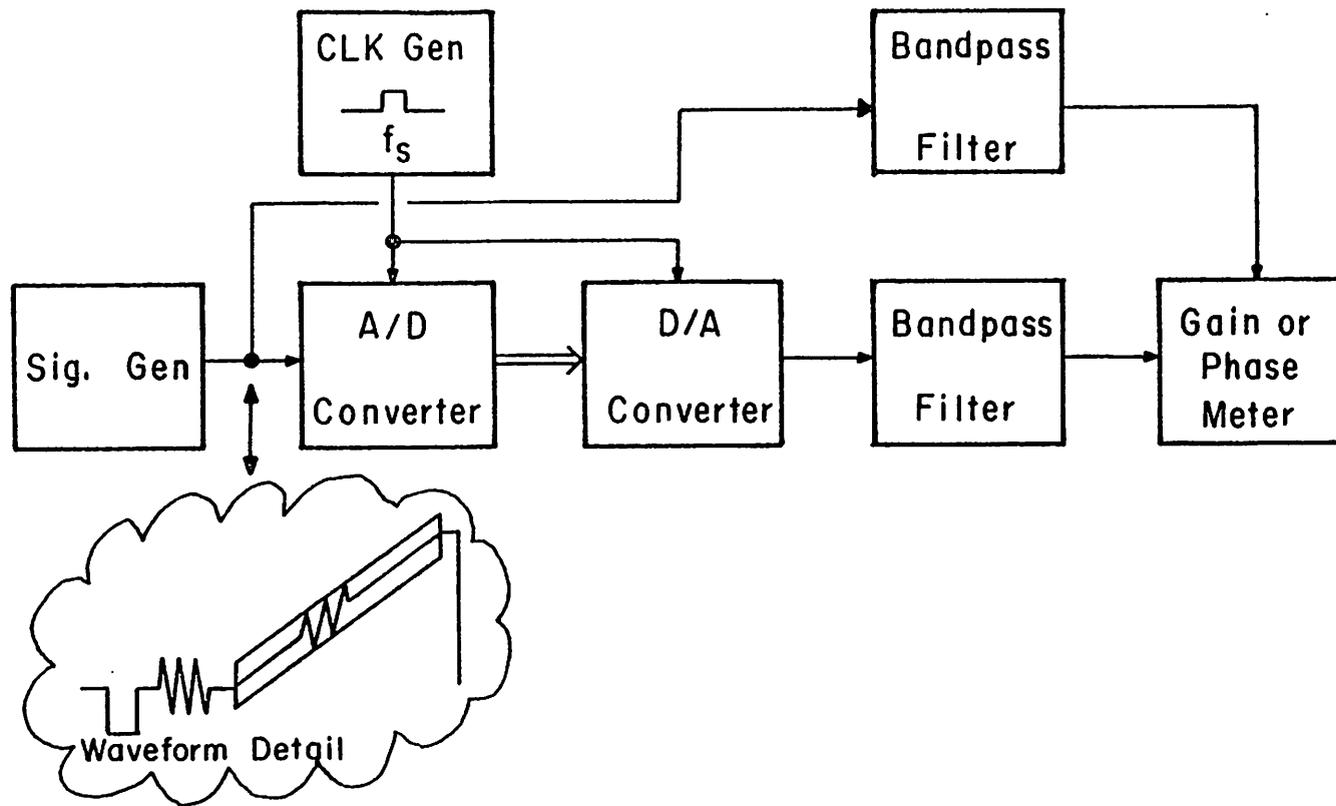


Fig. 3.8 Gain or phase linearity test.

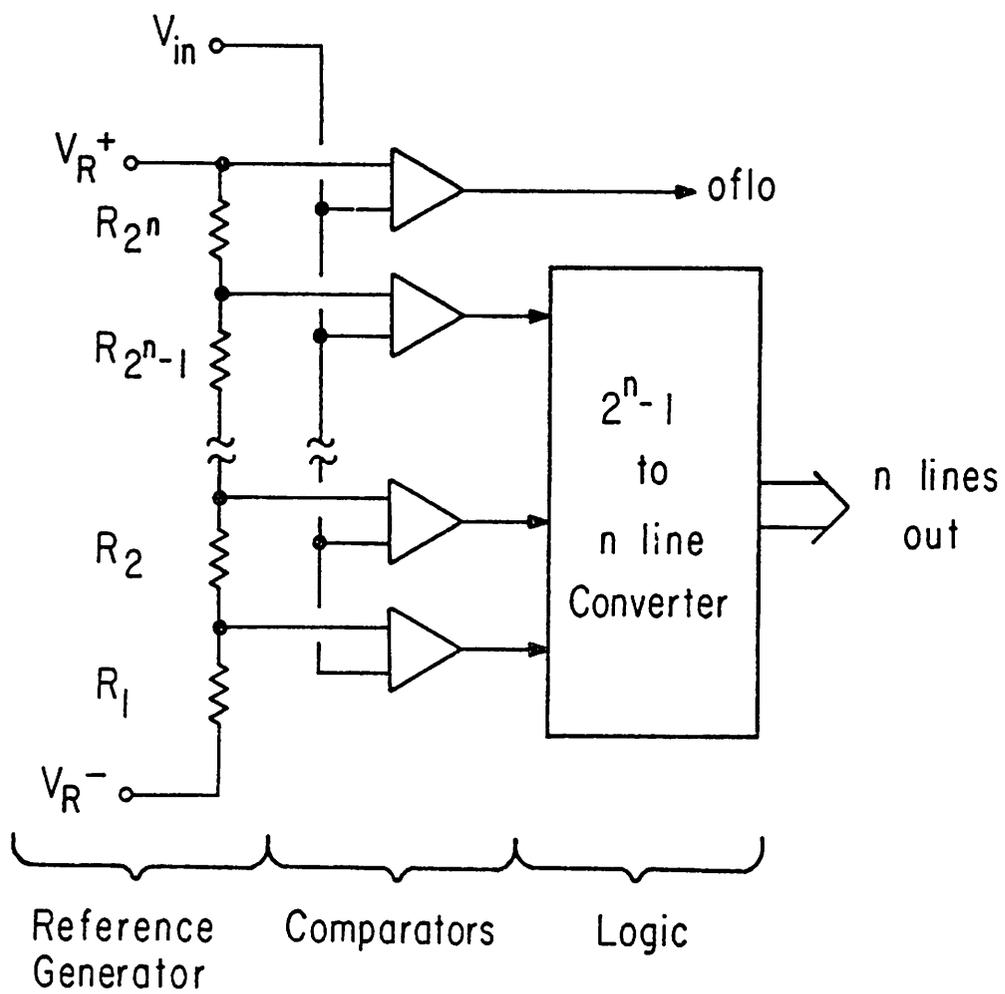


Fig. 3.9 Flash converter method.

3.2.1 Comparator Input Bias Currents

In flash circuits which are fabricated in a bipolar technology, the effect of comparator input bias current can be quite severe. This current may result in a substantial nonlinearity in the converter transfer characteristic if it is not compensated for in the chip design. This is particularly true in high resolution converters which must use a large number of comparators, and in converters fabricated in a high speed process which usually will display a low bipolar β . This problem may be reduced by using an input bias cancellation technique [33-35], although this usually requires the use of a pnp device which is typically slow or not available at all (triple diffused processes⁵). Alternately high resistor string bias currents, single or few tap trimming, and the use of a tapered resistor string [36] may be used. None of these latter techniques are particularly desirable, however, in that they force a high power dissipation, increased testing complexity (or use, depending upon where trimming is performed), and very accurate knowledge of input bias currents over temperature. This entire problem may be eliminated by the use of an MOS input device on the comparator, either as part of a combination bipolar/MOS circuit [37,38] or as part of an all MOS design. Both MOS and bipolar designs, however, will be susceptible to the problems discussed below.

3.2.2 Resistor Matching

Assuming that the reference source is just a string of series resistors, the voltage at any tap "k," may be given by

⁵Both vertical and lateral pnp devices are available in a triple-diffused (3-D) bipolar process, however, the vertical devices are all common collector, and the lateral devices usually exhibit a very low β due to the absence of a buried layer.

$$V_K = \frac{\sum_{i=1}^k R_i}{\sum_{i=1}^{2^n} R_i} \cdot (V_{\text{ref}}^+ - V_{\text{ref}}^-) - V_{\text{ref}}^- \quad (3.9)$$

where it is assumed that the k-th tap is above the kth resistor, and n is the bit resolution of the converter. This indicates that the integral nonlinearity about the nominal line may be given by (in LSBs)

$$\text{INL} = \max \left| \frac{\sum_{i=1}^k R_i}{\sum_{i=1}^{2^n} R_i} - \frac{k}{2^n} \right| \cdot 2^n \quad \begin{matrix} k = 2^n - 1 \\ k = 0 \end{matrix} \quad (3.10)$$

and the differential nonlinearity by

$$\text{DNL} = \max \left| \frac{R_k}{\sum_{i=1}^{2^n} R_i} - \frac{1}{2^n} \right| \cdot 2^n \quad \begin{matrix} k = 2^n \\ k = 1 \end{matrix} \quad (3.11)$$

As the worst case linearity is actually a function of all of the component mismatches, the required component matching for a specified performance level tends to be probabilistic in nature. If the resistors are assumed to have some distribution of values, converter performance in various static and dynamic tests may be simulated. A computer program has been written to perform just this simulation. This program, which is listed in Appendix A, assumes a gaussian distribution of component variations where the standard deviation of error is given by the user. By modifying

the input probability distribution, worst case linearity or signal to noise performance may be determined. This program may also be used to simulate the performance of converter arrays, as discussed in Chapter 6.

Simulated worst case linearity error for 100 flash converters of 7-bit resolution, as determined by this program, is indicated in Fig. 3.10. In this figure, the average and standard deviation of the worst case linearity is shown for error about the best fit line, and error about the nominal characteristic. As is evident from the figure, a σ_R/\bar{R} of approximately 5% is required to obtain less than 1 LSB error about both lines most of the time. Calculated full scale SNR performance for the same set of converters is shown in Fig. 3.11. Note that converters which display an average 1/2 LSB INL, exhibit an average SNR of about 41.6 dB.

In real flash converters the effect of linear processing gradients is potentially as, or more severe, than seemingly random variations. This is particularly true since most converters of this type are designed with a linear, or folded linear layout. The effect of a linear resistor gradient on maximum converter error is shown in Fig. 3.12a for an unfolded converter. Similarly, the effects of a resistor gradient upon the error of converters which are folded once and twice, are indicated in Fig. 3.12b. This same linear resistor gradient has been used for simulation of full-scale SNR performance and is shown in Fig. 3.13 for each of the converter topologies mentioned. As is apparent from all of these figures, interdigitating or folding of the converter layout will substantially reduce the degradation of converter performance due to a linear resistor gradient, even if the "folding" is only done once or twice.

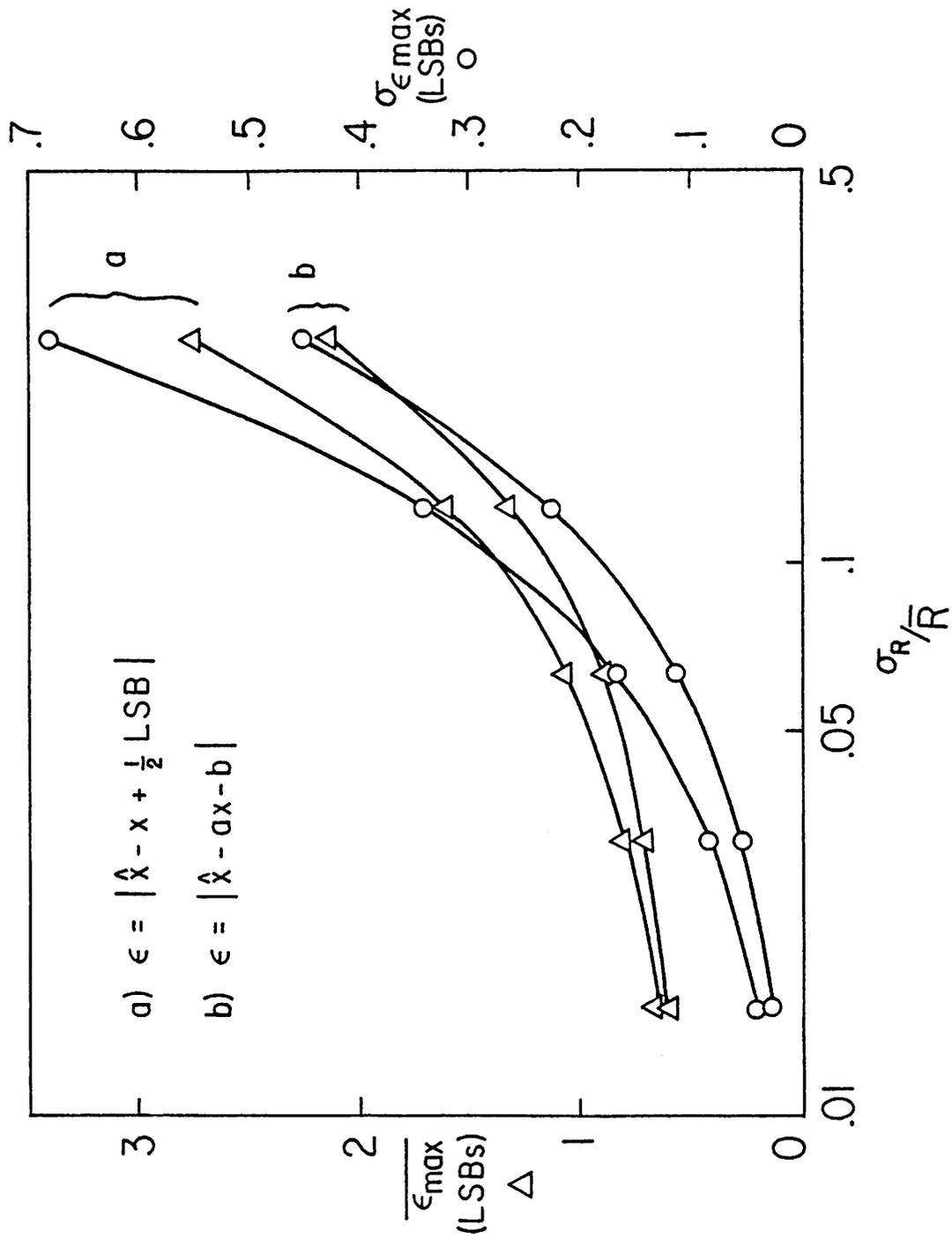


Fig. 3.10 Simulated maximum quantizing error of 100 7-bit flash converters as a function of resistor mismatch. Resistors are assumed to be normally distributed about their nominal value.

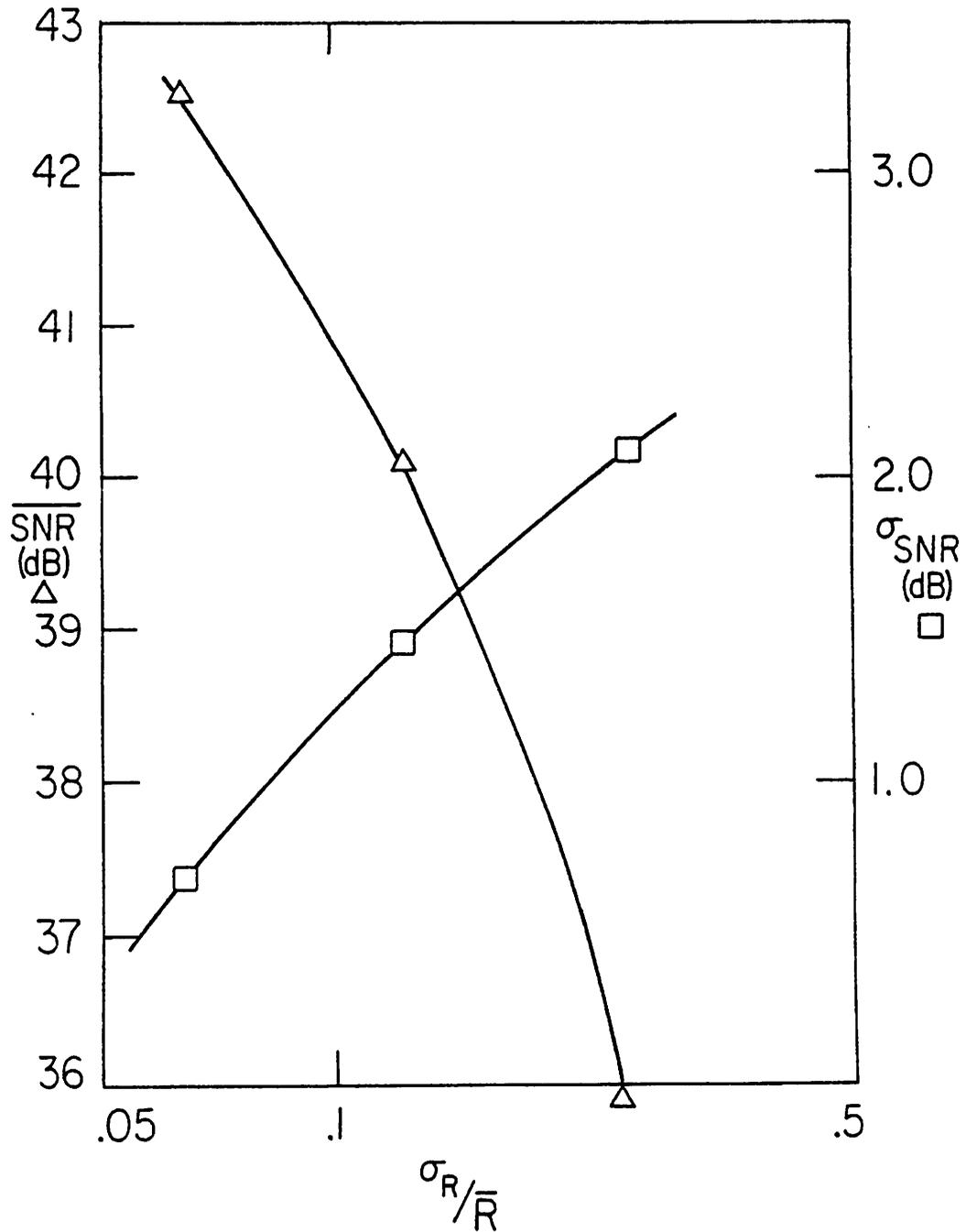


Fig. 3.11 Simulated SNR performance of 100 7-bit flash converters as a function of resistor mismatch.

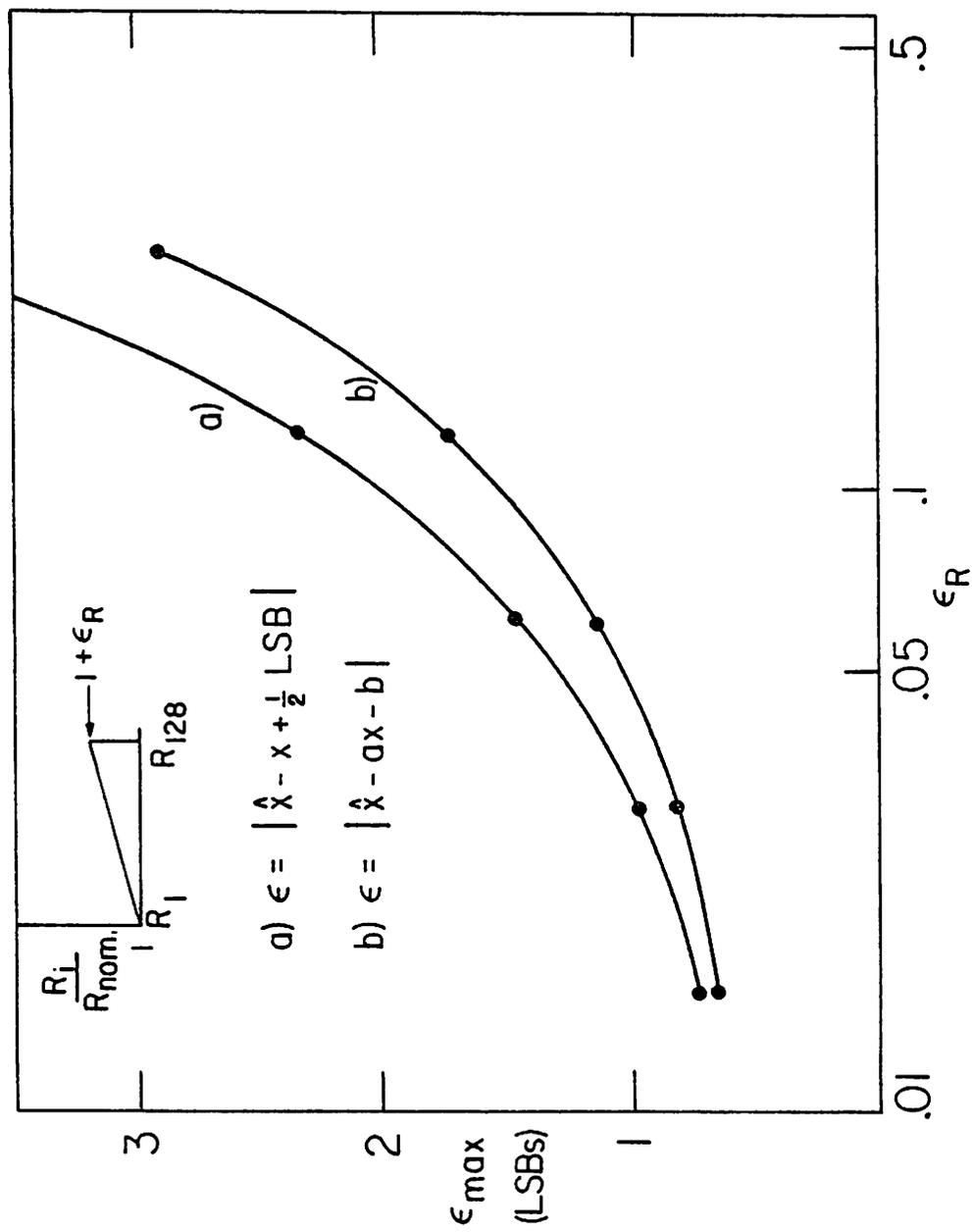


Fig. 3.12a Calculated maximum quantizing error of a 7-bit flash converter which has a linear resistor gradient.

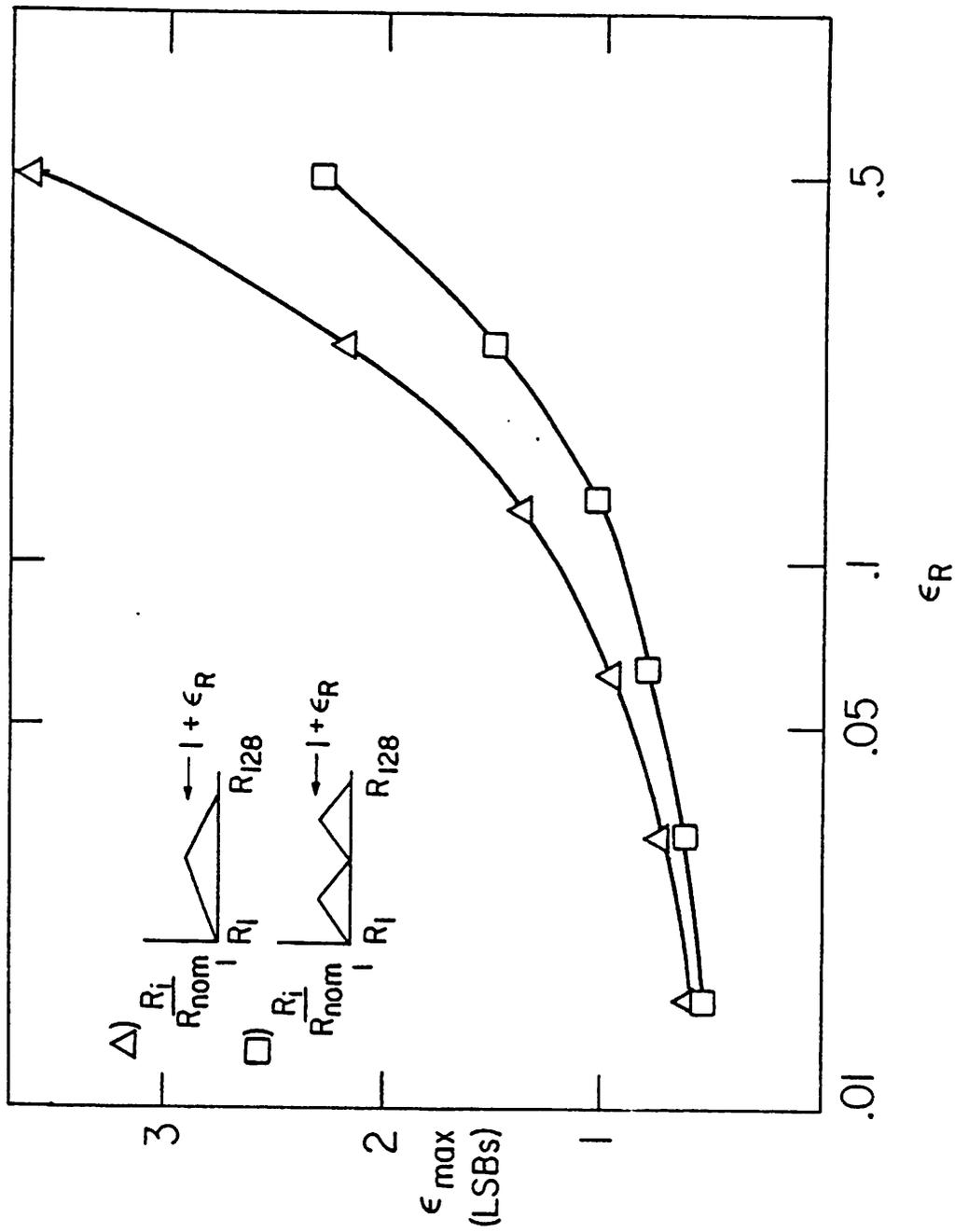


Fig. 3.12b Calculated maximum best-fit quantizing errors of "folded" 7-bit flash converters as a function of linear resistor gradient.

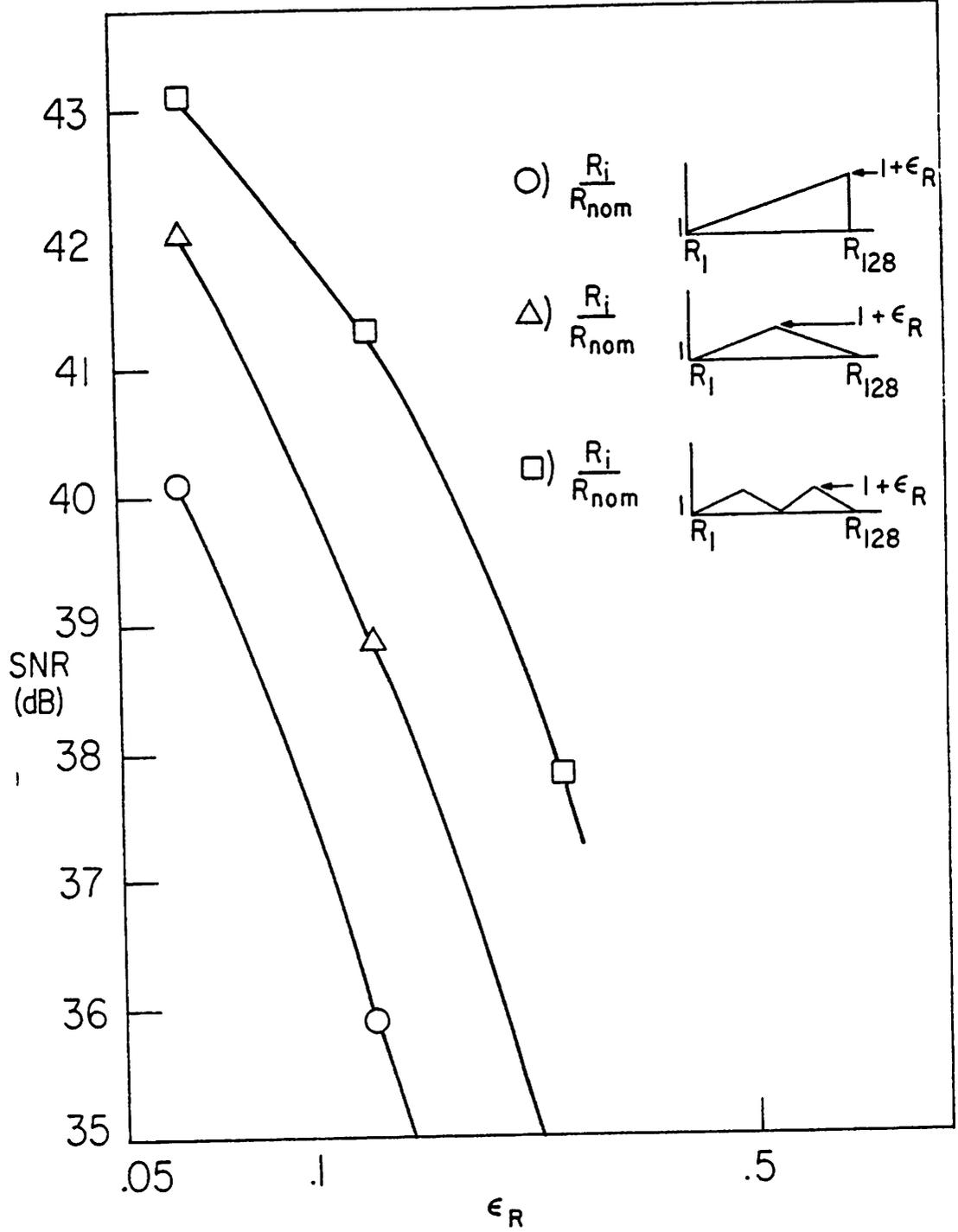


Fig. 3.13 Calculated SNR performance of 7-bit flash converters as a function of linear resistor gradient and "folding."

3.2.3 Comparator Intrinsic Offset

If the $2^n - 1$ comparators in a flash circuit exhibit a constant DC offset, the effect on the converter's DC transfer function will be a simple offset shift by the same amount. If, however, the comparators exhibit a variation in offsets, converter nonlinearity will result. As this is in addition to resistor nonlinearity, the ultimate matching requirements of the comparators and resistors for a specific performance level are actually coupled together. In fact, the linearity expressions are now given by (in LSBs)

$$\text{INL} = \max \left| \frac{\sum_{i=1}^k R_i}{\sum_{i=1}^{2^n} R_i} + \frac{V_{osk}}{(V_{ref^+} - V_{ref^-})} - \frac{k}{2^n} \right|_{\substack{k=2^n-1 \\ k=0}} \cdot 2^n \quad (3.12)$$

and

$$\text{DNL} = \max \left| \frac{R_k}{\sum_{i=1}^{2^n} R_i} - \frac{(V_{osk-1} - V_{osk})}{(V_{ref^+} - V_{ref^-})} - \frac{1}{2^n} \right|_{\substack{k=2^n \\ k=1}} \cdot 2^n \quad (3.13)$$

Note that for cases in which the DNL is greater than 1 LSB, a situation potentially much worse than a simple missing code may develop. If the combination of resistor nonlinearity and comparator offset is sufficiently large that comparator k indicates that $V_{in} > V_k$ but comparator $k-1$ still reads that $V_{in} < V_{k-1}$, a disallowed internal logic state will occur. Depending upon the specific implementation, this may result in

neither code being represented at the converter output. In fact, in circuits which employ XOR gates and a ROM, the truncated algebraic sum of the two codes is even possible.

If the comparator offsets in a flash circuit are assumed to be normally distributed with mean V_{OS} , the required offset variance σ_{OS}^2 may be determined for a real converter that is to have no missing codes a certain percentage of the time. The mean and variance of the sum of any two comparator offsets is given by simply $2V_{OS}$ and $2\sigma_{OS}^2$ respectively. A comparator offset induced missing code may develop if for any i , $|V_{OS} - V_{OS_{i+1}}| > 1$ LSB. Further, because there are nearly 2^n comparator pairs in an n -bit converter, the probability of a missing code occurring at all, is approximately 2^n times the probability of occurrence in any single pair. Thus, for example, if it is required that no missing codes occur 62% of the time in an 8 bit converter which has a 2V dynamic range, then it is straightforward to show that the required offset distribution must be such that,

$$3\sqrt{2\sigma_{OS}^2} \leq 1 \text{ LSB} \quad (3.14)$$

or

$$\sigma_{OS} \leq 1.87 \text{ mV} \quad (3.15)$$

This indicates that tight control of offsets is essential if good linearity is to be obtained with good yield. The requirement of 1.87mV can probably be met in bipolar circuits if care is exercised in the comparator design (as evidenced by [39]), but this may be difficult to achieve in MOS converters even with offset cancellation techniques.

3.2.4 Dynamic Ladder Tap Loading

A potentially severe problem in flash circuits, arises due to dynamic loading of the precision ladder structure. As each tap from the resistor string is ultimately connected to a comparator, any input charge-pumping which occurs during normal comparator operation may modify apparent resistor linearity. This charge-pumping may occur due to a comparator offset cancellation cycle, or from comparator re-generation. Dingwall has shown that charge-pumping from offset cancelled comparators (see Chapter 5) may be modeled as a lumped reactance, and further, that for comparators in which $n \geq 6$ the ladder may be represented as a distributed R-Z circuit, as shown in Fig. 3.14 [40]. He has analyzed this network with respect to network linearity, and found that the maximum fractional error ϵ_{\max} is given by

$$\epsilon_{\max} = [V(x_{\max}) - x_{\max} \cdot V_{\text{ref}}]/V_{\text{ref}} \quad (3.16)$$

where x_{\max} is the position of maximum error described by

$$x_{\max} = \frac{1}{K} \cosh^{-1} \left[\frac{V_{\text{ref}} - V_{\text{in}} \cdot k \cdot e^{-Kx_{\max}}}{2 \cdot K \cdot C_2} \right] \quad (3.17)$$

The function $V(x)$ in Eq. (3.16) is given by

$$V(x) = \frac{V_{\text{ref}}(e^{Kx} - e^{-Kx})}{e^K - e^{-K}} + V_{\text{in}} \left[1 - e^{-Kx} - \frac{1 - e^{-K}}{e^K - e^{-K}} \right] \quad (3.18)$$

where

$$K = 2^n \sqrt{\frac{R_{\text{TAP}}}{|Z(\omega)|}} \quad (3.19)$$

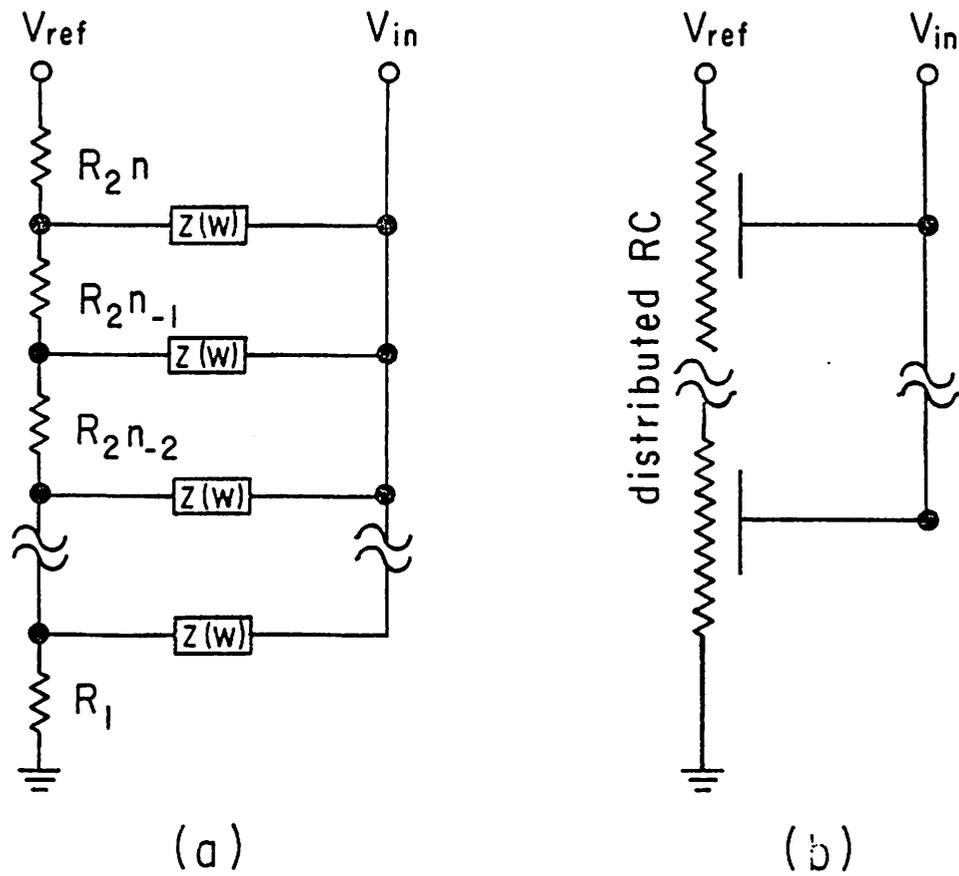


Fig. 3.14 Effect of comparator reset cycles on flash converter linearity. (a) Discrete model. (b) Distributed equivalent for a large number of tapes.

$$C_2 = \frac{V_{\text{ref}} - V_{\text{in}}(1 - e^{-K})}{e^K - e^{-K}} \quad (3.20)$$

and

$$C_1 = (C_2 + V_{\text{in}}) \quad (3.21)$$

In the above equations, R_{TAP} is the resistance between taps in the reference ladder, $|z(\omega)|$ is the comparator input reactance at the comparator reset frequency and the worst case V_{in} for the ladder shown in Fig. 3.14 is at $V_{\text{in}} = 0V$. The tap loading requirements ($z(\omega)/R_{\text{TAP}}$) for 1/2 LSB linearity errors due to this source alone, are shown in Fig. 3.15 for commonly occurring resolution levels. Note that for a constant comparator input capacitance, the required reference power goes as approximately the square of the resolution level, i.e., 8 bit resolution converters will require 16 times the reference power of a 6 bit device for an equivalent conversion rate. This will have profound effects upon non-SOS converters which will suffer from non-negligible input parasitic capacitance.

An alternative flash operation cycle would be to only offset cancel the comparators during system dead times, rather than between every conversion cycle [41]. This should be feasible in video systems which normally do not use information transmitted during the horizontal line sync., or frame sync. periods, although this may not be possible or convenient in many data acquisition systems.

3.2.5 Clock Skew Between Comparators

If the comparators within a flash converter are not clocked at precisely the same time, a clock induced phase skew will develop between different comparators. As will be shown in Chapter 6, this results in a form of aliasing distortion, creating signal related noise spikes

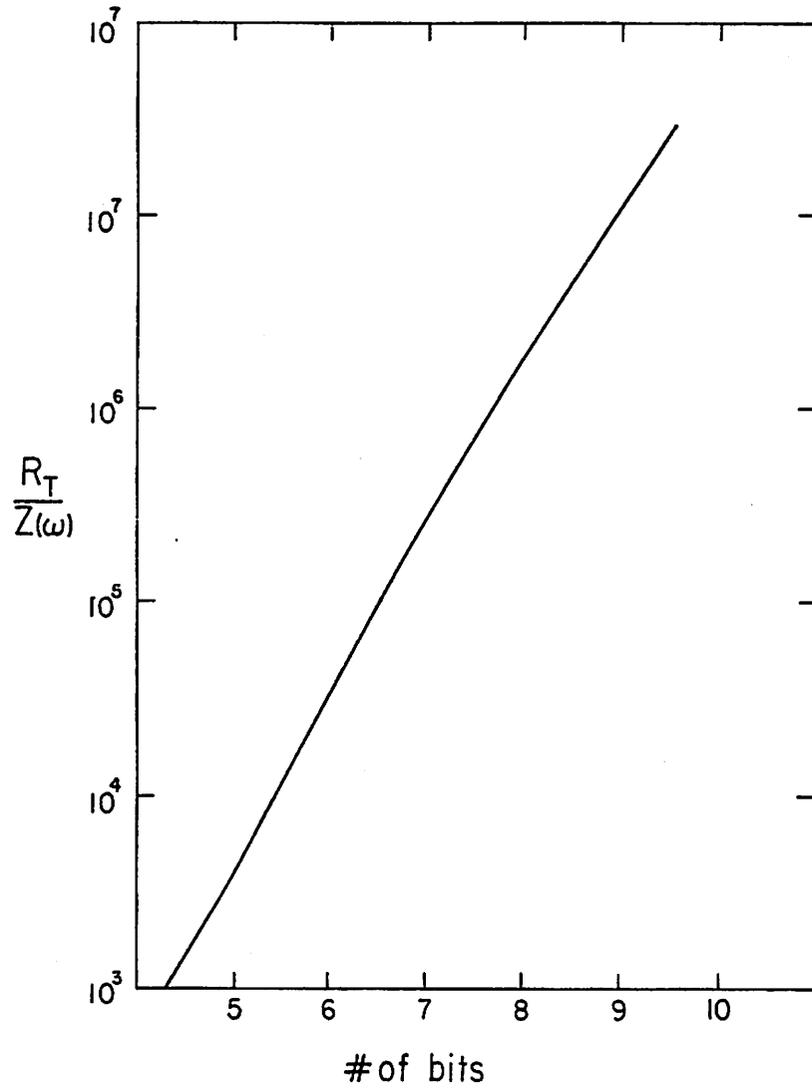


Fig. 3.15 Minimum allowable $R_T/2(\omega)$ as a function of bit resolution for 1/4 LSB error due to charge pumping alone.

which are not usually harmonically related to the input frequency. The error power resulting from clock time skew for a sinusoidal input of frequency ω and amplitude A , may be approximated by

$$E_p \approx \frac{A^2 \omega^2 \sigma_t^2}{2} \quad (3.22)$$

where σ_t is the variance of sampling period. For a S/N ratio of 40dB in a full scale sinusoidal signal of frequency ω , this indicates that

$$\sigma_t \leq .0016 \cdot 2\pi/\omega \quad (3.23)$$

which for a 4.2 MHz video signal becomes $\sigma_t \leq 380$ psec. In systems of even lower bandwidth this requires that transmission and loading effects between different comparators be well matched in order to prevent phase skew effects from degrading converter performance.

3.2.6 Comparator Phase Delay

In many flash circuit applications, the converter is used without a preceding sample/hold circuit. In these cases the input is directly tied to the comparators, and all conversions are performed "on the fly." A potential problem with this approach is the effect of finite comparator bandwidth upon the reconstructed signal. In particular, because of this limited bandwidth, the comparators will cause nonlinear phase effects in the transmitted signal. In situations which must use relatively high comparator gain (prior to regeneration) this has a profound effect upon the required comparator bandwidth for a stated phase requirement. In fact, for a single pole comparator with pole frequency p ,

$$\phi_{err}(s) = \tan^{-1} \left(\frac{s}{p} \right) \quad (3.24)$$

Thus for a 3° maximum phase shift at the maximum input frequency, it must be the case that $p/s \geq 19$. For a 4.2 MHz video signal this

translates to a required pole frequency of nearly 80 MHz. Further, variations in comparator bandwidth in the array may create a reconstructed signal phase which is not constant over different signal levels, or which is phase "noisy." The high bandwidths discussed are difficult to achieve in MOS circuits (especially at low power levels) as the required comparator gains are typically fairly high. As a point of reference, the stated -3dB frequency of the comparators in the only existing 8-bit monolithic flash converter (which is a bipolar circuit [42]) is 40 MHz.

3.3 Successive Approximation Technique

Perhaps the most common A/D conversion algorithm is the n-step sequential, or successive approximation technique. In this approach n-sequential approximations to the input voltage are made for an n-bit conversion, each typically a factor of two more accurate than the approximation made previously.⁶ In general, this algorithm can take on either of the forms shown in Fig. 3.16. The first of these, the routine employing a successively divided reference, is used in most successive approximation converters which employ n-stages of binary weighted components (or the equivalent, e.g., R-2R ladders) for an n-bit conversion. The second algorithm which employs a multiplied remainder scheme, is used in circuits which only have a few precision components, but which require the use of amplifiers [44]. As will be shown in Section 3.4, this latter technique is also useful in easily realizable pipeline converters.

As a converter employing a successive approximation algorithm uses typically n (or n+1) cycles for an n-bit conversion, very high conversion

⁶Successive Approximations by factors of less than two are possible, and indeed quite useful in converters employing nonlinear coding (such as companding) or adaptive calibration algorithms [43].

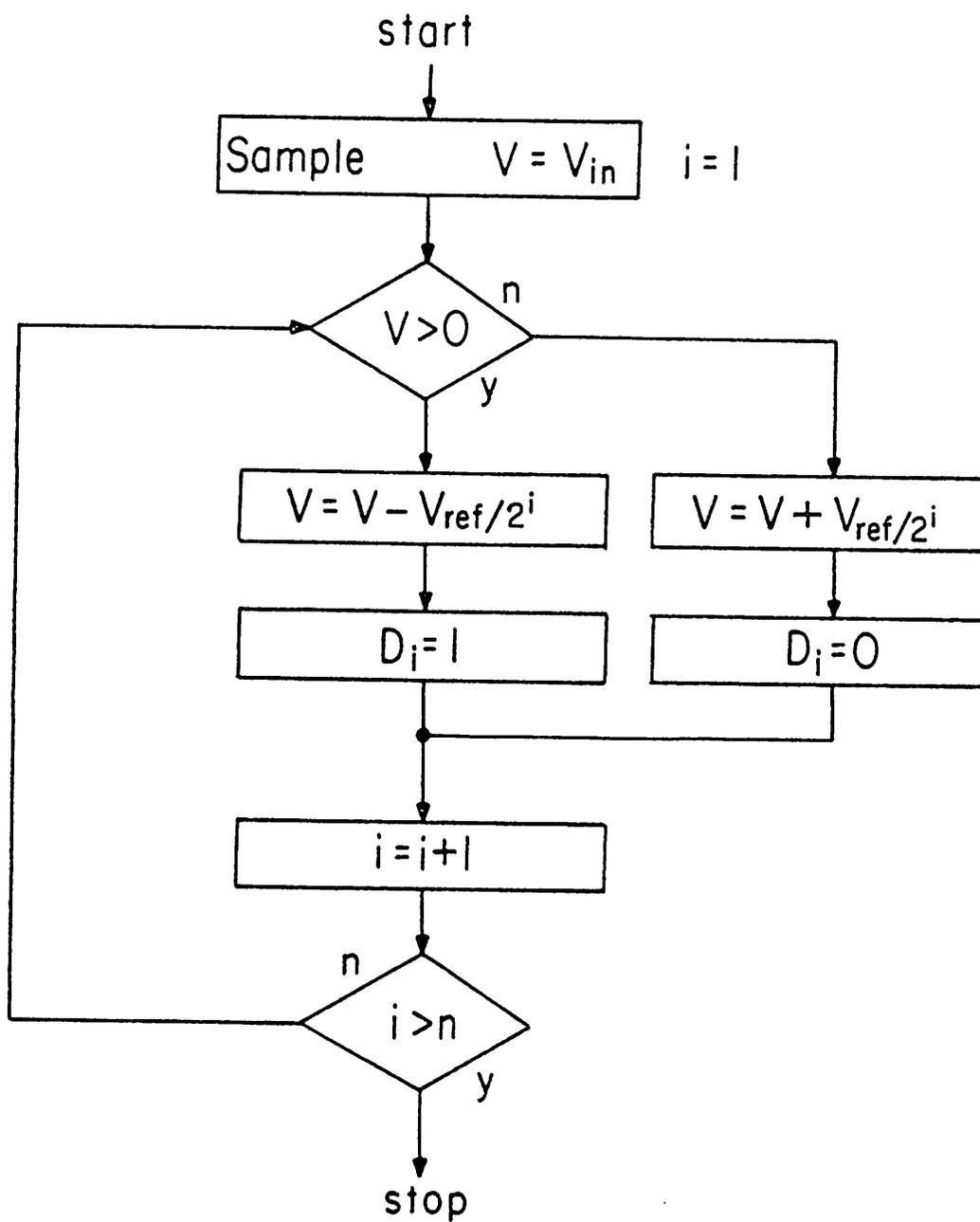


Fig. 3.16a Successive approximation algorithm using reference division.

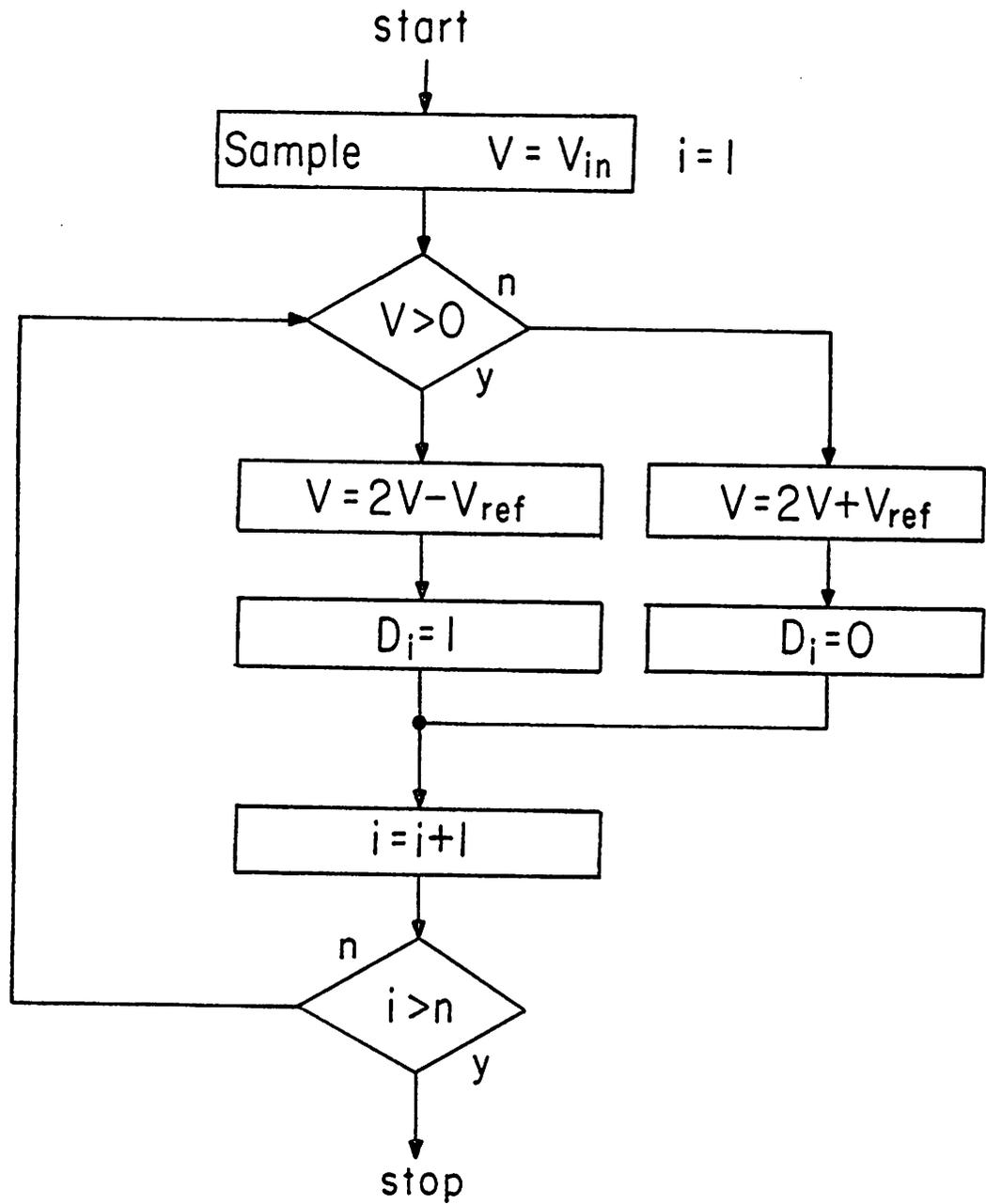


Fig. 3.16b Successive approximation algorithm using remainder multiplication.

rates may only be obtained in circuits fabricated in very fast processes. Although a high bandwidth monolithic converter has been realized at least once using this technique in a fast bipolar process [45], and appears to be possible in a few chip implementations built around a recently discussed high speed D/A [46], MOS implementations using this algorithms will be difficult or impossible to realize at conversion rates above a few MHz. This technique in one form appears to be quite useful in array configurations, however, as discussed in Chapter 6,. This particular method, which employs precision binary-weighted capacitors, is discussed in the next section.

3.3.1 Weighted Capacitor Methods

A relatively new A/D converter technique that has gained considerable acceptance, is the weighted-capacitor charge-redistribution method. This technique, first introduced by McCreary and Gray [47] typically uses binary-weighted capacitors to carry out a divided reference algorithm similar to that given in Fig. 3.16a. An implementation of a circuit of this type which directly carries out this algorithm is shown in Fig. 3.17. Converter operation proceeds in the sequence shown in this figure. More specifically, converter sampling is performed by tying all of the capacitor bottom plates to the input voltage and shorting the top plate to a fixed voltage such as ground, or the comparator offset point. At the end of the sampling period the bottom plates are reconnected to analog ground while the top plates are allowed to float. This results in a voltage present upon the top plate which is roughly proportional to $-V_{in}$. The relationship would be exact except for parasitic capacitances present on the top plate, which are usually somewhat voltage variable.

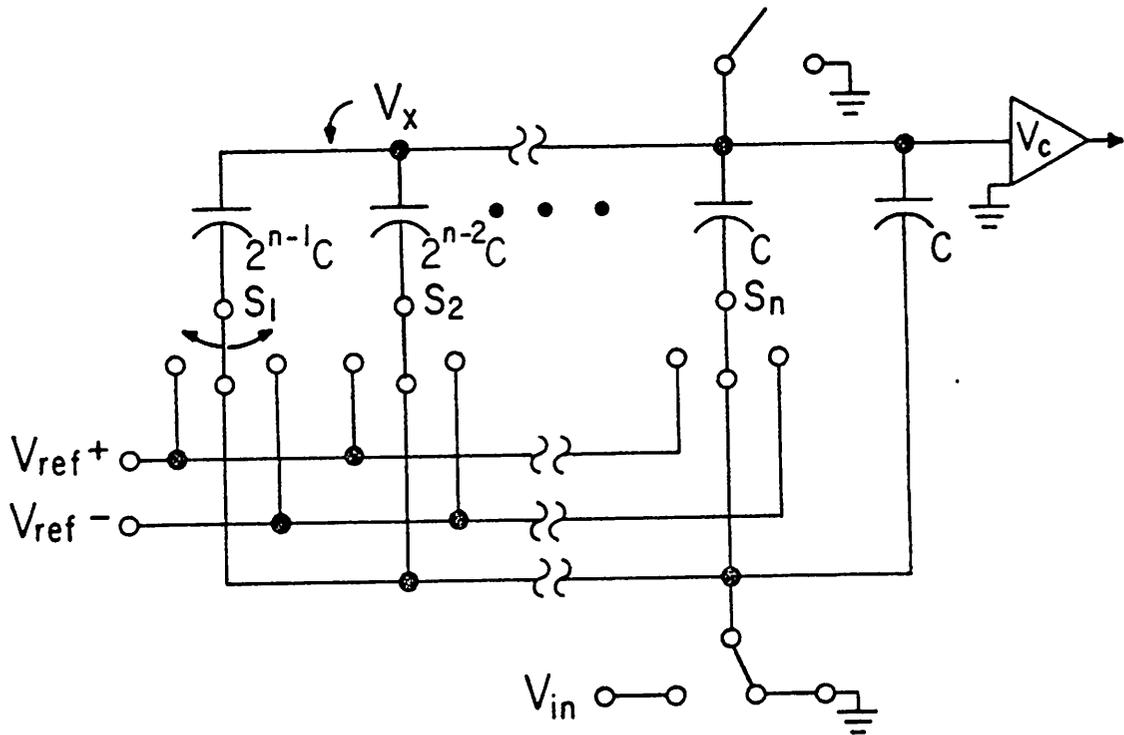


Fig. 3.17c Successive approximation cycling:

1. Determine sign of V_x
 If $V_x < 0$; Switch S_1 to V_{ref+} , $D_1 = 0$
 If $V_x > 0$; Switch S_1 to V_{ref-} , $D_1 = 1$

$$V_x = -V_{in} \pm V_{ref}/2$$

2. Determine sign of V_x

⋮

Converter operation proceeds by sequentially flipping each capacitor bottom plate to either V_{ref}^+ or V_{ref}^- (starting with the largest capacitor) depending upon the sign of V_x . This sequence continues until n capacitors have been switched, at the rate of one capacitor per time slot, for an n -bit plus sign conversion.

As this conversion method requires references to each polarity, and as it is designed for operation with bipolar input signals, it is not necessarily useful in all applications. A modification of this basic scheme which only uses one reference polarity and which operates on unipolar input signals, is shown in Fig. 3.18. In this circuit converter sampling is done as before, by tying the top plates of the capacitors to a fixed voltage such as ground, and the bottom plates to V_{in} . After the sampling period, the bottom plates are all switched to ground and subsequently switched to V_{ref} at the rate of one capacitor per clock period. After a capacitor has been flipped to V_{ref} , a sign determination is made upon the voltage present on the capacitor top plates. If the voltage is now above ground (or in many implementations, above the comparator offset point), the bottom plate is returned to ground during the next cycle. If, however, the top plate voltage is still below ground (or below the comparator offset point) the bottom plate is left at V_{ref} during the remainder of the conversion cycle. Note that this procedure differs from the one presented previously in that sign testing here only begins after the first capacitor has been flipped, while previously it was performed before. Thus this latter procedure provides one-bit less resolution than the circuit of Fig. 3.17, for the same number of capacitors. This drawback is acceptable in many applications, however, as the bit cell is usually simpler and only one reference polarity is required.

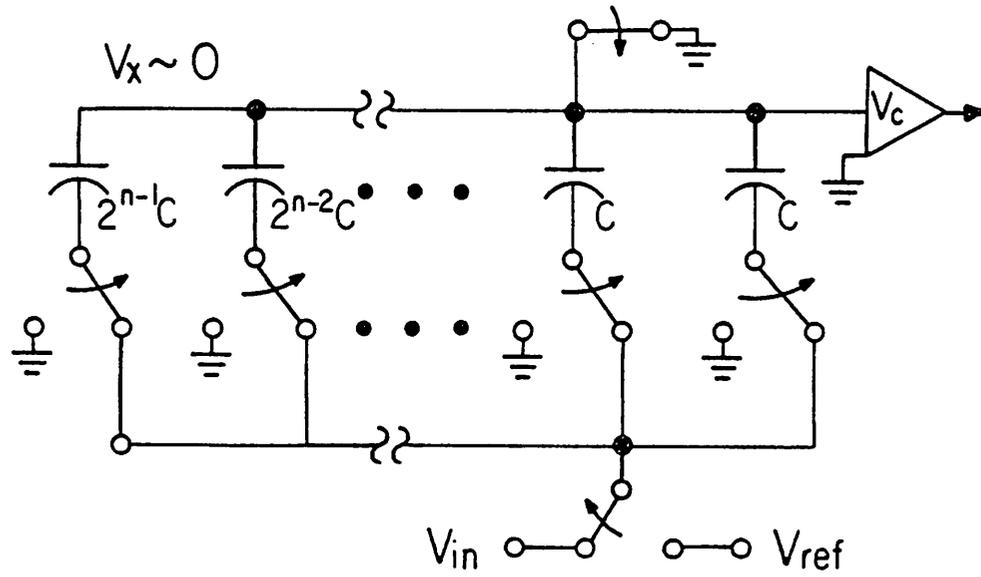


Fig. 3.18a Unipolar input converter: Sample mode.

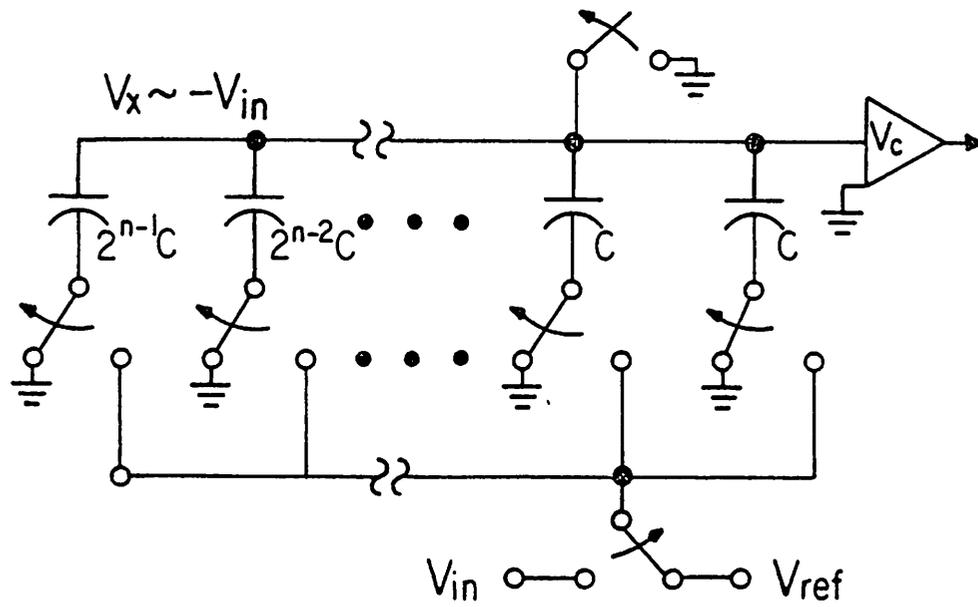


Fig. 3.18b Hold mode.

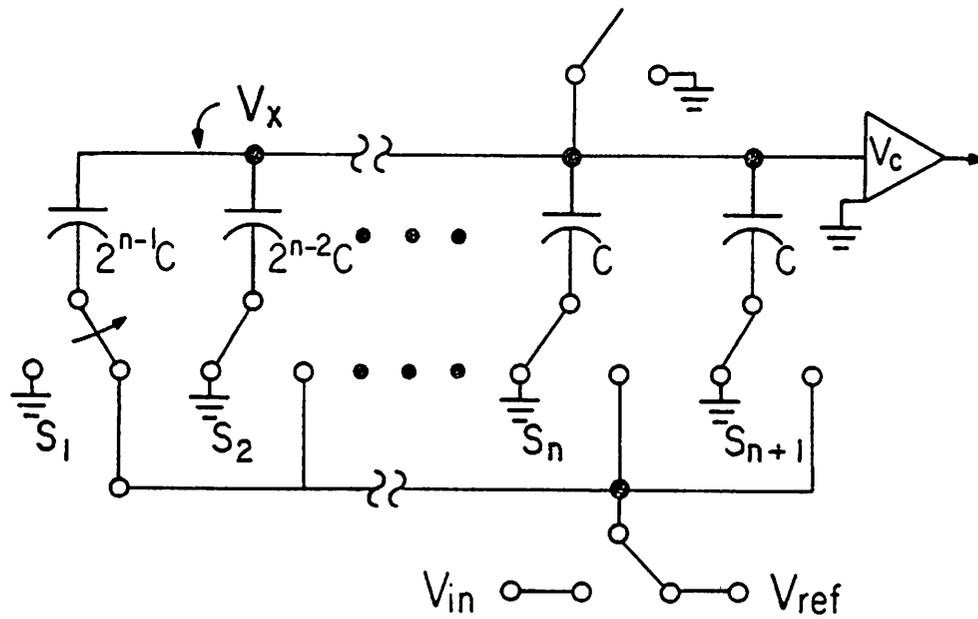


Fig. 3.18c Successive approximation cycling:

1. Switch S_1 to V_{ref}
 Determine sign of V_x
 If $V_x < 0$; Leave S_1 on, $D_1 = 1$
 If $V_x > 0$; Switch back, $D_1 = 0$

$$V_x = -V_{in} + D_1 V_{ref}/2$$

2. Switch S_2 to V_{ref}

⋮

This circuit may be modified yet again to allow bipolar inputs to be accepted, but with still just a single polarity reference. This method is illustrated in Fig. 3.19. In this approach the largest capacitor is not connected to the input during the sample mode, but instead, has its bottom plate tied to the reference voltage. At the end of the hold interval the bottom plate is connected to ground if the voltage present on the top plate is greater than zero, otherwise it is left at V_{ref} . The remainder of the conversion sequence is then the same as in the unipolar case. Note that the end result of this procedure is a compression of the input range by nominally a factor of two, and a shift in the transfer characteristic by nominally $V_{ref}/2$.

3.3.2 Capacitor Matching Requirement

The transition voltage between digital code D and $D-1$ may be given by

$$V(D) = \frac{\sum_{i=1}^n D_i V_{ref}}{\sum_{i=1}^{n+1} C_i} \quad (3.25)$$

for the unipolar converter of Fig. 3.18, and

$$V(D) = \frac{\sum_{i=2}^n D_i C_i V_{ref}}{C_1} - \frac{C_1 D_1 V_{ref}}{\sum_{i=2}^n C_i} \quad (3.26)$$

for the bipolar converter of Fig. 3.19 where D_i indicates the i -th bit of code representing the number D . Using these results, the matching requirements for various distributions of capacitor error may be determined via simulation, as was done in the flash case. It has been shown

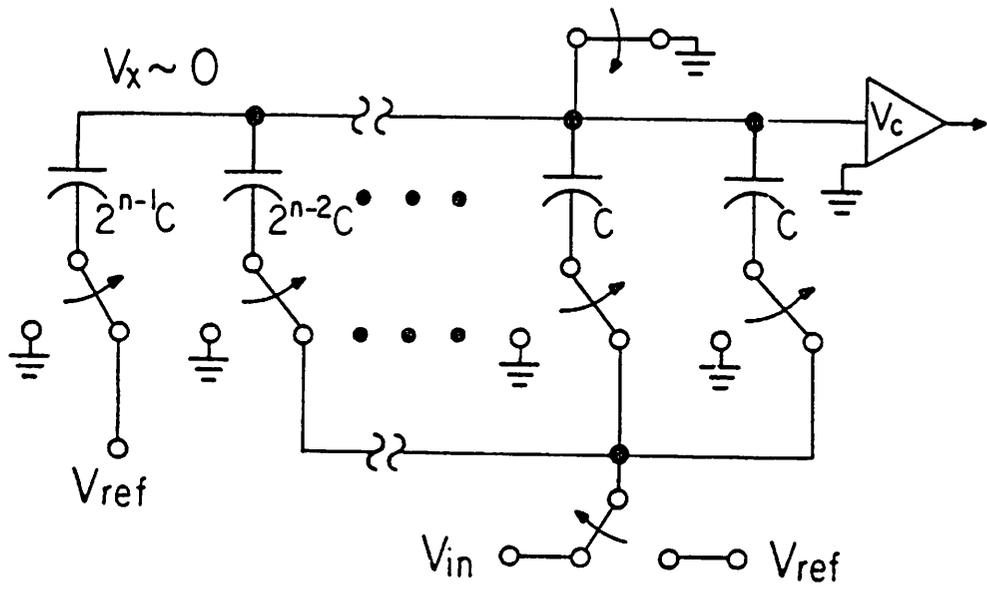


Fig. 3.19a Bipolar input converter: Sample mode.

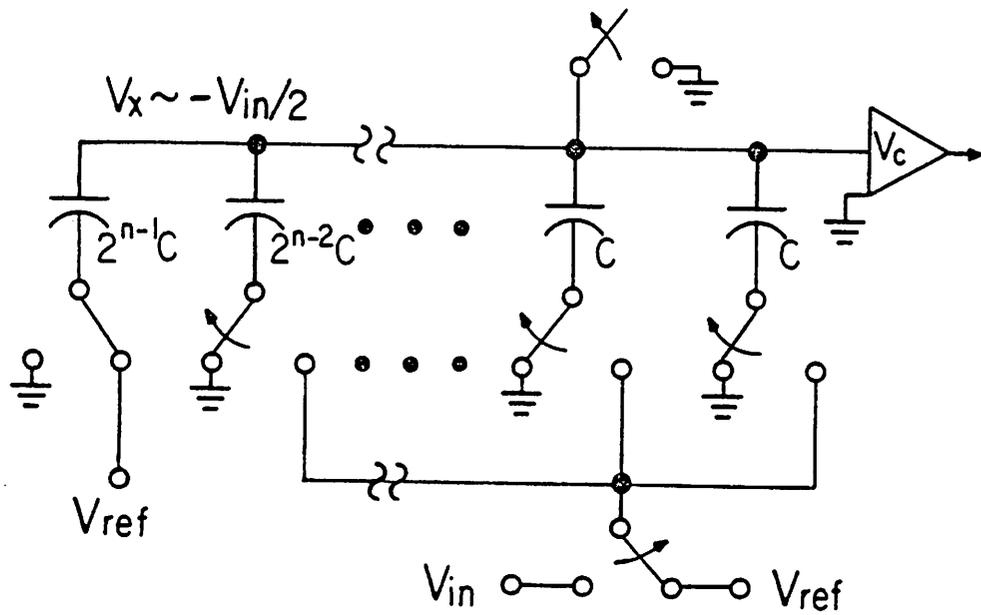


Fig. 3.19b Hold mode.

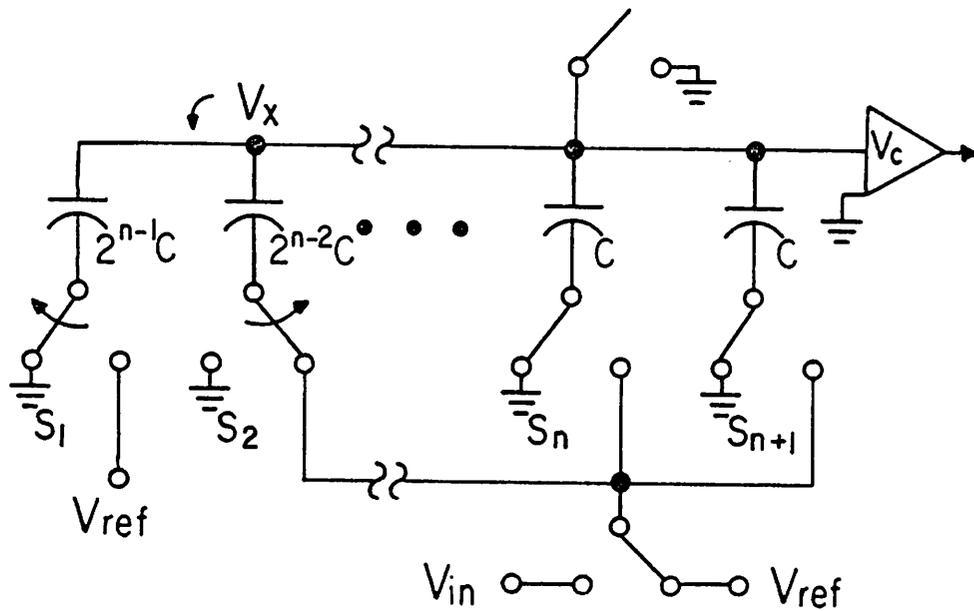


Fig. 3.19c Successive approximation cycling:

1. Determine sign of V_x
 If $V_x < 0$; Leave S_1 on, $D_1 = 0$
 If $V_x > 0$; Switch back, $D_1 = 1$

$$V_x = -V_{in}/2 - D_1 V_{ref}/2$$

2. Switch S_2 to V_{ref}
 Determine sign of V_x
 If $V_x < 0$; Leave S_2 on, $D_2 = 1$
 If $V_x > 0$; Switch back, $D_2 = 0$

$$V_x = -V_{in}/2 - D_1 V_{ref}/2 + D_2 V_{ref}/4$$

3. Switch S_3 to V_{ref}

⋮
⋮
⋮

experimentally that a least one type of precision capacitor (metal over $n+$) exhibits a nearly normal distribution of error [48]. If this is assumed as true, and if it is further assumed that all of the binary weightings are obtained by simply grouping the required number of unit cells together, the typical linearity and SNR performance is easily calculated (via program in Appendix A). The average and standard deviation of worst case error about the nominal and best fit line is shown in Fig. 3.20 for a 7-bit unipolar converter, and in Fig. 3.21 for a single reference bipolar type, for a number of different spreads in unit capacitor size. As is apparent from Fig. 3.20, the unipolar converter's fit to the nominal characteristic is nearly as good as to the best fit line. This contrasts to the bipolar case in Fig. 3.21 and the flash converter considered previously (Fig. 3.10). In these last two converter types, ratio errors will tend to modify best fit gain or offset, as well as result in nonlinearity. Of further interest in these figures, is the relative sensitivity of $\overline{\epsilon_{\max}}$ to precision component ratio errors. For an average maximum error of 1 LSB, σ_C/\overline{C} is approximately .06 for the unipolar converter, but is only .04 in the bipolar case. This compares to a required σ_R/\overline{R} of approximately .06 for the flash converter. Thus, the bipolar input converter exhibits a slightly higher component matching requirement than either of the other two types.

The signal to noise ratio performance for the weighted-capacitor circuits is indicated in Fig. 3.22 for the same distributions of capacitor errors. As is apparent, the unipolar converter is again the slightly better performer of the two, but is not quite as good as the flash converter, whose performance was indicated in Fig. 3.11.

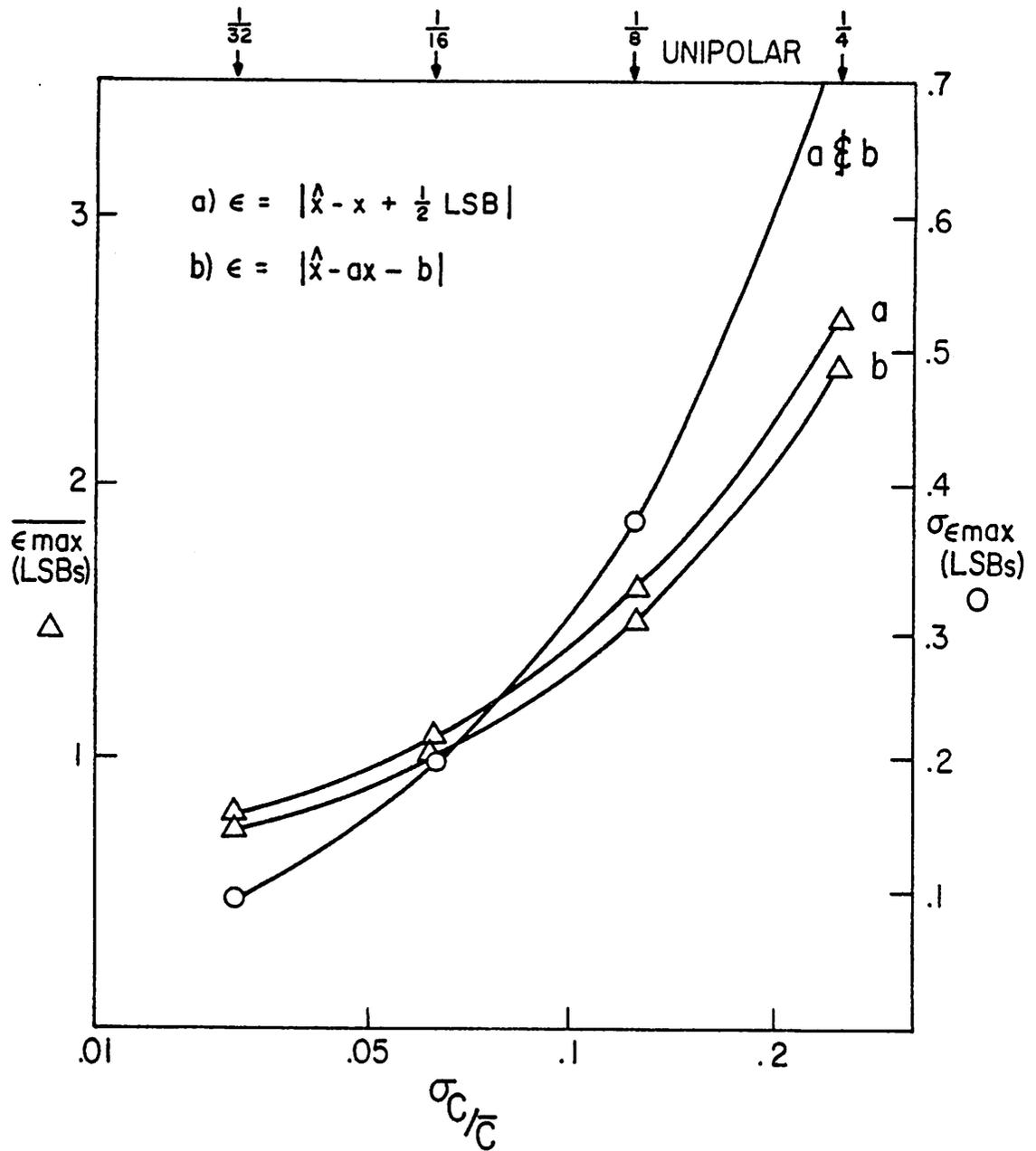


Fig. 3.20 Simulated worst case linearity error of 100 unipolar input weighted-capacitor A/D converters as a function of unit capacitor mismatch.

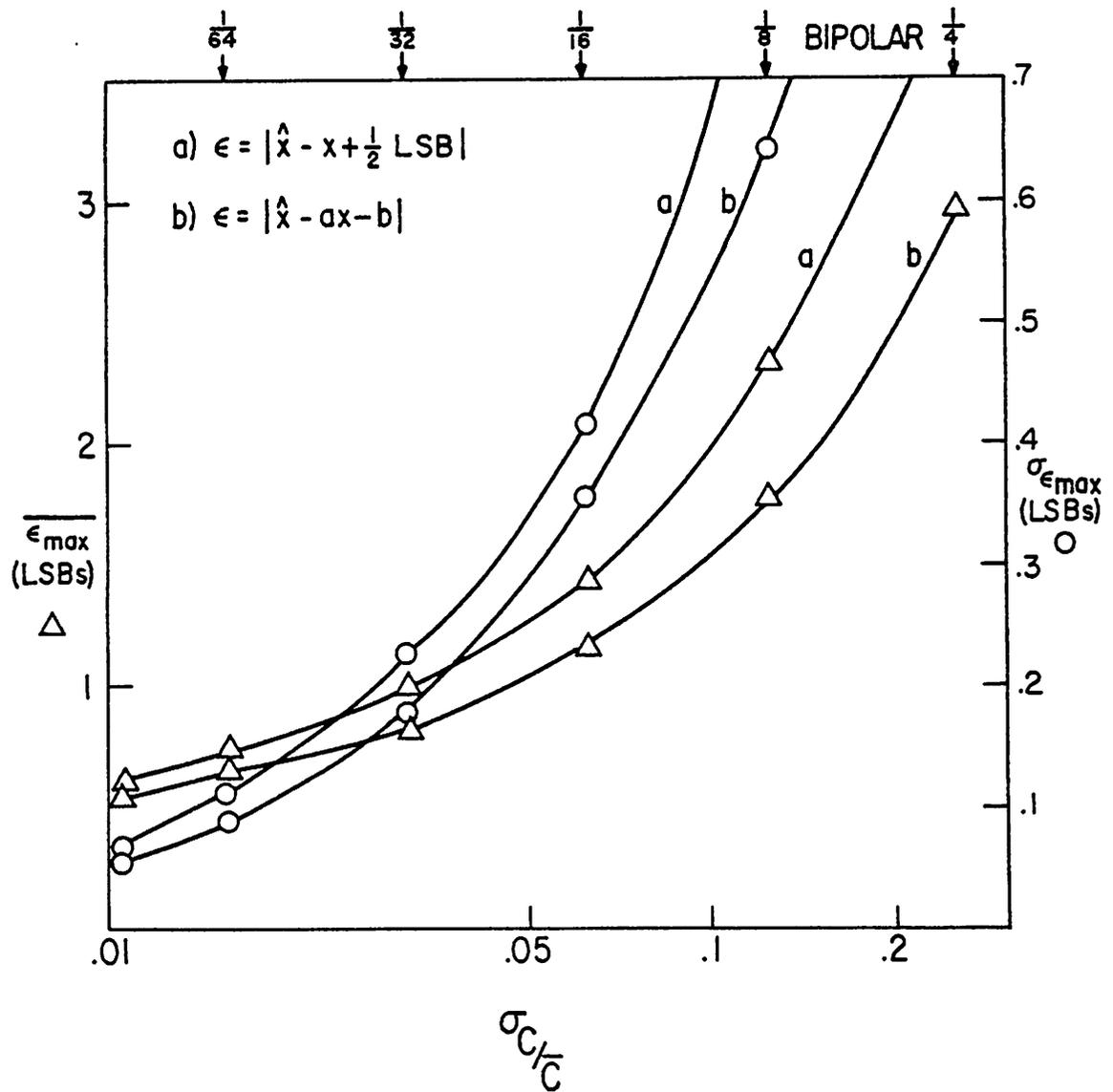


Fig. 3.21 Simulated worst case linearity error of 100 bipolar input weighted-capacitor A/D converters as a function of unit capacitor mismatch.

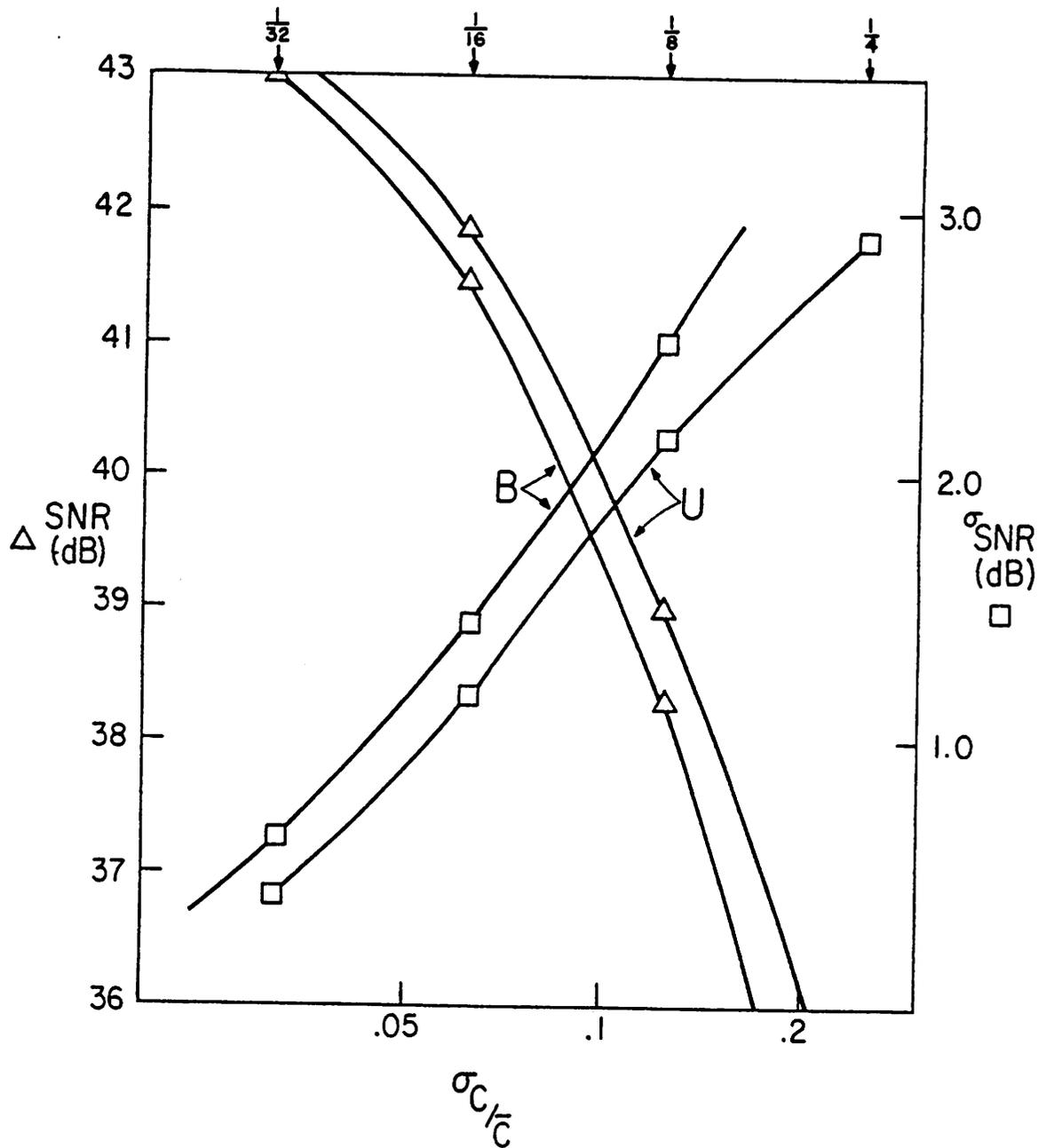


Fig. 3.22 Simulated SNR performance of 100 unipolar (a) and 100 bipolar (b) input weighted-capacitor A/D converters as a function of unit capacitor mismatch.

3.3.3 Speed Limitations

The ultimate speed limitations of real binary-weighted capacitor circuits, tends to be dominated by simple R-C time constants associated with the largest (most significant) capacitors. Depending upon the implementation, the worst case charging problem is usually during the converters sampling period. It is during this time slot that most or all of the capacitors are being tied across V_{in} and ground, or V_{in} and the comparator offset point. An equivalent circuit for this situation is shown in Fig. 3.23a. In this figure it is assumed that the switches to each capacitor are scaled with the capacitor size. This usually isn't practical or desirable beyond the most significant few bits, but gives a conservative estimate of the required charging time for a stated precision level. Resistor R_C in this figure represents the top plate switch resistance, or the equivalent input impedance of the comparator during it's offset cancellation cycle, while R_S is the switch impedance of the input bus selector. This network may be modeled as a simple R-C circuit as shown in Fig. 3.23b, where the equivalent time constant becomes,

$$\tau \approx (R_C + R_S + R/2) \sum_{i=1}^{n+1} C_i \quad (3.27)$$

$$\approx R_{eq} C_T \quad (3.28)$$

which for a settling accuracy of 1/2 LSB indicates that a minimum acquisition time is needed of the amount

$$\tau_{aq} \geq (n+1) R_{eq} C_T \ln 2 \quad (3.29)$$

If instead of a constant DC level, the input signal is a sinusoid of angular frequency ω , it is easy to show that this network produces an amplitude roll-off of the amount

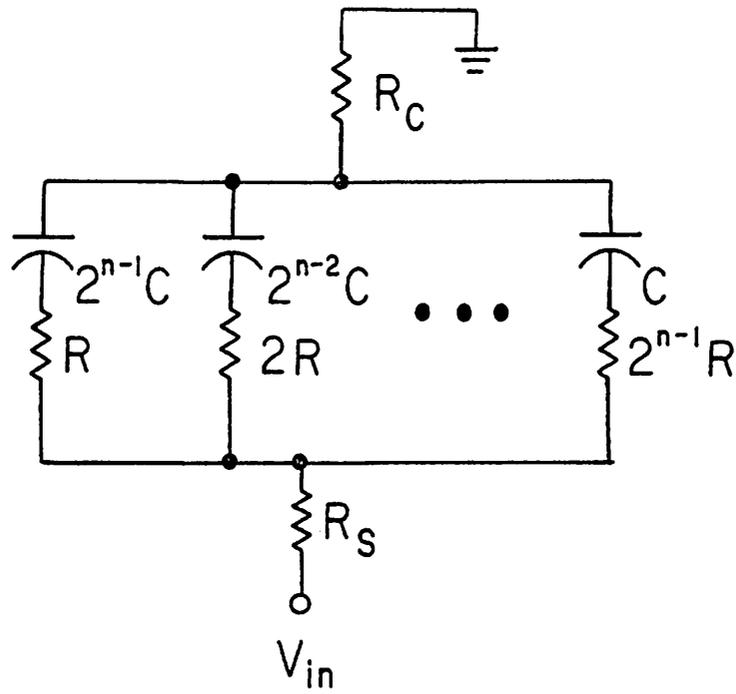


Fig. 3.23a Weighted-capacitor R-C charging model.

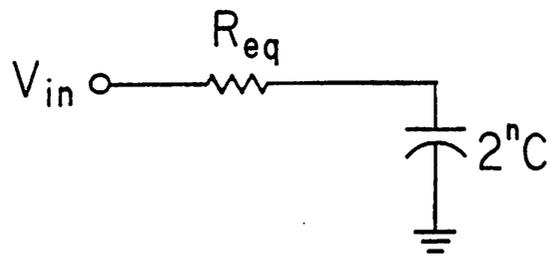


Fig. 3.23b Lumped R-C equivalent.

$$\frac{V'_{in}}{V_{in}} \approx \frac{1}{1 + \omega^2 \tau_{aq}^2} \quad (3.30)$$

where V'_{in} is the sampled value across the capacitors. This network also produces a phase shift in the sampled quantity of the amount

$$\phi_{err} \approx \tan^{-1}(\omega \tau_{aq}) \quad (3.31)$$

As indicated previously, nonlinear phase shift may be deleterious to system operation. Both the phase shift and amplitude rolloff effects may be minimized by using the bipolar double-reference circuit of Fig. 3.17, which uses the smallest amount of capacitance for a given resolution level. This arrangement requires more complicated bit cells, however, which are typically larger and slower than those of either of the other two circuits. Because of this, use of this circuit will not necessarily result in a higher conversion rate. Alternately, the bipolar input converter of Fig. 3.19 may be used. This circuit has the advantage of simple bit cells and is generally faster than the unipolar converter, as the MSB capacitor only goes through one switch to a voltage source, rather than through a pair of switches in series. Although this circuit is not as easily analyzed as its unipolar relation, it is possible to show that the use of this technique has reduced the required sampling time by 20%-30% in a real circuit [49].

Another approach for reducing the charging time constants, which is quite effective in high resolution converters, is to use a split-capacitor array as shown in Fig. 3.24. Although this circuit is functionally equivalent to a binary-weighted array, it requires a substantially smaller total capacitance. This circuit is different from the previously discussed weighted-capacitor circuits, however, in that converter linearity is dependent upon the parasitic capacitance on the capacitor top plates

(in this case in the low weight array). In fact the minimum unit capacitor size in this circuit will usually be determined by the diffusion capacitance from the top plate grounding switch. This capacitance will tend to produce a gain error in the weight of the lower precision array if it is relatively constant, and a nonlinear "bowing" if it is voltage variable. The constant gain error may be compensated for by modifying the weight of coupling capacitor C_x , while voltage variable effects may be minimized by employing a complementary switch which may be designed to have a small voltage coefficient of capacitance in a given bias region. Similarly, the voltage excursions on V_y may be minimized by intentionally employing a dummy MOS capacitor in parallel with C_p and scaling C_x to maintain a constant lower array gain. This will tend to reduce the non-linearity error created by a given voltage variable switch capacitance.

One final means of improving operation during the sample mode is to pre-sample the input with a separate sample and hold circuit which has a small aperture time and good high frequency properties. This sample and hold circuit may subsequently be used to charge the capacitor array over as long of a time as necessary. Although this results in an apparently slower conversion rate, the actual conversion rate may be faster if shorter time slots may be used for the remainder of the successive approximation cycle. This will occur in synchronous converters if the sampling period was forcing the use of longer time slots than were otherwise necessary. In addition, this procedure will reduce the roll-off and phase shift effects caused by directly coupling the capacitor array to the input. In particularly split-array converters, this sample and hold circuit is usually easier to design than circuits designed for use with flash converters, as the required capacitive loads are usually much less.

3.4 Pipeline Converters

A characteristic of the converter types previously presented, is that very little of the total circuitry is "active" at any given time. For example, only two of the $2^n - 1$ comparators in a flash circuit will actually determine the output code. Similarly, in a successive approximation converter, only one or two of the n different bit cells are involved in the conversion process during any time slot. It would be desirable if a technique could be developed which would use more of the converter circuitry a greater percentage of the time. This in theory at least, should allow a higher converter bandwidth to be realized.

One approach is to modify the previously presented successive approximation algorithms into the form presented in Fig. 3.25. In this method distinct circuitry exists for each bit of the conversion process, but no bit cell is idle during any time slot. This is accomplished by having each bit cell operating on a different sample during each clock period and synchronously passing the conversion remainder onto subsequent stages. For an n -bit converter this results in an effective throughput of one conversion per clock cycle, but with a time delay of typically n periods. A difficulty with this approach, however, particularly in high resolution converters, is the problem of achieving good isolation between bit cells without introducing distortion. In general this requires the use of high gain linear amplifiers between stages, although buffering with simpler techniques (e.g., source followers) appears possible in low resolution applications.

Three possible implementations of a pipeline are presented in the next two sections. The first two of these involve the divided reference algorithm of Fig. 3.16a, while the last approach uses the multiplied

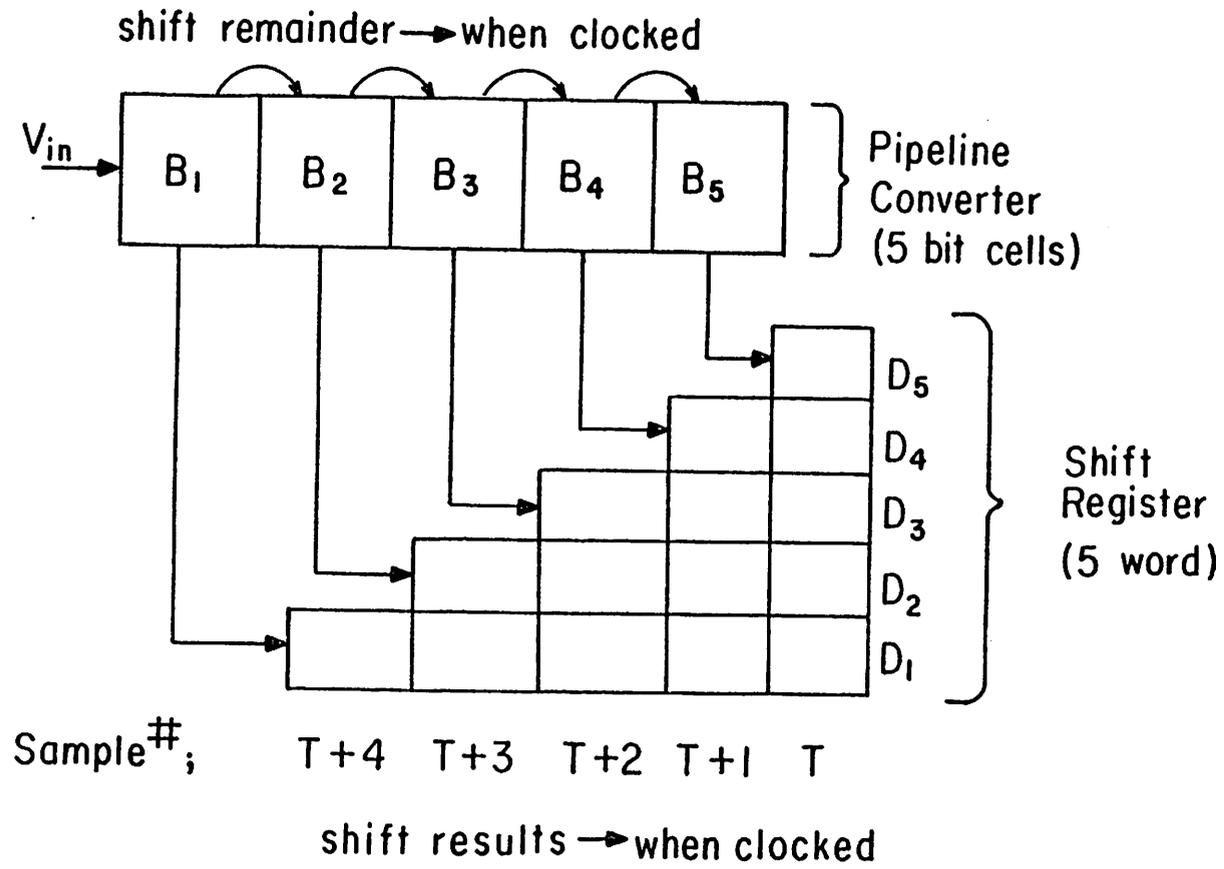


Fig. 3.25 Pipeline converter method.

remainder algorithm of Fig. 3.17b. As will be demonstrated, the divided reference methods are useful in low resolution converters (although they require great care in the design), while the multiplied remainder technique is potentially useful in even high precision applications.

3.4.1 Divided Reference Pipeline

A straightforward implementation of a pipeline which uses the divided reference successive approximation algorithm, is shown in Fig. 3.26. In this circuit a succession of precision capacitor pairs are used to generate the necessary reference voltage fractions and the required additions and subtractions. Isolation between stages is provided by voltage followers which allow the capacitors of subsequent stages to be charged without a voltage loss. The voltage followers may be low distortion, low offset source followers (or the equivalent, see Chapter 6) or fast settling linear amplifiers connected in a unity gain configuration.

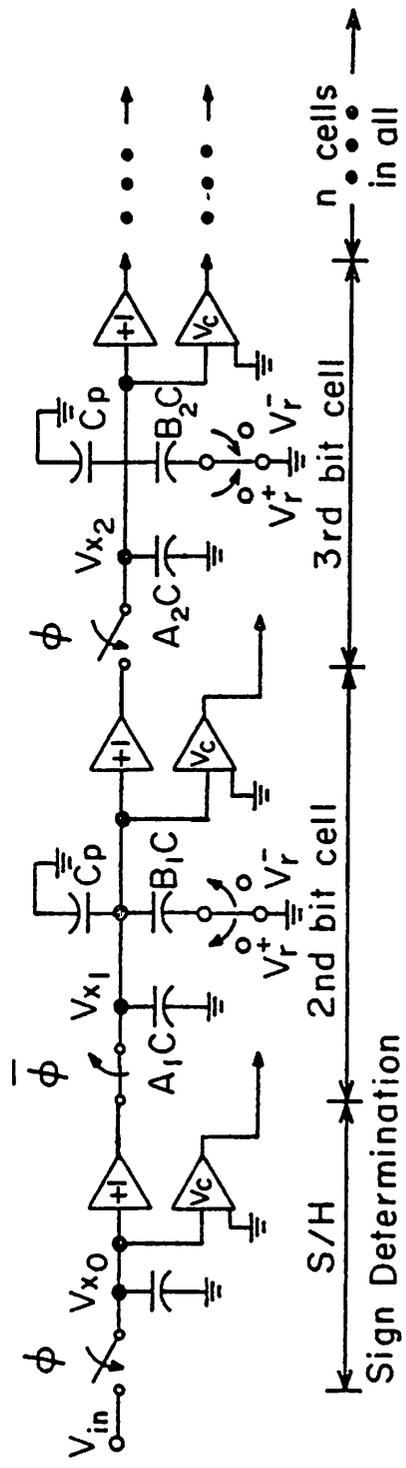
A difficulty with this circuit is that it is sensitive to parasitic capacitance on each integrating node (represented by the lumped capacitance C_p). The effect of this capacitance, if constant, is a gain roll-off of the summed reference fraction. Specifically, after the first summation, the voltage present on V_{x1} is given by

$$V_{x1} = V_{in} \pm \frac{B_1 C V_{ref}}{(A_1 + B_1)C + C_p} \quad (3.32)$$

which for proper converter operation requires that

$$\frac{B_1 C}{(A_1 + B_1)C + C_p} = 1/2 \quad (3.33)$$

or



ϕ	V_{x0}	V_{x1}	V_{x2}
1	V_{in}	X	X
0	"	$V_{in1} \pm B_1 C V_r / (A_1 + B_1) C + C_p$	X
1	V_{in2}	"	$V_{in1} \pm B_1 C V_r / ((A_1 + B_1) C + C_p) \pm B_2 C V_r / ((A_2 + B_2) C + C_p)$
0	"	$V_{in2} \pm B_1 C V_r / (A_1 + B_1) C + C_p$	"
1	V_{in3}	"	$V_{in2} \pm B_1 C V_r / ((A_1 + B_1) C + C_p) \pm B_2 C V_r / ((A_2 + B_2) C + C_p)$

Fig. 3.26 Divided reference pipeline employing ratioed capacitors and voltage followers. (Top) Pipeline schematic. (Bottom) Time-state definitions where time progresses downward.

$$A_1 = B_1 - C_p/C \quad (3.34)$$

If it is desired that $A_1 + B_1 = 2$, then

$$A_1 = 1 - C_p/2C \quad (3.35)$$

and

$$B_1 = 1 + C_p/2C \quad (3.36)$$

In fact, for any stage i , which is to generate a reference fraction of the amount $V_{ref}/2^i$,

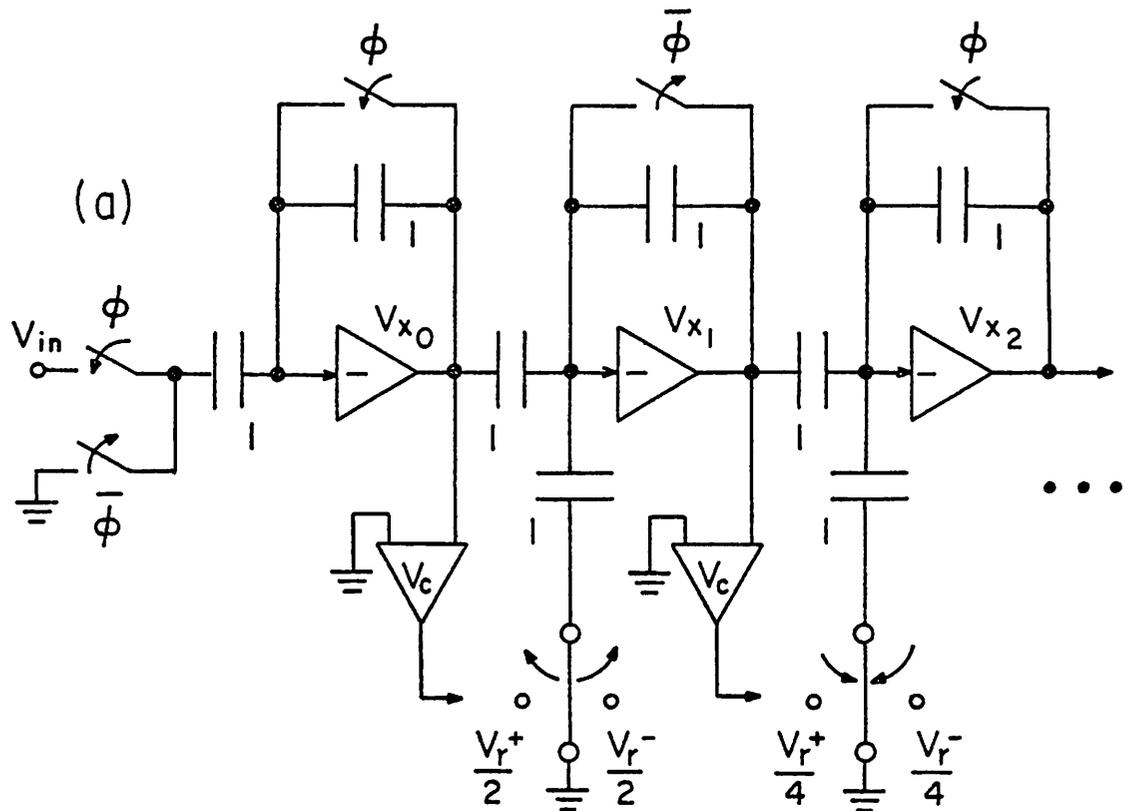
$$A_i = \frac{2^{i+1}C - 2C - C_p}{2^i C} \quad (3.37)$$

and

$$B_i = \frac{2C + C_p}{2^i C} \quad (3.38)$$

A problem arises in this circuit, however, if C_p has a voltage dependent component (as it invariably does). This voltage dependence is usually caused by the junction depletion capacitance of the switch on the capacitor top plate. As mentioned previously, this dependence may be reduced by employing a complementary top plate switch which may be designed to exhibit a small voltage dependence of capacitance over a modest voltage range. In precision applications, however, which require converter resolutions to greater than 5 or 6 bits, this procedure may not be sufficient to ensure adequate linearity. In these cases the circuit may be modified, as shown in Fig. 3.27.

In this circuit each stage of the converter is based upon a capacitively coupled multiplier/summer circuit which uses a fast settling inverting amplifier. This circuit may be used with constant sized



(b)

	ϕ	V_{x0}	V_{x1}	V_{x2}
	1	0	X	0
	0	V_{in1}	0	X
time ↓	1	0	$V_{in1} \pm V_r/2$	0
	0	V_{in2}	0	$V_{in1} \pm V_r/2 \pm V_r/4$
	1	0	$V_{in2} \pm V_r/2$	0

Fig. 3.27 Divided reference pipeline using linear amplifiers.
 (a) Pipeline schematic. (b) Time-state definitions.

capacitors and binary-scaled reference sources (generated by an R-2R ladder for instance), or used with only a positive and negative reference and scaled capacitors. As the amplifiers effectively eliminate the effect of voltage dependent parasitic capacitance on linearity, good linearity may be obtained with even large voltage swings. In addition, the amplifiers maintain a unity gain addition or subtraction of the reference voltage, thus eliminating the need for scaling of voltages by peculiar fractions. Operation of this circuit proceeds in the manner shown in Fig. 3.27b. As is apparent from the figure, each amplifier is reset during alternate time slots. This maintains the amplifiers within their normal operating regions (even in the presence of leaky junctions) and allows each stage to operate as an isolated sample and hold. As in the previous circuit, the time required for a complete n -bit conversion is only $n/2$ clock cycles, due to switches on adjacent stages being run from opposite clock phases. Addition or subtraction of the fractional reference voltage is accomplished most quickly if both positive and negative references are available. In this case the capacitors to the reference fractions are simply kept at ground during the amplifier reset cycles and switched to plus or minus $V_r/2^i$, depending upon whether subtraction or addition is desired. This same function may be achieved with but one reference polarity if the reference input capacitor can be precharged to ground or $V_r/2^i$ during the amplifier reset cycle, based upon the comparator result from the previous stage. This requires more time, however, and is thus not as attractive in high speed applications.

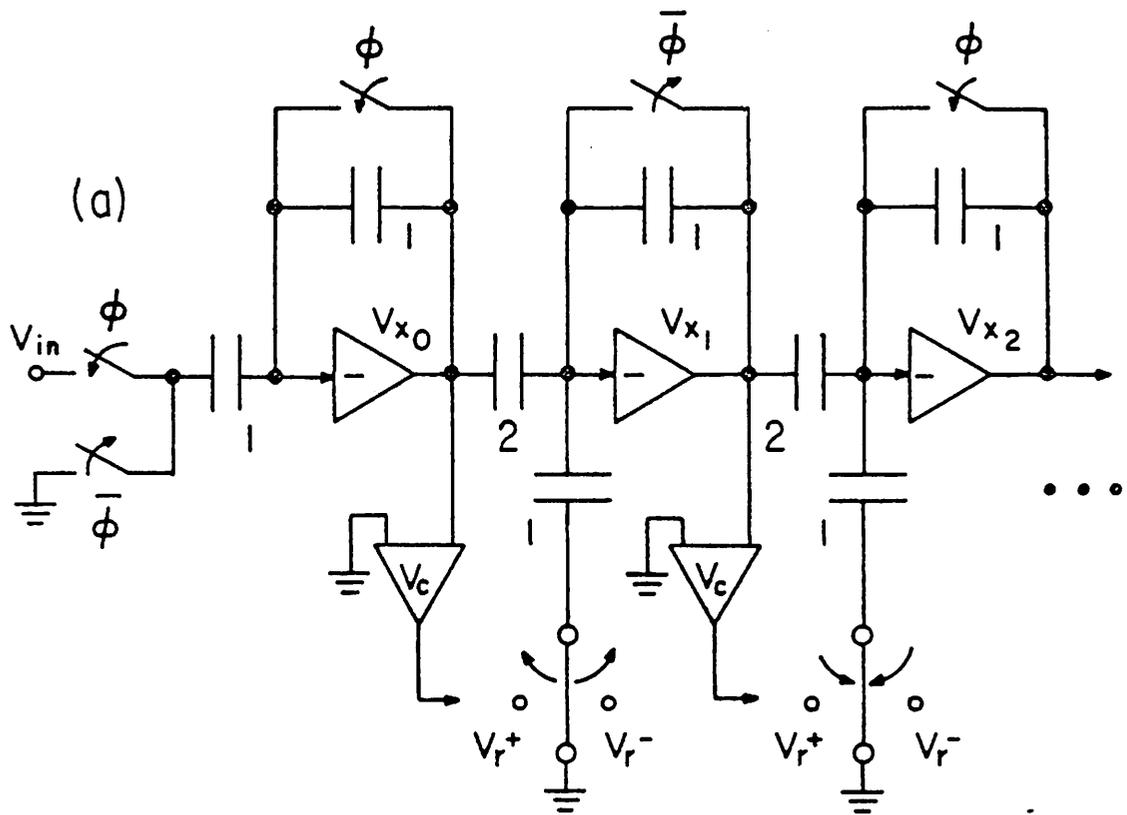
Practically speaking, the circuit of Fig. 3.27 has several problems. First, the generation of fractional reference voltages tends to make

the circuit somewhat cumbersome, whether the fractions are generated by capacitor weightings or R-2R circuits. Second, reset switch feedthrough and amplifier noise appears as an unattenuated error in the conversion remainder, which is made worse at every stage. This can easily result in poor linearity in the converter transfer characteristic. A preferred embodiment of a pipeline converter, which does not have such severe problems is discussed below.

3.4.2 Multiplied Remainder Pipeline

A slight modification of capacitor ratios in the circuit of Fig. 3.27, allows the multiplied remainder algorithm to be carried out.⁷ This circuit, which is shown in Fig. 3.28, is in many ways preferable to the divided reference pipeline converter in high resolution applications. As the circuit multiplies the remainder after each stage, no binary scaling of the reference voltage is necessary. Further, the remainder multiplication reduces the effect of switch feedthrough and amplifier noise by a factor of two at every stage. Thus the problem of feedthrough induced nonlinearity and amplifier related converter noise is principally limited to the first few stages. The required comparator resolution is also reduced after the first stage, possibly allowing the use of smaller and simpler comparators in the least significant bit positions.

⁷Note that certain embodiments of this algorithm are similar to the "stage by stage converter" described by Gordon [50] and invented by Waldhauer [51] and Jepperson [52].



(b)

	ϕ	V_{x_0}	V_{x_1}	V_{x_2}
	1	0	X	0
	0	V_{in_1}	0	X
time ↓	1	0	$2V_{in_1} \pm V_r$	0
	0	V_{in_2}	0	$2(2V_{in_1} \pm V_r) \pm V_r$
	1	0	$2V_{in_2} \pm V_r$	0

Fig. 3.28 Multiplied remainder pipeline. (a) Pipeline schematic. (b) Time-state definitions.

3.4.3 Pipeline Performance Limitations

As all of the pipeline circuits presented employ some form of linear buffer circuitry, the settling time of these circuits will most directly effect the maximum conversion rate. The design of fast settling circuits which are suitable for use in pipeline converters, and their ultimate speed limitations are discussed in Chapter 5.

The ultimate resolution of the first pipeline considered (shown in Fig. 3.26), is affected by the voltage variability of the parasitic capacitance on the top plate and drift in this capacitance from the nominal design value (used in determining the required capacitor ratios). Circuit linearity is also affected by buffer nonlinearity and offset, (particularly if the buffer is not a linear amplifier employing feedback), and switch feedthrough. The result of all of these problems taken together is that this circuit will probably not be useful in high speed converters of even modest resolution unless great care is taken in every aspect of the design.

The resolution of the improved divided reference circuit of Fig. 3.27, will typically be dominated by switch feedthrough effects, as the myriad of parasitic capacitance and buffer nonlinearity problems will have been eliminated. In general, however, the resolution of this circuit will be less than the multiplied remainder circuit of Fig. 3.28. As the noise and feedthrough effects of this last circuit of Fig. 3.28. attenuated at every stage, the result of these nonidealities upon overall circuit linearity is greatly reduced. In fact, with regard to switch feedthrough alone, it is easy to show that $1/2$ LSB INL will be obtained if

$$\sum_{i=2}^n V_{ft_i} / 2^{(i-2)} < 1 \text{ LSB} \quad (3.39)$$

where V_{ft_i} represents the feedthrough resulting from the i th stage. If the above constraint can be met or exceeded, the ultimate linearity obtainable will probably be limited by simple capacitor matching or comparator offset skews. Note that this latter problem, that of comparator offset skew, may be virtually eliminated by a variety of techniques which are discussed in Chapter 5. This multiplied remainder pipeline scheme, appears to offer considerable promise in high resolution converters, either by itself, or in a hybrid configuration as discussed in the next section.

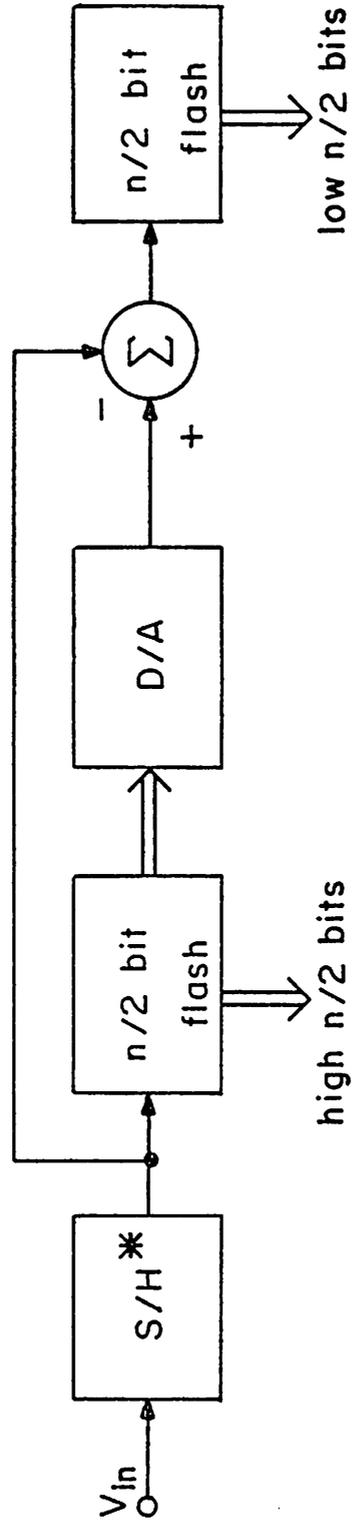
3.5 Hybrid Converters

Each of the previously discussed conversion techniques have characteristics which make their use desirable in certain systems. In general, however, every simple method for performing an A/D conversion becomes difficult to realize as either the required precision or the required speed increases. For example, a flash converter will require an inordinate die size for resolutions of greater than 6 or 7 bits while pipeline converters will increasingly suffer from switch feedthrough effects as the switch size is increased and the precision capacitor sizes are decreased in order to improve speed. In certain cases, some of these problems may be avoided by combining several converters (of the same or different types) into a single composite or hybrid converter. Several of these circuits are discussed in this section.

3.5.1 Two-Step Flash Converter

A fairly commonplace high-speed converter method is the 2-step or half flash technique. In this approach, a pair of flash circuits (of typically about $n/2$ bits resolution) are used in sequence to obtain a fast n -bit conversion, as shown in Fig. 3.29. Although this method only allows a bandwidth of approximately half that of the straight flash approach, it can be made considerably smaller in high resolution applications. This technique has been demonstrated in a partially monolithic current-input converter fabricated in a double-level metal bipolar process [53].

The design considerations in a converter of this type are in many ways similar to those of the simple flash converter with the addition of matters relating to the D/A and differencing circuit. The resistor linearity requirement upon the first flash circuit is the same as if it were a complete n -bit converter, even though its actual resolution is only approximately $n/2$ bits. The required comparator performance, however, is considerably reduced for the requirement of monotonic behavior. Problems relating to charge pumping within the precision ladders are also greatly reduced, as the number of comparators per flash circuit is so much smaller. Depending upon time delays present in the D/A and subtractor circuit, the half-flash circuit may have a greater need for an input sample and hold circuit than the straight flash approach. With a high frequency input signal the half-flash circuit of Fig. 3.29, if used without a sample and hold, can easily produce grossly distorted results. One method for alleviating this problem has been to combine the first flash circuit, the D/A converter and the subtractor into a single high speed circuit [54]. By minimizing the total delay through



*Optional

Fig. 3.29 Two-step flash converter method.

this path, acceptable performance has been obtained with even broadband input signals.

A slight modification of the basic half-flash circuit should allow improved bandwidth to be obtained. In the circuit as just presented, the first flash converter must wait until the D/A and subtractor have fully settled and the second flash converter has finished sampling before it can proceed with subsequent conversions. If an analog sampler is incorporated into the forward differencing path, and if a latch is placed after the first flash circuit, the overall converter circuitry effectively becomes pipelined (Fig. 3.30). This allows the operation of the first flash circuit to overlap with that at the DAC/subtractor and second flash circuits, resulting in an improved throughput. Further, this technique should greatly reduce the need for a separate sample and hold circuit prior to the first flash converter.

3.5.2 Flash-Pipeline Converter

As discussed in the previous sections of this chapter, the flash converter method becomes rather unwieldy as the required resolution increases above 6 or 7 bits. Similarly, for the pipeline converters discussed in Section 3.4, the effect of switch feedthrough becomes an increasingly serious problem at high resolution levels. An interesting converter approach which takes advantage of the favorable aspects of each converter type, yet which is not deleteriously affected by the problems mentioned above, is shown in Fig. 3.31. In this circuit a flash front-end is used in conjunction with a high-speed pipeline to obtain a single fast converter. Operation of this converter begins by having the input stage of the pipeline, and each of the flash comparator

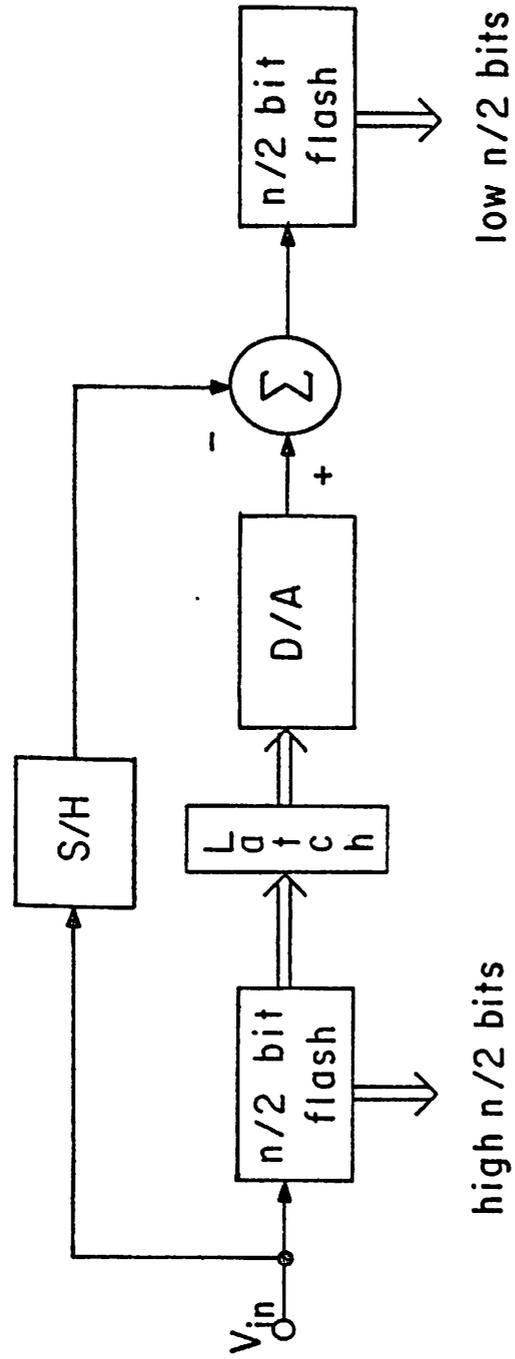


Fig. 3.30 Pipelined two-step flash converter.

circuits "track" the input voltage. After the track mode, the flash converter proceeds to determine which voltage is closest but not greater than the input voltage. After this has occurred, the input capacitor to the pipeline is switched to the tap in the middle of the next higher resistor segment. This creates an amplified difference voltage which can subsequently be digitized by the pipeline circuitry.

Note that because the resistor string is inherently monotonic, the converter as a whole will display a monotonic characteristic if the pipeline circuitry does not display a nonmonotonicity by itself. Although the charging times of the pipeline input capacitor being switched to the appropriate ladder tap will probably limit the flash part of the converter to a relatively small number of bits, this will extend the useful range of the pipeline method by an equivalent amount. Further, because flash circuits of low resolution are small, this technique will not greatly increase the die size from that required by the pipeline circuitry alone.

CHAPTER 4

PRECISION ELEMENTS

Nearly all analog-digital conversion circuits require a precision analog multiplier or divider, which is at least as accurate as the converter itself. Thus a 10-bit conversion circuit will usually require precision components which match to a minimum of 10-bits (and usually more). A common means of implementing this precision multiplier/divider circuit is with a pair (or more) of matched resistors or capacitors. Various resistor considerations including matching, voltage and temperature dependent behavior are presented in Sections 4.1-4.5, while capacitor related material is covered in Sections 4.6-4.9.

4.1 Resistor Matching Considerations

Depending upon the specific integrated circuit (IC) process, diffused, thin film or polysilicon resistors may be available. In certain cases, MOSFETs themselves may be used as a source of matched transresistance. Consider the general pair of matched resistors shown in Fig. 4.1. If the relative conductance is assumed homogeneous in the y-z plane, then the value of either resistor may be given by:

$$R = \frac{L}{W \int_0^{x_L} \sigma(x) dx} \quad (4.1)$$

$$= \frac{L}{W x_L \bar{\sigma}} \quad (4.2)$$

where $\bar{\sigma}$ is the average conductivity in the resistor. If an uncertainty δ

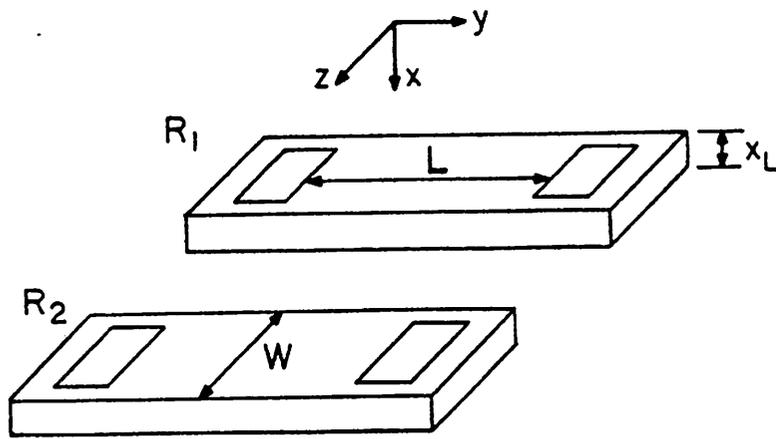


Fig. 4.1 Matched resistor pair.

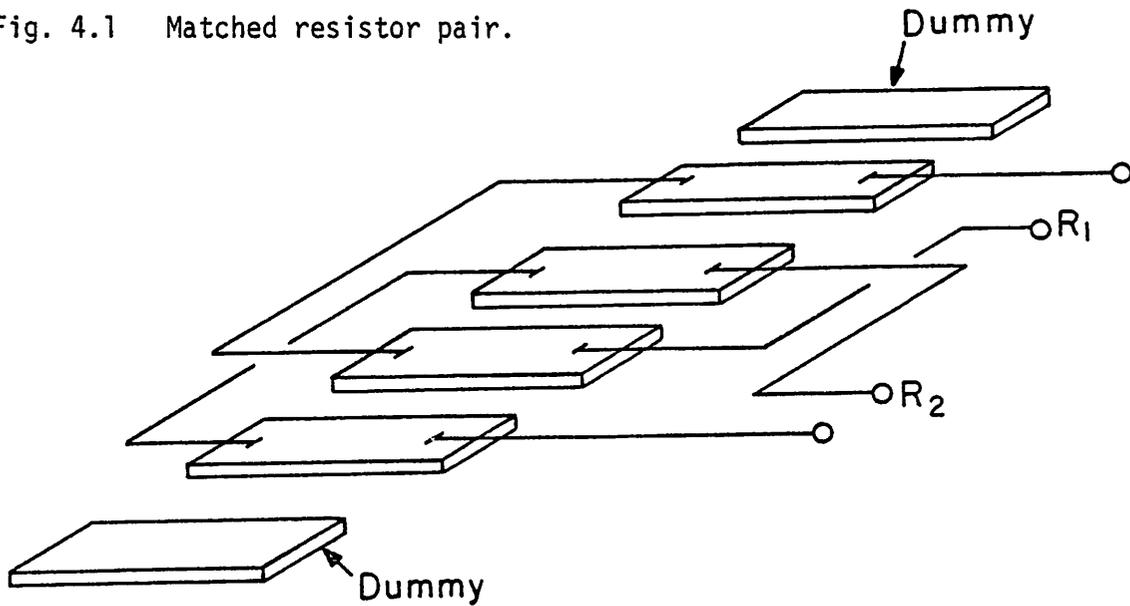


Fig. 4.2 Interdigitated resistor pair. Cross-coupled structure reduces the effect of linear gradients in resistor matching.

is assigned to each resistor parameter, a potential error will result in the final resistor value of the amount

$$\frac{\delta R}{R} \approx \left[\left(\frac{\delta W}{W} \right)^2 + \left(\frac{\delta L}{L} \right)^2 + \left(\frac{\delta x_L}{x_L} \right)^2 + \left(\frac{\delta \bar{\sigma}}{\bar{\sigma}} \right)^2 \right]^{1/2} \quad (4.3)$$

where it is usually the case that the relevant uncertainty is the spread within the precision components on a single die, and not the variation across different wafers or lots. Further, it is generally true that $\delta W/W$ will dominate $\delta L/L$, as most resistors have an aspect ratio greater than one ($L/W > 1$). Variations in W and L may be due to dopant inhomogeneities and crystal defects (particularly in diffused resistors) or variations in deposition rate and uniformity (thin-film or poly resistors) or grain size (poly resistors). In addition, $\bar{\sigma}$ is quite sensitive to stress effects [55], which may seriously affect component matching after packaging. Resistor matching of from .23% for diffused resistors to .12% for implanted resistors [56] has commonly been reported, although one report indicates that diffused resistor matching to .01% is at least possible [57]. Typically, resistor components which must match to greater than 8 or 9 bits are trimmed by any of a number of techniques to the required accuracy [58-60]. As a point of reference, .1% resistor matching in 50 μm wide resistors, indicates an edge uncertainty of less than .05 μm . The effect of random edge variations may only be reduced by using larger resistor geometries, although linear gradients may be compensated by using an interdigitized structure such as that shown in Fig. 4.2. The purpose of the "dummy" resistors at each end of the array, is to maintain a uniform environment for all of the resistors during processing.

4.2 Characteristics of Planar Diffused Resistor Structures

Of the variety of resistor types available in an IC process, a diffused structure is often the most advantageous to use, as no process modification is necessary and the resistivity may be well controlled (at least in ion-implanted types). The planar characteristics of diffused resistors with gaussian impurity profiles are examined in this section, while additional effects, resulting from sidewall outdiffusion are considered in Section 4.4.

In diffused resistors, the effective dimensions are somewhat bias dependent, due to the junction depletion region extending a voltage variable distance into the resistor. This situation is illustrated in Fig. 4.3. The resistors value may still be determined from Eq. (4.2), however, corrections must be made for x_L .

An implanted resistor which has been diffused into the background material usually exhibits a gaussian impurity profile

$$N_c(x) \approx \frac{Q_I}{\sqrt{\pi DT}} e^{-x^2/4DT} \quad (4.4)$$

where Q_I is the implant dose in ions/cm², D is the effective impurity diffusivity over the length of the drive in, and T is the drive in time. For a background doping of N_B ions/cm³ of an opposite impurity type, this results in a junction depth of

$$x_j \approx \left[-4DT \ln \left(\frac{N_B}{Q_I} \sqrt{\pi DT} \right) \right]^{1/2} \quad (4.5)$$

The effective resistor depth is actually reduced from the junction depth by the portion of the depletion region on the resistor side of the metallurgical junction. Several workers have developed relatively simple

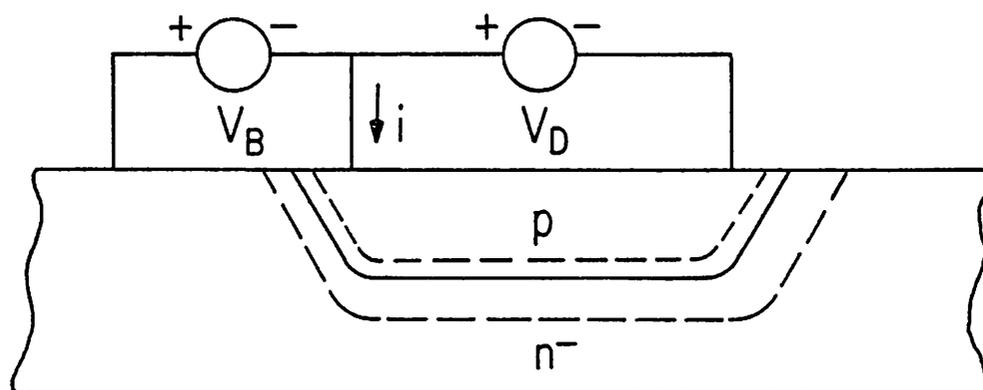


Fig. 4.3 Cross-section of a diffused resistor under bias, showing incursion of the depletion region into the diffusion.

means for evaluating the various depletion widths in certain types of diffused junctions [61-62], however, neither is as complete or as accurate as the original treatment by Lawrence and Warner [63]. Using a similar procedure (Appendix B.1), it may be shown that the electric field in the resistor (or left) depletion region, is given by

$$E_1(x) = \frac{qN_s}{\epsilon_s} \left[\sqrt{\pi DT} \left[\operatorname{erf} \left(\frac{x}{\sqrt{4DT}} \right) - \operatorname{erf} \left(\frac{x_L}{\sqrt{4DT}} \right) \right] - \frac{N_B}{N_s} (x - x_L) \right] \quad (4.6)$$

while in the substrate (or right) side

$$E_2(x) = \frac{qN_s}{\epsilon_s} \left[\sqrt{\pi DT} \left[\operatorname{erf} \left(\frac{x}{\sqrt{4DT}} \right) - \operatorname{erf} \left(\frac{x_R}{\sqrt{4DT}} \right) \right] - \frac{N_B}{N_s} (x - x_R) \right] \quad (4.7)$$

where x_L and x_R represent the left and right edges of the depletion region. For a specified x_R (or x_L), the opposite side of the junctions depletion region may be determined by solving for the case $E_1(x_j) = E_2(x_j)$. The electrostatic potential on each side of the junction, is then found to be (Appendix B.1)

$$V_1 = \frac{-qN_s}{\epsilon_s} \left[\sqrt{\pi DT} x_j \left(\operatorname{erf} \left(\frac{x_j}{\sqrt{4DT}} \right) - \operatorname{erf} \left(\frac{x_L}{\sqrt{4DT}} \right) \right) + 2DT \left(e^{-x_j^2/4DT} - e^{-x_L^2/4DT} \right) - \frac{N_B}{2N_s} (x_j - x_L)^2 \right] \quad (4.8)$$

and

$$\begin{aligned}
V_2 = \frac{qN_s}{\epsilon_s} \left[\sqrt{\pi DT} x_j \left(\operatorname{erf} \left(\frac{x_j}{\sqrt{4DT}} \right) - \operatorname{erf} \left(\frac{x_R}{\sqrt{4DT}} \right) \right) \right. \\
+ 2DT \left(e^{-x_j^2/4DT} - e^{-x_R^2/4DT} \right) \\
\left. - \frac{N_B}{2N_s} (x_j - x_R)^2 \right] \quad (4.9)
\end{aligned}$$

The total potential across the junction is then simply

$$V_T = V_1 + V_2 \quad (4.10)$$

which equals the combined built-in voltage and applied reverse bias, or

$$V_T = \phi + V_B \quad (4.11)$$

where

$$\phi = \frac{kT}{q} \ln \left[\frac{N(x_L)N(x_R)}{n_i^2} \right] \quad (4.12)$$

By iteration, the above may be solved for x_L and x_R given an applied V_B (Appendix B.3).

To complete the resistivity determination, it is necessary to assume a specific form for the mobility and solve Eq. (4.1). Defining the mobility to be¹

$$\mu_c(x) = KN_c^m(x) \quad (4.13)$$

where K is always positive and typically $0 > m > -.5$, Eq. (4.1) then becomes

$$R \approx \frac{L}{W} \left[\int_0^{x_L} qKN_s^m e^{-mx^2/4DT} \left(N_s e^{-x^2/4DT} - N_B \right) dx \right] \quad (4.14)$$

¹This is a similar expression to that used by Irwin [64] for piecewise mobility approximations; $\mu_c(x) = K[N_c(x) - N_B]^m$. Irwin's expression leads to more involved integrations, however, and is apparently less applicable when $N_c(x) \approx N_B$.

or in terms of the conductivity ($G = 1/R$)

$$G = \frac{WqKN_s^m}{L} \left[\int_0^{X_L} N_s e^{-(m+1)x^2/4DT} dx - \int_0^{X_L} N_B e^{-mx^2/4DT} dx \right] \quad (4.15)$$

As

$$\int_0^x e^{-a^2 t^2} dt = \frac{\sqrt{\pi}}{2a} \operatorname{erf}(ax) \quad (4.16)$$

then

$$\int_0^{X_L} e^{-(m+1)x^2/4DT} dx = \sqrt{\frac{\pi DT}{(m+1)}} \operatorname{erf}\left(X_L \sqrt{\frac{(m+1)}{4DT}}\right) \quad (4.17)$$

Further,

$$\int_0^x e^{t^2} dt = \sum_{i=0}^N \frac{x^{2i+1}}{i!(2i+1)} + e \quad (4.18)$$

where

$$|e| < \frac{1}{N!(2N+1)} \quad \text{for } x \leq 1 \quad (4.19)$$

Thus

$$\int_0^{X_L} e^{-mx^2/4DT} dx \approx \sqrt{\frac{4DT}{-m}} \sum_{i=0}^N \frac{\left(X_L \sqrt{\frac{-m}{4DT}}\right)^{2i+1}}{i!(2i+1)} \quad (4.20)$$

so finally

$$G \approx \frac{WqKN_s^m}{L} \left[N_s \sqrt{\frac{\pi DT}{(m+1)}} \operatorname{erf}\left(X_L \sqrt{\frac{(m+1)}{4DT}}\right) + N_B \sqrt{\frac{4DT}{-m}} \sum_{i=0}^N \frac{\left(X_L \sqrt{\frac{-m}{4DT}}\right)^{2i+1}}{i!(2i+1)} \right] \quad (4.21)$$

The bias voltage coefficient may be obtained by differentiating the above, or more easily, by differentiating Eq. (4.1). Specifically

$$\frac{1}{R} \frac{R}{\partial V_B} = \frac{L}{WR} \frac{\partial x_L}{\partial V_B} \frac{\partial}{\partial x_L} \left[\int_0^{x_L} \sigma(x) dx \right]^{-1} \quad (4.22)$$

$$= - \frac{L}{WR} \frac{\partial x_L}{\partial V_B} \frac{\sigma(x_L)}{x_L^2 \sigma^2} \quad (4.23)$$

$$= -\rho_s(V_B) \frac{\partial x_L}{\partial V_B} \sigma(x_L) \quad (4.24)$$

where $\rho_s(V_B)$ is given by Eq. (4.21) ($\rho_s(V_B) = W/(LG(V_B))$) and $\partial x_L/\partial V_B$ must be determined numerically.

Plots of calculated ρ_s and normalized bias voltage coefficients for several CMOS planar well resistors (which assume $W \gg x_j$) are shown in Fig. 4.4.

4.3 Characteristics of Diffused Resistors Due to Sidewalls

In diffused resistors where it is not the case that $W \gg x_j$, the effects of impurity outdiffusion from the sides and ends of the mask cut must be considered. This outdiffusion may modify both the resistivity and the voltage coefficient, in some cases dramatically. A simple approximation for this situation is to assume that the outdiffused sidewalls are cylindrical, while the four corners are spherical [65], as shown in Fig. 4.5.

For predepositions from a gaseous source, the total deposited impurity " Q_T ," will usually increase, due to enhanced diffusion away from the surface near the mask edges. For implanted depositions, however, the total dose remains the same, as this is only determined by the mask opening size and implant density. This results in a lower average impurity concentration as the dose is now distributed over a larger volume (planar region plus sidewalls). This situation will be considered in detail.

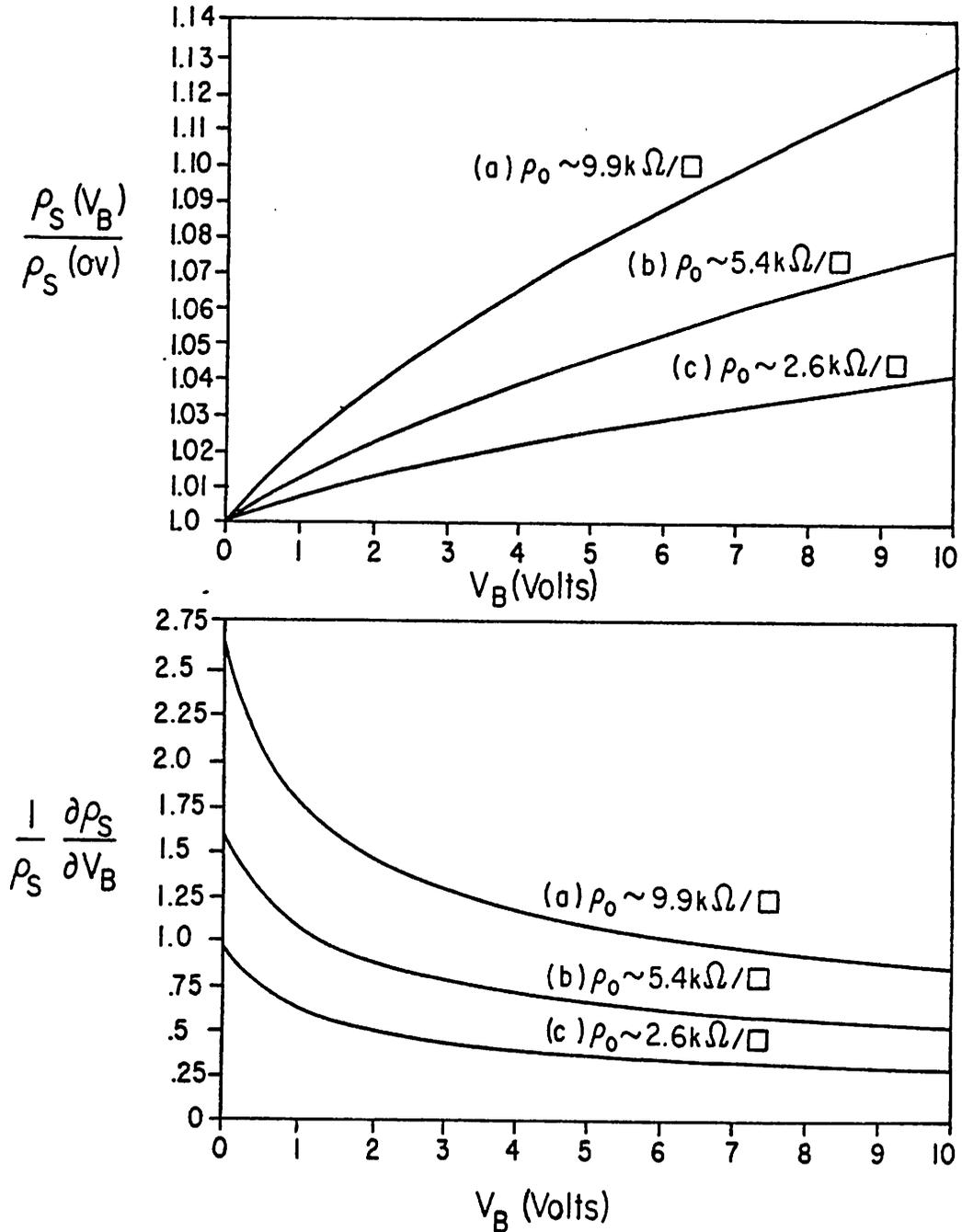


Fig. 4.4 Normalized planar resistivity bias dependence (top) and bias voltage coefficient (bottom) for three typical CMOS well resistors. In curve (a); $Q_I = 1.75 \times 10^{12} / \text{cm}^2$, $DT = 1.95 \times 10^{-8} \text{ cm}^2$, and $N_B = 5 \times 10^{14} / \text{cm}^3$. Multiply values by 2 for curve (b) and by 4 for curve (c).

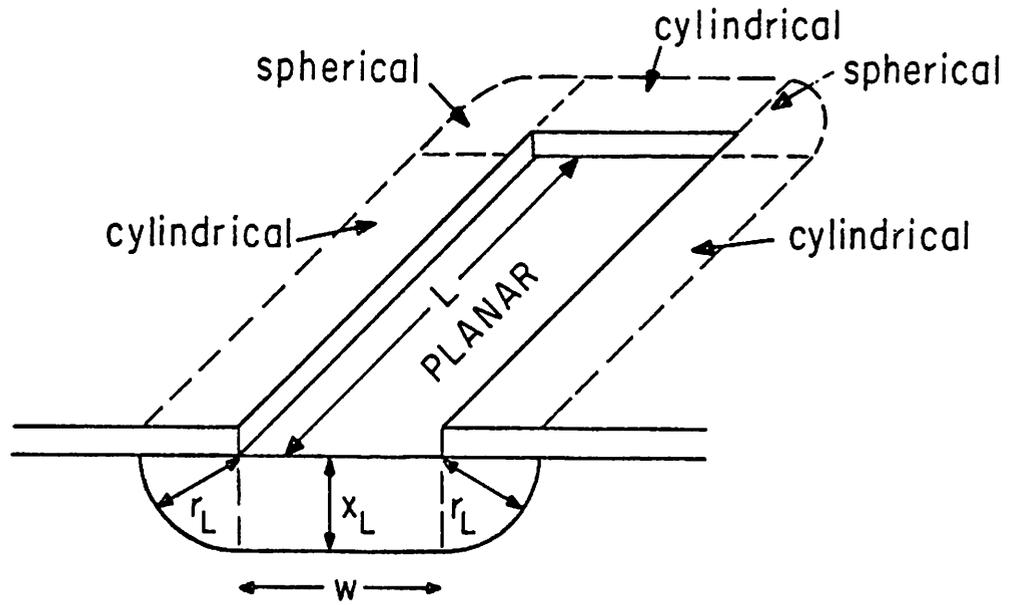


Fig. 4.5 Geometrically simple model of impurity outdiffusion in a diffused resistor.

For a mask opening of length L and width W ($L \gg W$),

$$2Q_{\text{side}} + Q_I LW = Q_T LW \quad (4.25)$$

where $Q_I LW$ is the total charge in the planar region after outdiffusion, Q_{side} is the charge present in each cylindrical sidewall and Q_T is the total ion dose in ions/cm². Note that the constraint $L \gg W$ allows elimination of the corner and endwall outdiffusion terms.

The total dose in one sidewall, is given by

$$Q_{\text{side}} = L \int_0^{\pi/2} d\phi \int_0^{\infty} r N_c(r) dr \quad (4.26)$$

which for a gaussian impurity distribution reduces to

$$Q_{\text{side}} = \pi N_s DTL \quad (4.27)$$

where

$$N_s = \frac{Q_I}{\sqrt{\pi DT}} \quad (4.28)$$

By combining terms,

$$Q_I = \frac{Q_T}{(1 + 2\sqrt{\pi DT}/W)} \quad (4.28)$$

Thus Q_I is reduced from Q_T by an amount which is dependent upon W . The junction depth will also be reduced from the planar case, and is given by

$$X_J = \left[-4DT \ln \left[\frac{N_B}{Q_T} \sqrt{\pi DT} \left(1 + \frac{2\sqrt{\pi DT}}{W} \right) \right] \right]^{1/2} \quad (4.29)$$

Plots of Q_I/Q_T and $X_J(W)/X_J(\infty)$ for various diffusion cases, are shown in Figs. 4.6 and 4.7.

The conductivity of a sidewall may now be determined in a manner similar to that used in the planar case. First, the depletion region

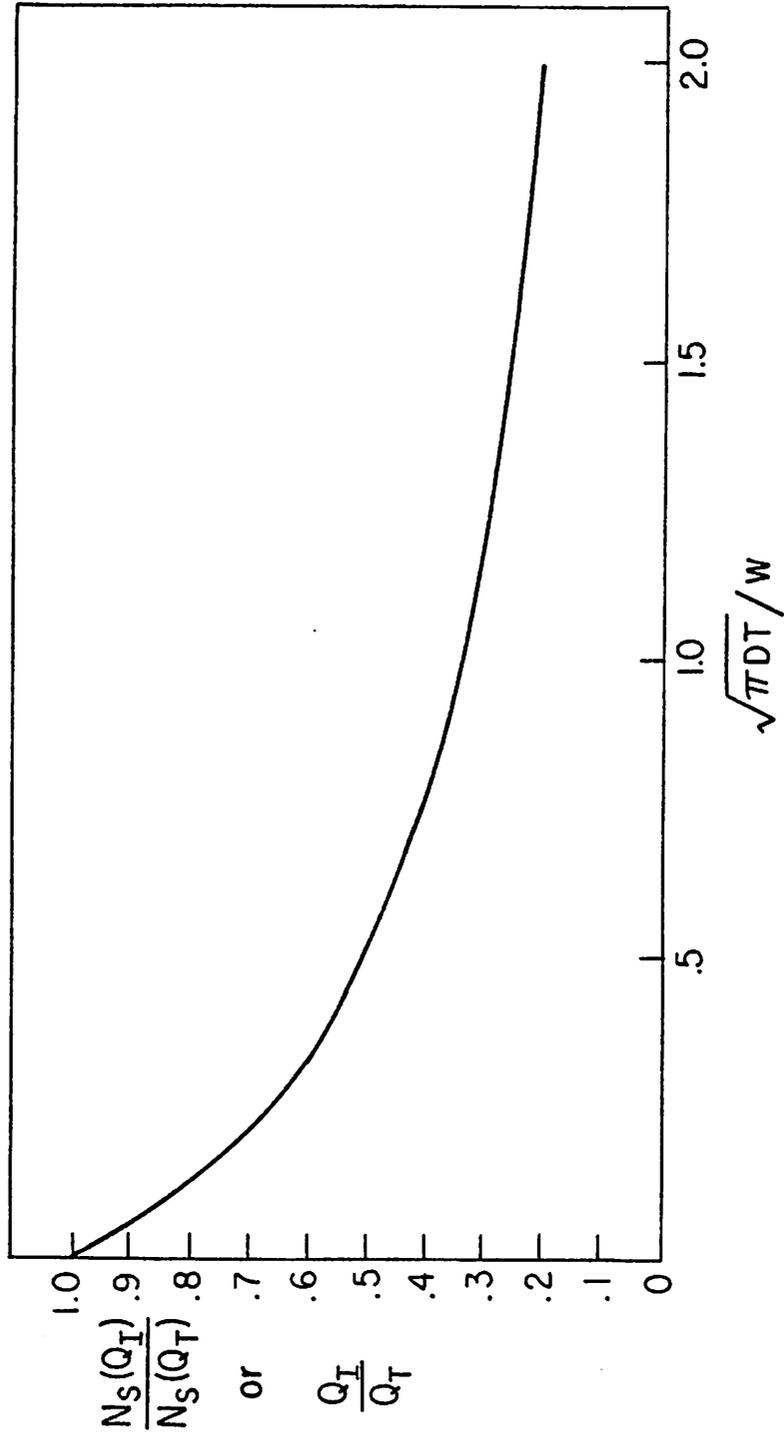


Fig. 4.6 Normalized surface concentration or effective implant dose as a function of $\sqrt{\pi D T} / w$.

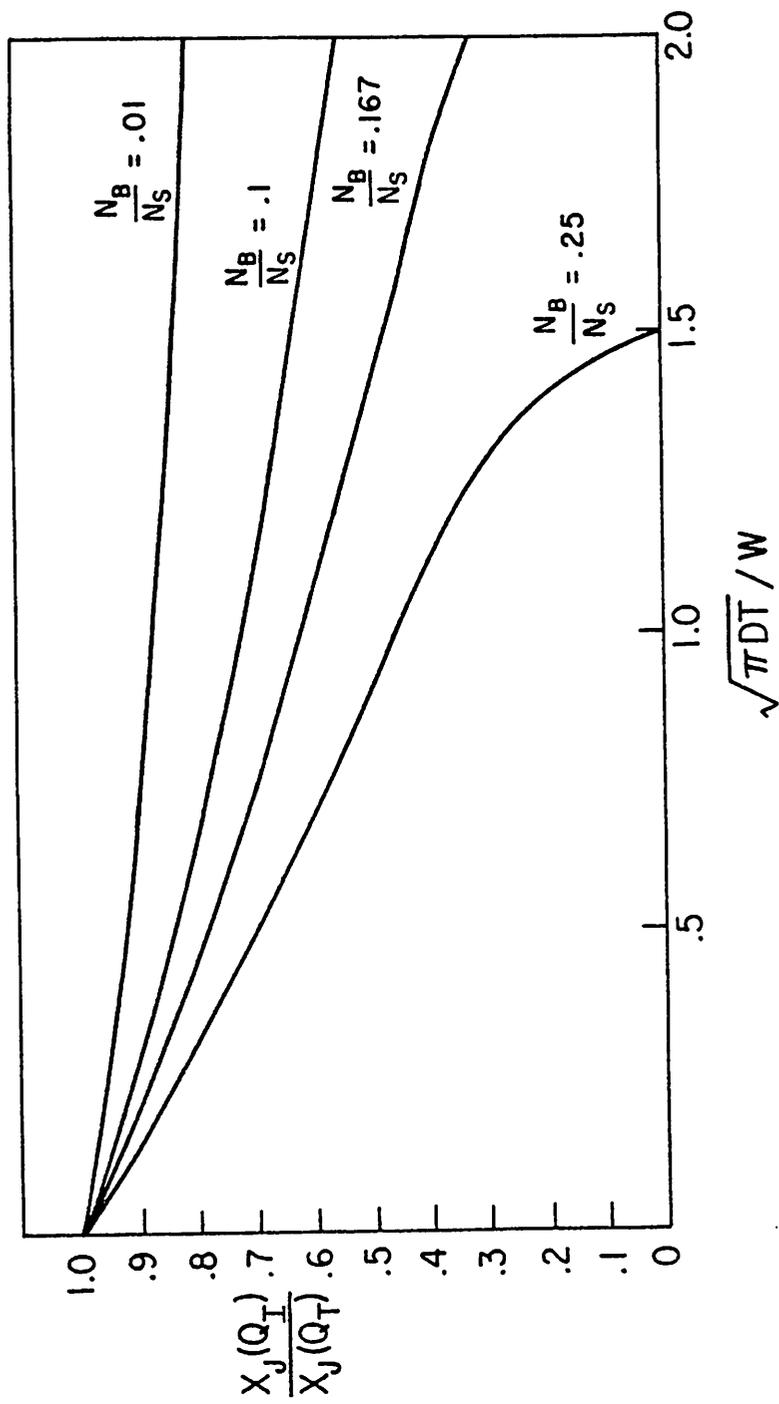


Fig. 4.7 Normalized junction depth as a function of $\sqrt{\pi DT} / W$ and planar N_B / N_S .

thickness on the resistor side of the junction must be determined. It may be shown (Appendix B.2), that for this case the electric field on the diffusion side of the junction is given by

$$E_1(r) = -\frac{2DTqN_s}{r\epsilon_s} \left[e^{-r^2/4DT} - e^{-r_L^2/4DT} \right] - \frac{qN_B}{2r\epsilon_s} (r^2 - r_L^2) \quad (4.30)$$

while on the substrate side

$$E_2(r) = -\frac{2DTqN_s}{r\epsilon_s} \left[e^{-r^2/4DT} - e^{-r_R^2/4DT} \right] - \frac{qN_B}{2r\epsilon_s} (r^2 - r_R^2) \quad (4.31)$$

where r_L and r_R indicate the inside and outside edges of the cylindrical depletion region. For a specified r_R (or r_L) the opposite side of the depletion region r_L (or r_R) may be determined by solving for the case $E_1(r_j) = E_2(r_j)$. The electrostatic potential on each side of the junction is then found to be (Appendix B.2)

$$V_1 \approx -\frac{q}{\epsilon_s} \left[-\frac{N_B}{4} (r_j^2 - r_L^2) + \frac{N_B}{2} r_L^2 \ln\left(\frac{r_j}{r_L}\right) + 2DT N_s e^{-r_L^2/4DT} \ln\left(\frac{r_j}{r_L}\right) - 2DT N_s [f(r_j) - f(r_L)] \right] \quad (4.32)$$

and

$$V_2 \approx \frac{q}{\epsilon_s} \left[-\frac{N_B}{4} (r_j^2 - r_R^2) + \frac{N_B}{2} r_R^2 \ln\left(\frac{r_j}{r_R}\right) + 2DT N_s e^{-r_R^2/4DT} \ln\left(\frac{r_j}{r_R}\right) - 2DT N_s [f(r_j) - f(r_R)] \right] \quad (4.33)$$

where

$$f(r) = \ln(r) + \sum_{i=1}^N \frac{(r^2/4DT)^i (-1)^i}{i! (2i)} \quad (4.34)$$

The total potential across the cylindrical junction is simply

$$V_T = V_1 + V_2 \quad (4.35)$$

which equals the combined built-in voltage and applied reverse bias, or

$$V_T = \phi + V_B \quad (4.36)$$

where

$$\phi = \frac{kT}{q} \ln \left[\frac{N(r_L)N(r_R)}{n_i^2} \right] \quad (4.37)$$

By iteration, the above may be solved for r_L and r_R given an applied V_B (Appendix B.3).

Assuming the same mobility dependence as in the planar case

$$\mu_c(r) \approx K N_c^m(r) \quad (4.38)$$

the sidewall conductivity then becomes

$$G_{\text{side}} \approx \frac{W\pi q K N_s^m}{2L} \left[N_s \int_0^{r_L} e^{-(m+1)r^2/4DT} r dr - N_B \int_0^{r_L} e^{-mr^2/4DT} r dr \right] \quad (4.39)$$

As

$$\int_b^c e^{ax^2} x dx = \frac{e^{ax^2}}{2a} \Big|_b^c \quad (4.40)$$

then

$$G_{\text{side}} \approx \frac{DTW\pi q K N_s^m}{L} \left[\frac{N_s}{(m+1)} \left[e^{-(m+1)r_L^2/4DT} - 1 \right] - \frac{N_B}{m} \left[e^{-mr_L^2/4DT} - 1 \right] \right] \quad (4.41)$$

The voltage coefficient of the above is obtainable by differentiating

$$\frac{\partial G_{\text{side}}}{\partial V_B} = \frac{\pi q K W r_L}{2L} \frac{\partial r_L}{\partial V_B} N_s^m \left[N_s e^{-(m+1)r_L^2/4DT} - N_B e^{-mr_L^2/4DT} \right] \quad (4.42)$$

where $\partial r_L / \partial V_B$ must be evaluated numerically.

The entire resistance and voltage coefficient of a diffused resistor with outdiffused sidewalls may now be determined. Specifically

$$R_T = \frac{R}{2R G_{\text{side}} + 1} \quad (4.43)$$

and

$$\frac{\partial R_T}{\partial V_B} = \frac{1}{[2R G_{\text{side}} + 1]} \left[\frac{\partial R}{\partial V_B} - \frac{2R}{[2R G_{\text{side}} + 1]} \left[G \frac{\partial R}{\partial V_B} + \frac{R \partial G}{\partial V_B} \right] \right] \quad (4.44)$$

Plots of normalized resistivity bias voltage coefficient as a function of applied bias and resistor width are shown in Fig. 4.8 for a typical CMOS well resistor.

Lastly, the voltage dependence for R_T with respect to V_D must be determined. Assuming

$$R_T = \frac{1}{W} \int_0^L \rho(y) dy \quad (4.45)$$

and assuming that a linear IR drop exists along the length of the resistor

$$\rho(y) = \rho + \frac{\partial \rho}{\partial V} \frac{y}{L} V_D \quad (4.46)$$

and as

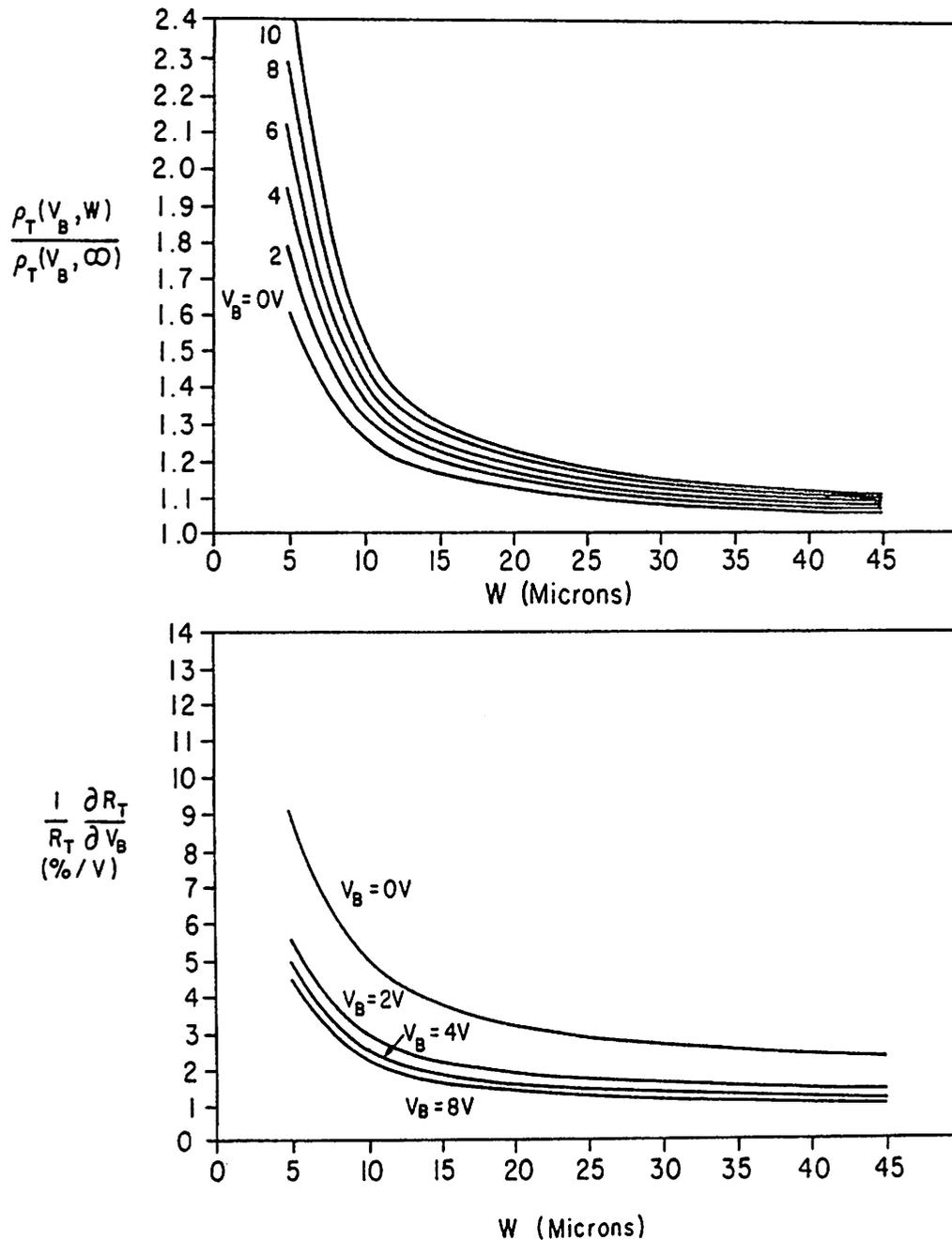


Fig. 4.8 (a) Normalized resistivity width dependence as a function of applied bias. (b) Normalized resistance bias voltage coefficient as a function of width and applied bias. For each figure; $Q_T = 3.5 \times 10^{12}/\text{cm}^2$, $DT = 3.91 \times 10^{-8} \text{ cm}^2$, $N_B = 1 \times 10^{15}/\text{cm}^3$ and $\rho_s(0V) = 5.4 \text{ k}\Omega/\square$.

$$\frac{\partial \rho}{\partial V} = \frac{\partial R_T}{\partial V_B} \frac{W}{L} \quad (4.47)$$

then

$$R_T = \frac{1}{W} \int_0^L \left[\rho + \frac{\partial R_T}{\partial V_B} \frac{Wy}{L^2} V_D \right] dy \quad (4.48)$$

so, for small values of V_D

$$\frac{\partial R_T}{\partial V_D} \approx \frac{1}{2} \frac{\partial R_T}{\partial V_B} \quad (4.49)$$

This expression actually indicates the average of the resistivities in the area near $y=L$ and the area near $y=0$, because of the assumption that the I-R drop along the resistor be linear in y .

4.4 Compensation of Voltage Dependence in a Diffused Monolithic Resistor

From the discussion in the preceding sections it is evident that voltage dependent behavior may substantially alter the matching characteristics of diffused resistors. In CMOS technologies, however, there are several techniques for reducing the size of these effects. In one method, the resistor to be matched is placed in an isolated well, which is biased at a variable voltage. Applications of this technique to a current to voltage converter is shown in Fig. 4.9. Using this technique [66], effective voltage coefficients which are on the order of those obtainable with polysilicon resistors have been achieved ($\sim .001\%V$).

Using a different technique, equivalent or lower, voltage coefficients should be obtainable. By again placing the precision resistor inside a well, but now inducing an I-R drop in the well which is similar to that in the precision resistor, the n^+ depletion region width is kept fairly constant. The well may either be placed in parallel with the n^+

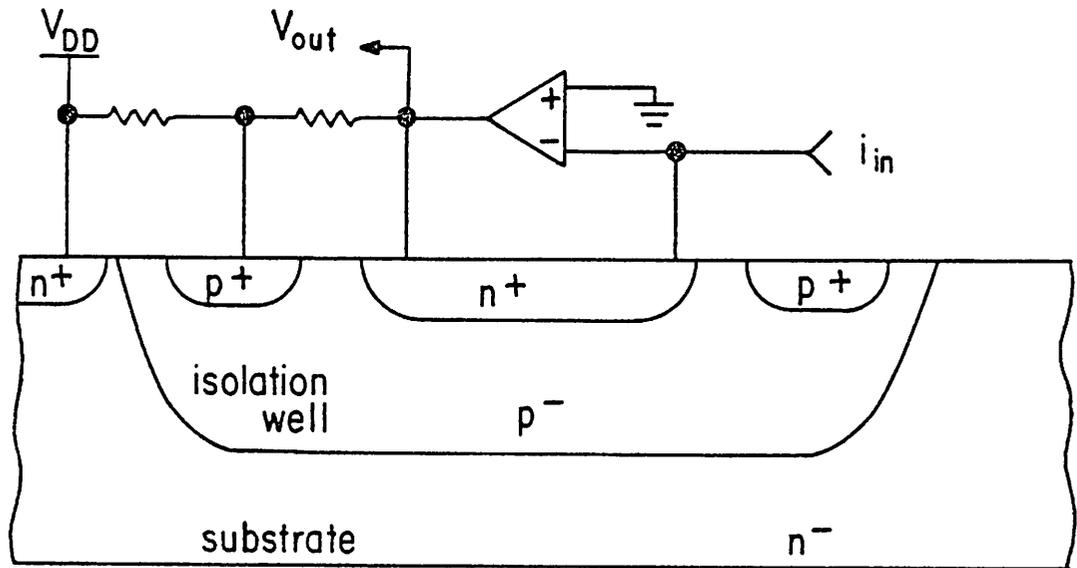


Fig. 4.9 Variable well bias may be used to reduce the voltage coefficient of a diffused resistor. A current to voltage converter is shown [66].

resistor, as shown in Fig. 4.10a, or biased independently as shown in Fig. 4.10b. In the former case, the voltage dependence of the well resistivity will usually dominate, while in the latter, the nonlinear voltage drop along the well caused by its voltage dependence, will create a small bias dependence in the $n+$ resistivity. These effects will be analyzed in detail.

Assuming that the well resistivity may be represented as

$$\rho_1 = K_1 (V_D + V_B + \phi_1)^n \quad (4.50)$$

where V_D and V_B are as defined in Fig. 4.3, then the voltage drop at any point along its length may be given by

$$\begin{aligned} V(Y) &= \int_0^Y I \rho_1(t) dt \\ &= I K_1 \int_0^Y \left(\frac{t}{L} V_D + V_B + \phi_1 \right)^n dt \end{aligned} \quad (4.51)$$

As the nominal I-R drop is simply

$$V_{\text{nom}}(Y) = \frac{Y}{L} V_D \quad (4.52)$$

this indicates a nonlinear error voltage of the amount

$$\begin{aligned} V_e(Y) &= \frac{IK_1 L}{V_B W_1^{(n+1)}} \left[\left(\frac{Y}{L} V_D + V_B + \phi_1 \right)^{n+1} - (V_B + \phi_1)^{n+1} \right] \\ &\quad - \frac{Y}{L} V_D \end{aligned} \quad (4.53)$$

where the well current is $V_D/R_1(L)$. As

$$R_1(L) = \frac{1}{W_1} \int_0^L K_1 \left(\frac{t}{L} V_D + V_B + \phi_1 \right)^n dt \quad (4.54)$$

this indicates that

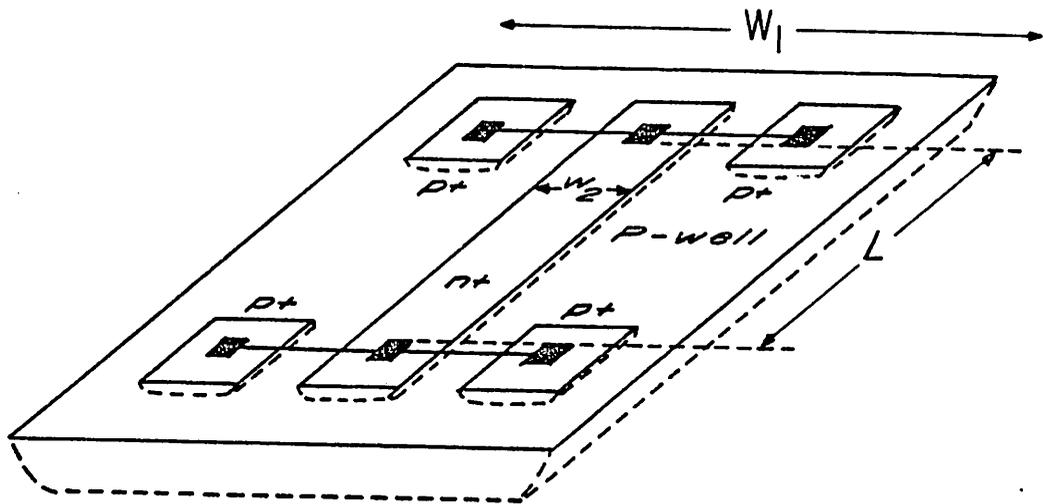


Fig. 4.10a Low voltage coefficient resistor.

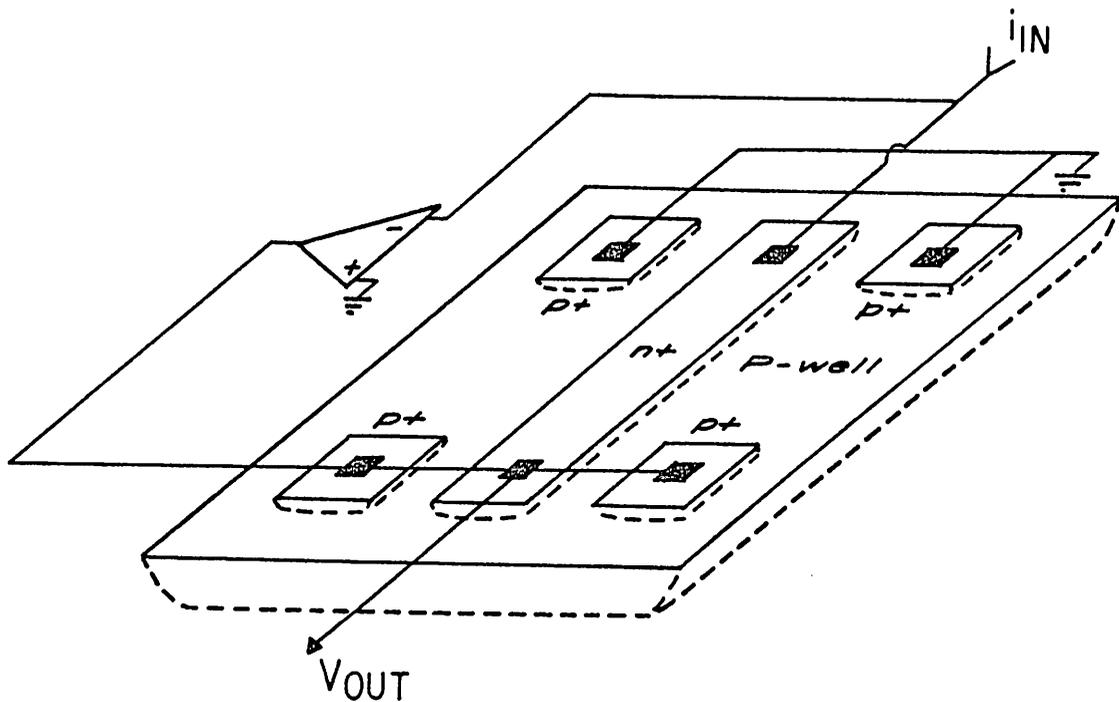


Fig. 4.10b Improved low voltage coefficient resistor configuration in a current to voltage converter.

$$I = \frac{V_D^2 W_1 (n+1)}{LK_1 [(V_D + V_B + \phi_1)^{n+1} - (V_B + \phi_1)^{n+1}]} \quad (4.55)$$

Assuming that the error voltage is small the $n+$ resistance may be given by

$$R_2(Y) = \frac{1}{W_2} \int_0^Y \rho_2 + \frac{\partial \rho_2}{\partial V} V_e(t) dt \quad (4.56)$$

$$\sim \frac{Y\rho_2}{W_2} + \frac{1}{W_2} \frac{\partial \rho_2}{\partial V} \left[\frac{IK_1 L}{V_D W_1 (n+1)} \left(\frac{L}{V_D (n+2)} \left(\left(\frac{YV_D}{L} + V_B + \phi_1 \right)^{n+2} - (V_B + \phi_1)^{n+2} \right) - Y(V_B + \phi_1)^{n+1} \right) - \frac{1}{2} \frac{Y^2 V_D}{L} \right] \quad (4.57)$$

and

$$R_2(L) = \frac{L\rho_2}{W_2} + \frac{L}{W_2} \frac{\partial \rho_2}{\partial V} \left[\frac{IK_1 L}{V_D W_1 (n+1)} \left(\frac{1}{V_D (n+2)} \left((V_D + V_B + \phi_1)^{n+2} - (V_B + \phi_1)^{n+2} \right) - (V_B + \phi_1)^{n+1} \right) - \frac{V_D}{2} \right] \quad (4.58)$$

In the first case considered (Fig. 4.10a), the total resistance is the parallel combination of $R_1(L)$ and $R_2(L)$, while in the second case (Fig. 4.10b), the total impedance is simply $R_2(L)$. Plots of absolute resistor error for each of these cases when applied to fairly conventional $n+$ and p -well resistors are shown in Fig. 4.11.

4.5 Temperature Coefficient of Resistors

Because electron and hole drift mobilities in silicon are temperature as well as concentration dependent, silicon resistors of any type will display temperature dependent behavior. This is also true of thin-film metallic resistors, although often to a smaller degree. A plot of electron and hole mobilities in silicon as a function of dopant concentration

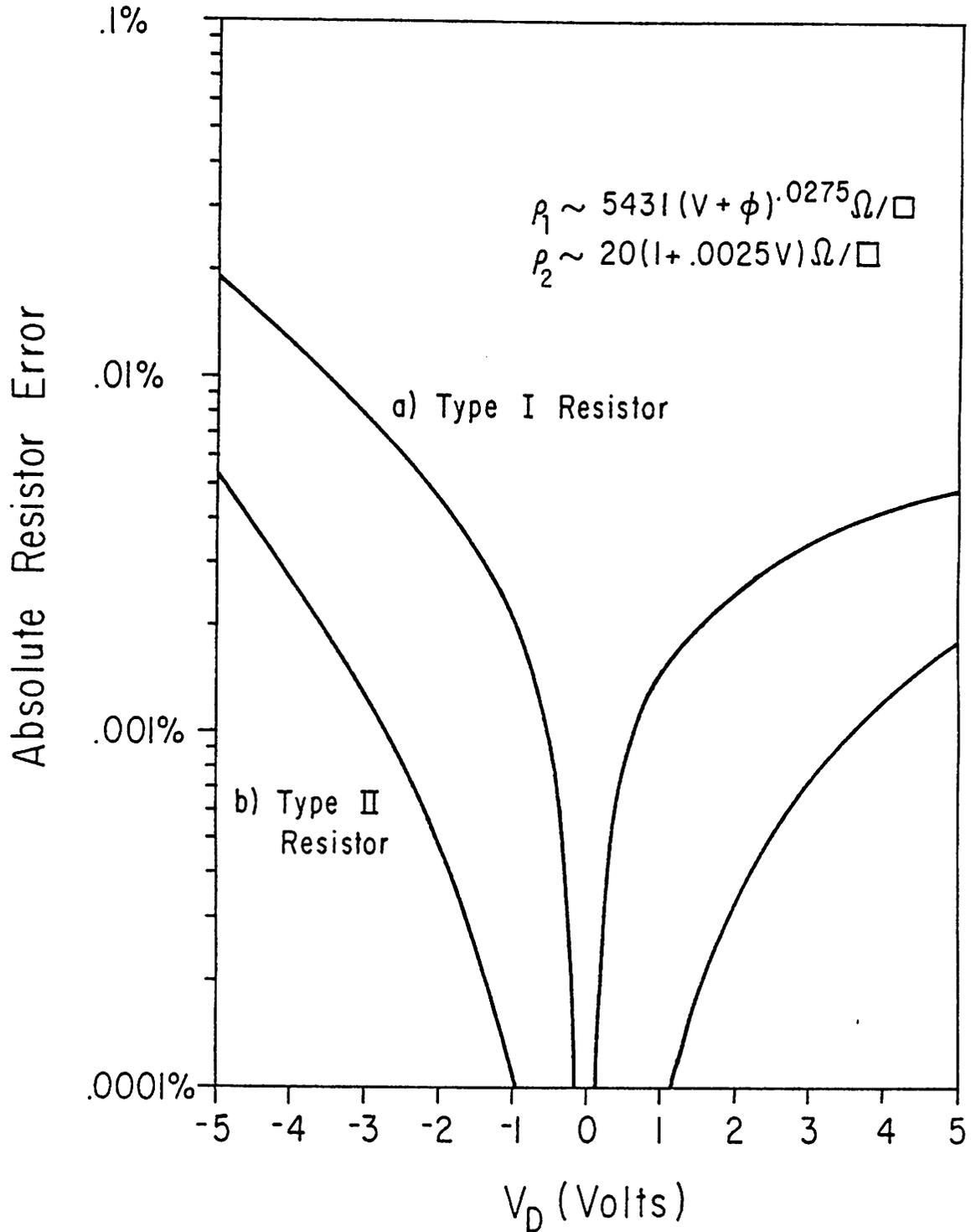


Fig. 4.11 Resistor error as a function of voltage drop for the resistor configurations shown in Fig. 4.10. The type I resistor is for R_1 (well resistor) in parallel with R_2 (n+ resistor), while in the type II resistor they are electrically isolated.

and temperature is presented in Fig. 4.12 [67]. For deposited resistors of uniform doping, the temperature dependence may be estimated directly from this figure. For diffused resistors a slight modification to the previous resistivity analysis, will allow resistance determination at any temperature. By assuming that the previously defined mobility coefficients K and m are not constant in a given region, but are in fact temperature dependent, all of the previous analysis will still remain valid. Thus it is now assumed that

$$\mu(N_c, T) = K(T)N_c^{m(T)} \quad (4.59)$$

where as before K and m are defined for a region of values of N_c . Note that K and m are easily obtainable from Fig. 4.12, by determining the best fit of the equation

$$\ln \mu = m \ln N_c + \ln K \quad (4.60)$$

in the temperature and doping range of interest.

4.6 Matching Consideration in Precision Capacitors

An alternative to the precision resistor, is the precision MOS capacitor, which has recently gained considerable acceptance. Depending upon the specific type of IC process employed, capacitors with metal or poly top plates and poly or diffusion bottom plates may be available, with usually a fairly thin SiO_2 layer acting as the intermediate dielectric. In certain processes, however, no precision capacitors are directly available (as in a single-poly self-aligned process for instance) which forces the use of precision resistors.

Consider the general pair of matched capacitors shown in Fig. 4.13. If the oxide thickness is assumed to be fairly uniform, and the top and

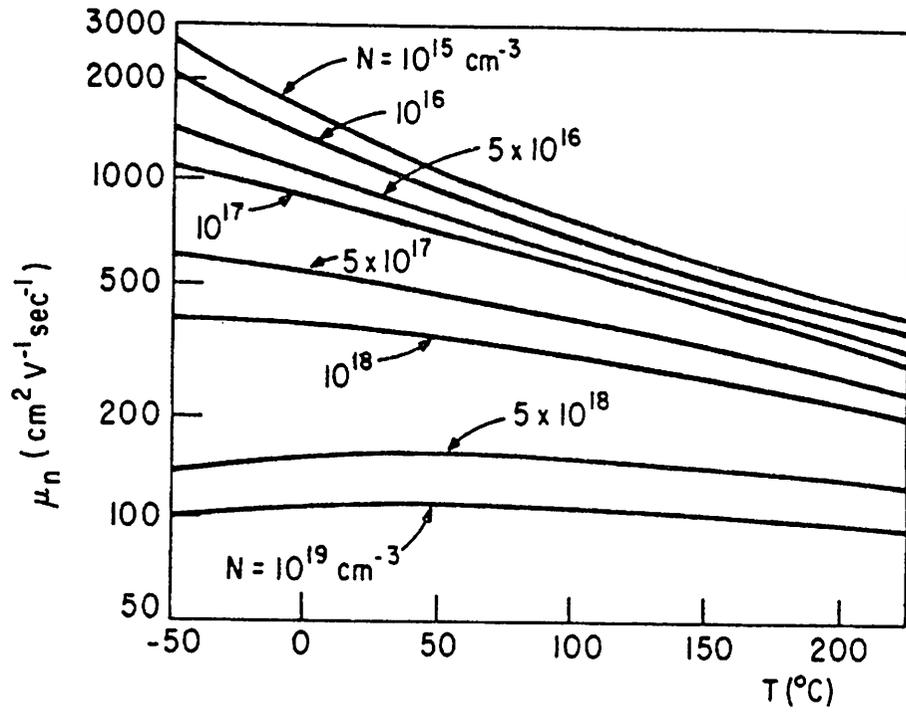


Fig. 4.12 Electron mobility as a function of doping concentration and temperature [67].

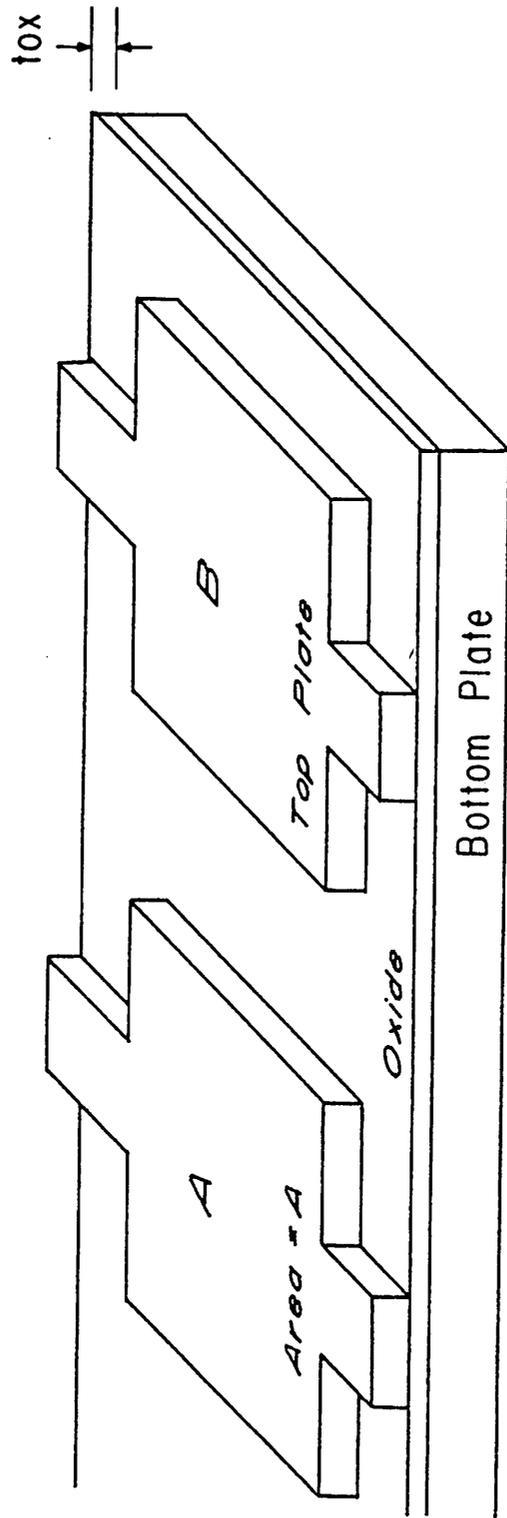


Fig. 4.13 A matched pair of MOS capacitors.

bottom plates are assumed to be either metallic or degenerately doped, the value of either capacitor may be given by

$$\begin{aligned} C &= \epsilon_{\text{ox}} \int \frac{1}{t_{\text{ox}}} dA \\ &= \frac{A\epsilon_{\text{ox}}}{\overline{t_{\text{ox}}}} \end{aligned} \quad (4.61)$$

where $\overline{t_{\text{ox}}}$ is the average oxide thickness in the capacitor, and A is the total overlap area. If an uncertainty δ is assigned to each capacitor parameter, a potential error will result in the final capacitor value of the amount

$$\frac{\delta C}{C} = \left[\left(\frac{\delta A}{A} \right)^2 + \left(\frac{\delta \overline{t_{\text{ox}}}}{\overline{t_{\text{ox}}}} \right)^2 \right]^{1/2} \quad (4.62)$$

For approximately rectangular capacitor structures, this becomes

$$\frac{\delta C}{C} = \left[\left(\frac{\delta W}{W} \right)^2 + \left(\frac{\delta L}{L} \right)^2 + \left(\frac{\delta \overline{t_{\text{ox}}}}{\overline{t_{\text{ox}}}} \right)^2 \right]^{1/2} \quad (4.63)$$

As in the resistor case, the relevant uncertainties in the above, usually indicate the spread within the precision components on a single die, and not the variation across different wafers or lots. Unlike the situation with resistors, the aspect ratio of capacitors is usually near unity ($L/W \approx 1$), thus $\delta W/W$ and $\delta L/L$ will normally contribute about equally to overall capacitance error. Variations in $\overline{t_{\text{ox}}}$ may be due to thermal gradients, dopant inhomogeneities or crystal defects which are present during processing (if thermally grown), or silane concentration gradients (if deposited). Capacitor matching of better than .5% in even moderate ratio situations [68] is commonly accepted, and one report indicates that an untrimmed .1% matching is achievable with high yield [69].

Typically, capacitors which must match to greater than 8 or 9 bits and/or

which must be made quite small, are trimmed in quantum amounts by either electrical [70] or fusible link techniques [71].

As was the case with precision resistors, truly random edge variations may only be compensated for by using larger device geometries, although the effect of linear gradients may be reduced by using an interdigitized or common centroid structure (Fig. 4.14a and 4.14b respectively).

4.7 Capacitance of Nonplanar Structures

Depending upon the specifics of an IC process, completed capacitors may not resemble the planar structures illustrated in Fig. 4.13. This is particularly true of capacitor types which are principally defined by a thin oxide mask cut, such as those shown in Fig. 4.15. In cases such as these, the total overlap capacitance is divided into 10 different areas of 4 basic types, as illustrated in Fig. 4.16a. The region of largest capacitance is usually one of planar thin oxide as before, however now, this area is surrounded by slanted sidewall and corner regions and a planar field oxide area. In some circuit applications the contributions of these areas to the total cell capacitance will be unimportant, as circuit operation may depend only upon the absolute ratio of capacitance, and all capacitors are fabricated from unit cells of the same size [72]. In other applications, however, which for reasons of speed must attempt to minimize overall capacitance [73], or in circuits which must use noninteger capacitance ratios (e.g., filters), all of these regions must be taken into account.

A cross section of a thin oxide cut in a nonplanar capacitor, is shown in Fig. 4.16b. As is apparent in the figure, some undercutting from the original mask opening will usually have taken place, and the sidewalls are typically slanted (affords good step coverage). In terms

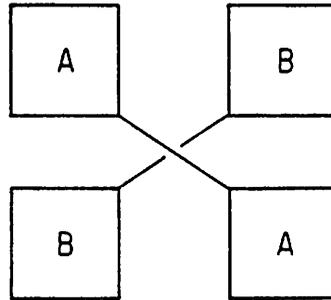


Fig. 4.14a Interdigitating of capacitors to compensate for linear oxide gradients ($C_A \approx C_B$).

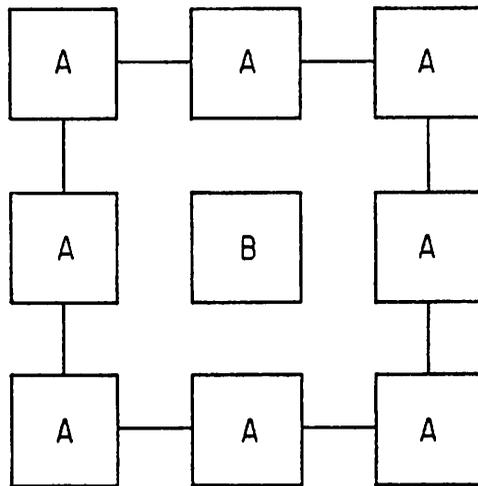


Fig. 4.14b Common centroid geometry ($C_A > C_B$).

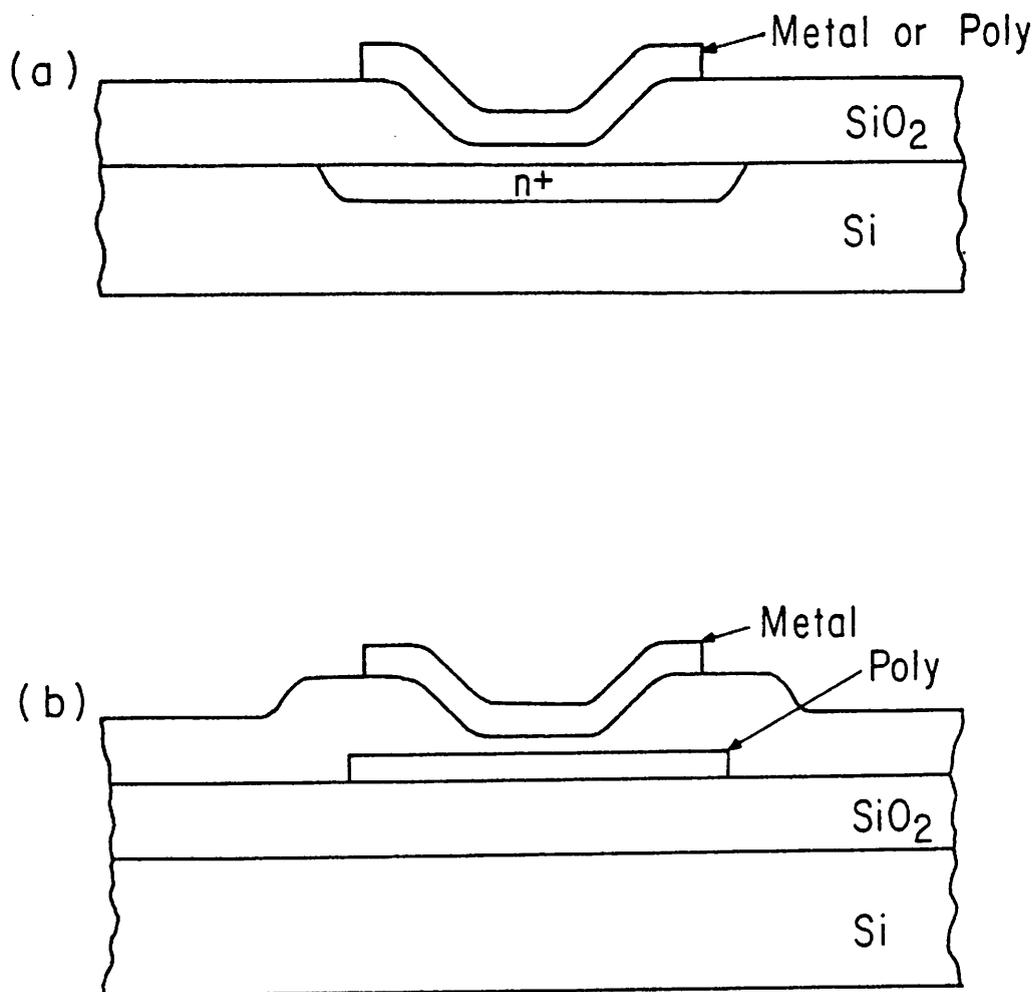


Fig. 4.15 Nonplanar capacitor structures. (a) Metal on poly top plate and diffused bottom plate structure is common in metal gate and slightly modified silicon gate processes. (b) Metal top plate and poly bottom plate is available in some modified silicon gate processes.

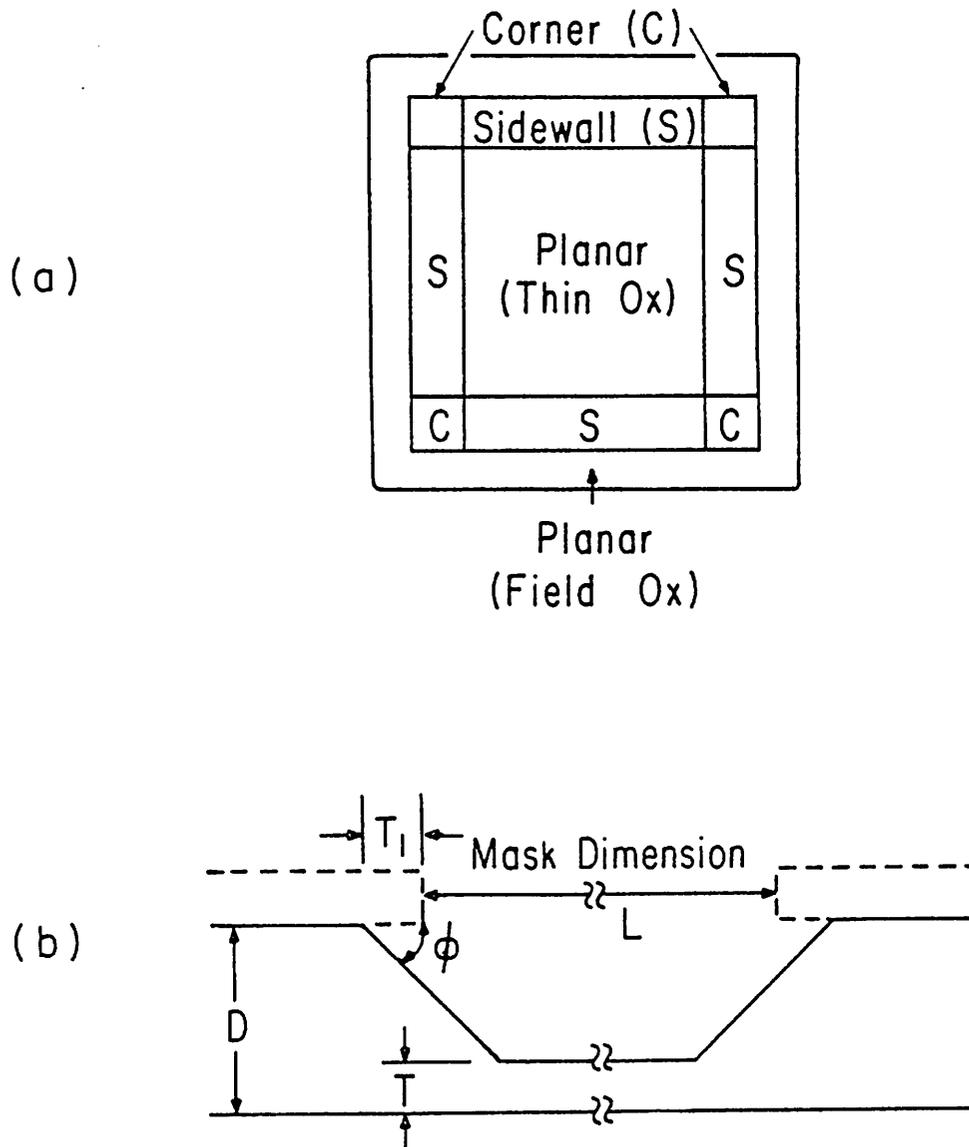


Fig. 4.16 Generalized nonplanar capacitor. (a) Top view showing division of capacitance into 10 different areas of 4 basic types. (b) Capacitor cross-section.

of the various dimensions noted in the figure, this results in a planar thin oxide region of linear dimension B, where

$$B = L + 2T_1 - 2(D - T) \tan \phi \quad (4.64)$$

and a total planar capacitance of the amount

$$C_{pt} = \frac{\epsilon_{ox}}{T} [L + 2T_1 - 2(D - T) \tan \phi]^2 \quad (4.65)$$

The capacitance of a sidewall region may be determined by analyzing the simplified cross section shown in Fig. 4.17a. Assuming that the sidewall is linear as shown, a differential capacitance may be defined of the amount

$$\delta C = \left(\frac{\epsilon_{ox}}{y(x)} \right) \delta x \delta y \quad (4.66)$$

so, the capacitance of a sidewall of length B and width T_2 is given by

$$\begin{aligned} C_s &= \int_0^B \int_0^{T_2} \left(\frac{\epsilon_{ox}}{D + \frac{(T-D)}{T_2} x} \right) dx dy \\ &= \epsilon_{ox} \frac{BT_2}{(T-D)} \ln \left(\frac{T}{D} \right) \end{aligned} \quad (4.67)$$

where T_2 may be defined in terms of the sidewall angle ϕ , by

$$T_2 = (D - T) \tan \phi \quad (4.68)$$

The capacitance of a corner section may be determined in a similar fashion to the sidewall, except that now, the section of length L actually becomes an effective length of $T_2 - x$ (see Fig. 4.17b) so

$$\begin{aligned} C_c &= \int_0^{T_2-x} \int_0^{T_2} \left(\frac{\epsilon_{ox}}{D + \frac{(T-D)}{T_2} x} \right) dx dy \\ &= \int_0^{T_2} (T_2 - x) \left(\frac{\epsilon_{ox}}{D + \frac{(T-D)}{T_2} x} \right) dx \end{aligned}$$

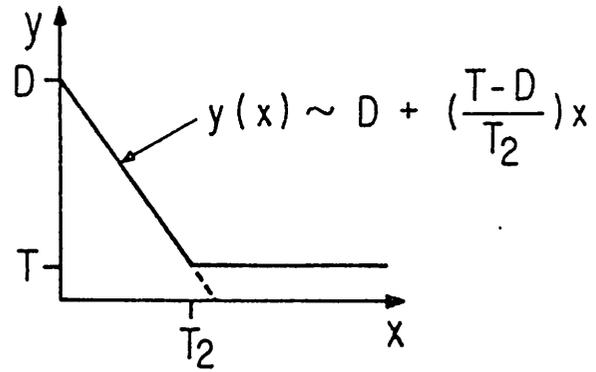


Fig. 4.17a Cross-section of one sloped sidewall.

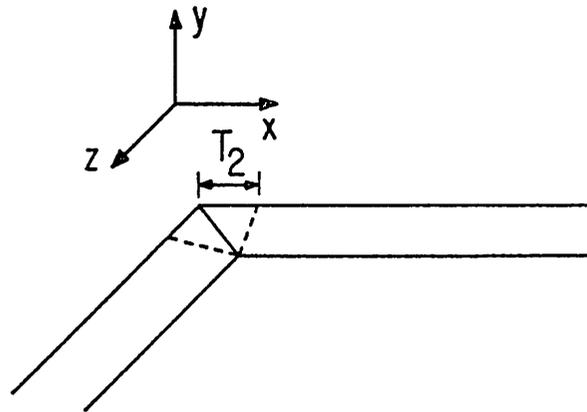


Fig. 4.17b Diagram of corner region.

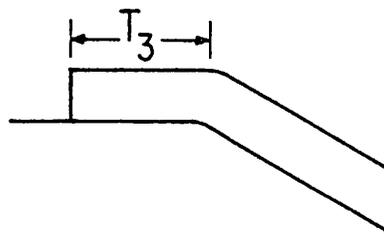


Fig. 4.17c Top plate overlap of field region.

$$= \epsilon_{\text{ox}} \left[(T-D) - \frac{DT_2^2}{(T-D)} \ln \left(\frac{T}{D} \right) \right] \quad (4.69)$$

Lastly the planar capacitance over the field region must be determined. Defining the top plate overlap to be that shown in Fig. 4.17c, the field capacitance is given by

$$C_{\text{pf}} = \frac{\epsilon_{\text{ox}}}{D} [4T_3 (L + 2T_1) + 4T_3^2] \quad (4.70)$$

where the capacitance due to interconnect tabs has not been included.

Thus the total capacitance of a structure of this type may be given by

$$C_T = C_{\text{pt}} + 4C_s + 4C_c + C_{\text{pf}} + C_{\text{tab}}$$

where C_{tab} is any interconnect capacitance, and all the other variables are as previously defined.

4.8 Voltage Coefficient of MOS Capacitors

For capacitors which employ a doped semiconductor top or bottom plate, the preceding capacitance calculations have been slightly simplified. For calculation of voltage and temperature dependence, a number of additional effects must be considered. Because a doped semiconductor does not behave as a metallic material, it will exhibit a finite surface capacitance. That is, a semiconductor which functions as a plate of a capacitor will be unable to supply charge exactly at the oxide interface, but only some distance from it. For a capacitor which employs one metal, and one semiconductor plate, this results in an effective total capacitance of

$$C_T = \frac{C_{\text{ox}} C_s}{C_{\text{ox}} + C_s} \quad (4.71)$$

where C_{OX} is the oxide capacitance and C_S is the effective silicon surface capacitance. A number of techniques have been developed for the evaluation of C_S under different conditions [74]. In general, however, it is simplest to assume that the silicon surface is non-degenerate and fully-ionized [75], which enables the use of Maxwell-Boltzman statistics.

The total voltage applied across an MOS capacitor is given by

$$V_T = V_{FB} + \phi_s - \frac{Q_s}{C_{OX}} \quad (4.72)$$

where V_{FB} is the flatband voltage, ϕ_s is the surface potential and Q_s is the charge density in the silicon. For the case of an n+ bottom plate, this latter quantity can be expressed as [76]

$$Q_s = \frac{-U_s}{|U_s|} [2\epsilon_{si} kT N_D (e^{U_s/kT} - 1)]^{1/2} \quad (4.73)$$

The surface capacitance is then given by

$$C_s = \frac{q}{kT} \frac{\partial Q_s}{\partial U_s} \quad (4.74)$$

which has been shown to be [77]

$$C_s = \frac{q}{kT} \frac{U_s}{|U_s|} \frac{[2\epsilon_{si} kT N_D]^{1/2} e^{U_s/kT} [e^{U_s/kT} - 1]}{[e^{U_s/kT} - 1 - U_s/kT]^{1/2}} \quad (4.75)$$

where U_s is the dimensionless quantity, $U_s = \phi_s q/kT$. The normalized total voltage coefficient may now be determined. Specifically

$$\alpha = \frac{1}{C_T} \frac{\partial C_T}{\partial V_T} \quad (4.76)$$

or eventually

$$\alpha \approx \frac{C_{OX}}{C_S} \frac{\partial C_S}{\partial U_s} \left[\frac{\partial V_T}{\partial U_s} \right]^{-1} \quad (4.77)$$

where $\partial C_s / \partial U_s$ may be found from Eq. (4.74), and $\partial V_T / \partial U_s$ from Eqs. (4.72) and (4.73),

The total voltage coefficient may then be found to be [78]

$$\alpha = \frac{C_{ox}^2}{\epsilon_{si} q N_D} \left[\frac{e^{2U_s} - 2U_s e^{U_s} - 1}{(e^{U_s} - 1)^3} \right] \quad (4.78)$$

or near $V = V_{FB}$ ($U_s \rightarrow 0$)

$$\alpha = \frac{C_{ox}^2}{3q \epsilon_{si} N_D} \quad (4.79)$$

As is apparent from the above, higher doping levels should result in a nearly linear improvement in voltage coefficient. As the impurity concentration approaches very high levels ($N_D \geq 2 \times 10^{19}$); however, the assumption regarding non-degeneracy is no longer valid. In this region a more involved capacitance model must be used, which assumed only partial donor ionization and which includes impurity band broadening effects (IB). The predicted results of both of the above models, and a third which also assumed Fermi-Dirac statistics but with complete ionization (FI) are shown with experimental data in Fig. 4.18 [79].

Note that although the above analysis using Maxwell-Boltzman statistics is rather simplified, it still provides reasonable agreement with the experimental data in even degenerately doped cases. Of perhaps greatest interest in this figure, is the very low voltage coefficient of MOS capacitors with even a modestly doped semiconductor plate. These voltage coefficient figures compare favorably with those of thin-film resistors or diffused resistors which are used with a bias compensation technique. These figures are even more impressive when coupled with temperature characteristics which are described in the next section.

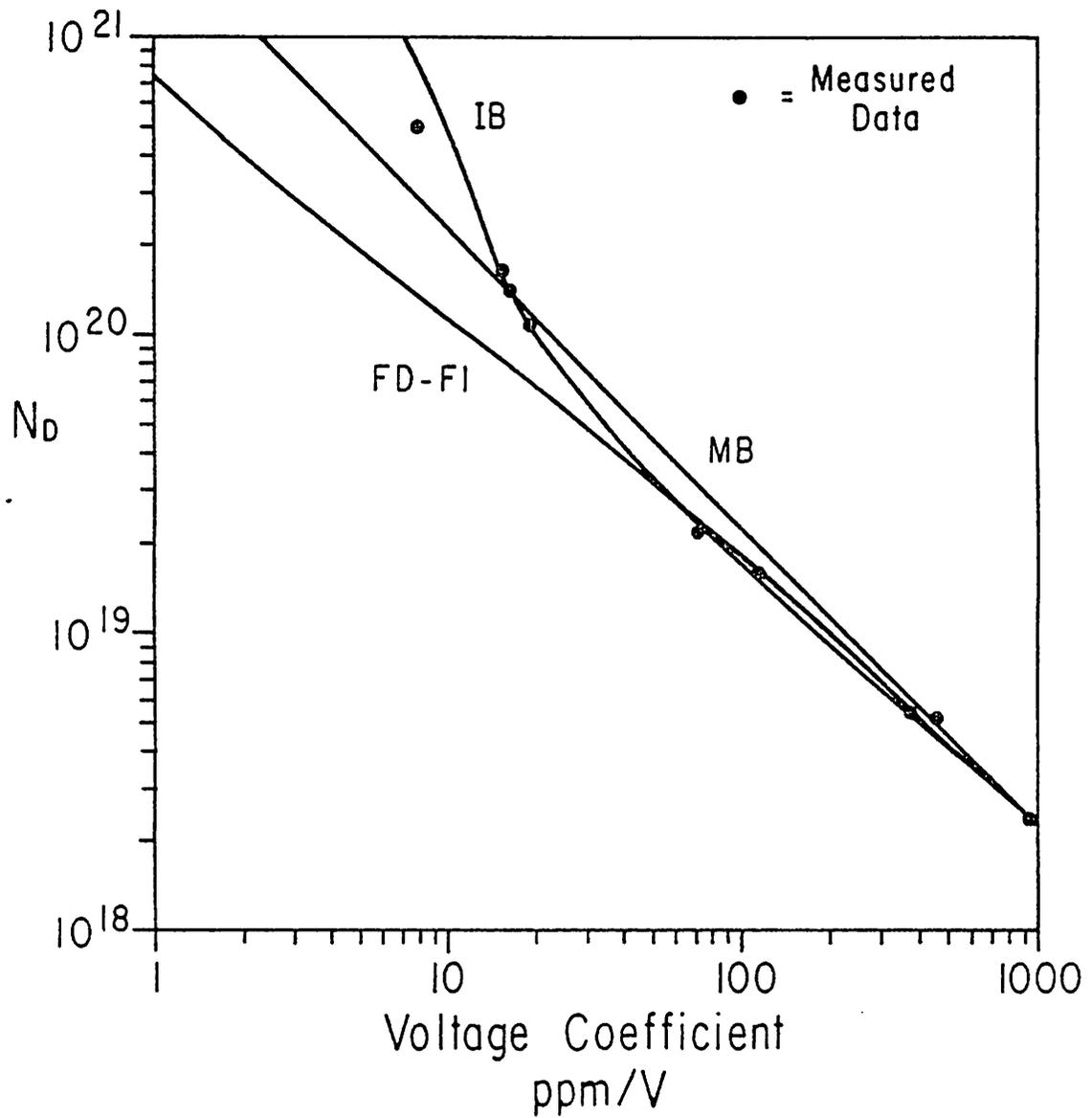


Fig. 4.18 Measured and predicted voltage coefficient of capacitance for different bottom plate impurity concentrations [79]. See text for label definitions.

4.9 Temperature Coefficient of MOS Capacitors

The temperature dependence of an MOS capacitor may be determined by using similar techniques to those used in the previous section.

The normalized temperature coefficient of capacitance may be given by

$$\beta = \frac{1}{AC_T} \frac{\partial AC_T}{\partial T} \quad (4.80)$$

in which C_T is the total capacitance per unit area, and where A is the total capacitor area. The expression for the temperature coefficient may be expanded to give

$$\begin{aligned} \beta &= \frac{1}{A} \frac{\partial A}{\partial T} + \frac{1}{C} \frac{\partial C}{\partial T} \\ &= \left[\frac{1}{A} \frac{\partial A}{\partial T} - \frac{1}{t_{ox}} \frac{\partial t_{ox}}{\partial T} \right] + \frac{C_{ox}}{C_s} \frac{C_s}{\partial T} + \frac{1}{\epsilon_{ox}} \frac{\partial \epsilon_{ox}}{\partial T} \end{aligned} \quad (4.81)$$

$$= \beta_{th} + \beta_{sc} + \beta_{\epsilon_{ox}} \quad (4.82)$$

where the first term corresponds to physical modification of the relevant capacitor dimensions with temperature, the second indicates modification of the effective silicon surface capacitance, while the last indicates temperature dependence of the oxide dielectric constant. McCreary has shown that the first of these terms for a circular capacitor may be expressed by [79]

$$\beta_{th} = \frac{2\alpha_{si} - (1 + \nu_{ox})\alpha_{ox}}{(1 - \nu_{ox})} \quad (4.83)$$

where α_{si} and α_{ox} are the linear thermal expansion coefficients of silicon dioxide and ν_{ox} is the Poisson's ratio for SiO_2 .

The temperature dependence of the silicon surface capacitance may be obtained by differentiating Eq. (4.75). This differentiation is made considerably easier by first expanding $e^{\frac{U}{\psi_s}}$ about zero and ignoring all but first order terms. This approximation is valid about the flatband voltage, and results in a space-charge temperature dependence of

$$\beta_{sc} \approx \frac{C_{ox} k}{2q(\epsilon_{si} N_D kT)^{1/2}} \left[1 - T \frac{1}{\epsilon_{si}} \frac{\partial \epsilon_{si}}{\partial T} \right] \quad (4.84)$$

The value of $\beta_{\epsilon_{ox}}$ may be determined from previously reported data, which has been given as between + 15 ppm/°C [80] and + 21 ppm/°C [79]. Assuming values of 2.8 ppm/°C and .5 ppm/°C for the linear expansion coefficients of Si and SiO₂ [81], a value of .163 for ν_{ox} [82], and a value of + 250 ppm/°C for $\frac{1}{\epsilon_{si}} \frac{\partial \epsilon_{si}}{\partial T}$, this indicates a β_{th} of + 6.0 ppm/°C and a β_{sc} of - 1.5 ppm/°C in a capacitor with 1000 Å of oxide and a surface doping of 1.6×10^{20} [79]. This then indicates a total β of between 20 and 26 ppm/°C. This approximate value is verified by the experimental results shown in Fig. 4.19 [79]. Also included on this plot are the predicted results from a more complete model which includes degeneracy effects [83].

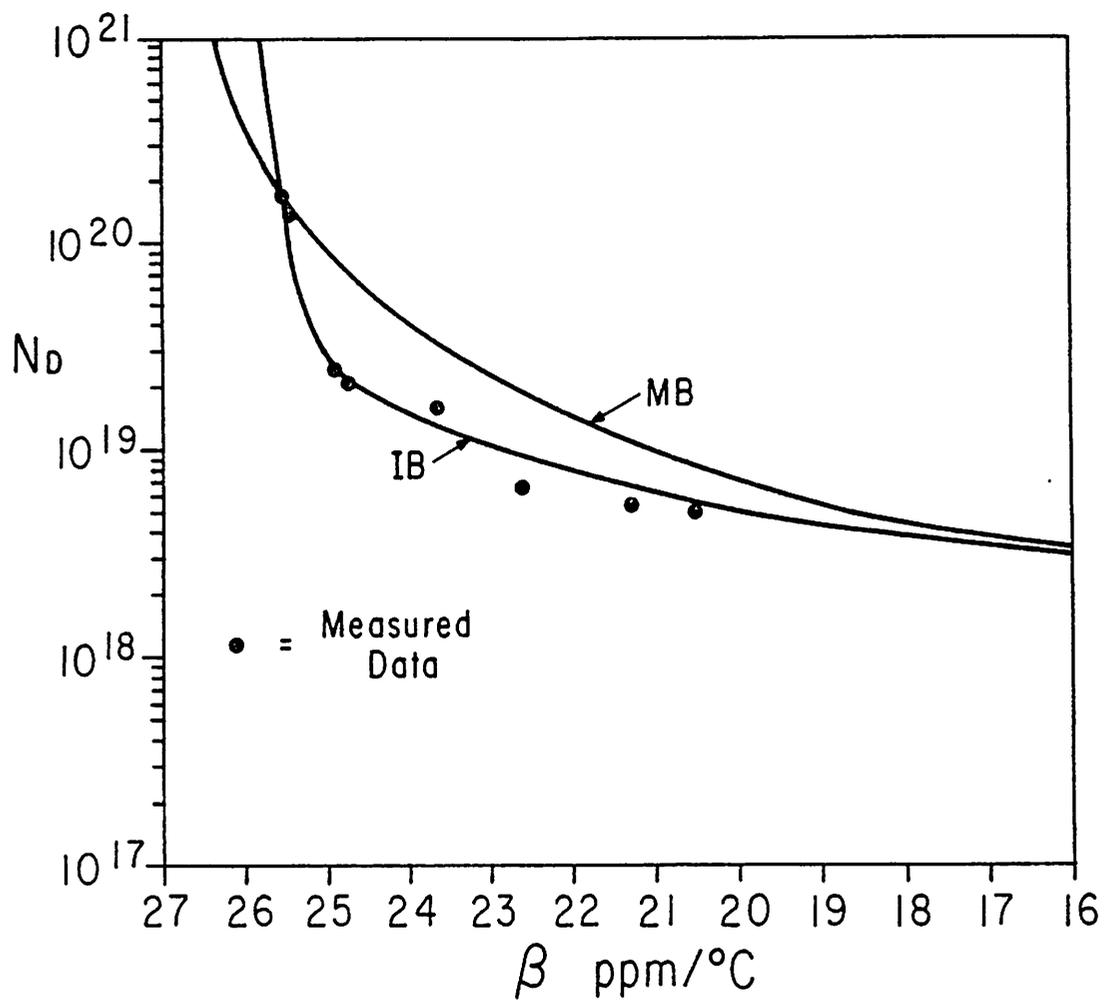


Fig. 4.19 Measured and predicted temperature coefficient of capacitance for different bottom plate impurity concentrations [79].

CHAPTER 5
ANALOG CIRCUIT BUILDING BLOCKS

In this chapter, a family of general purpose analog building blocks is described. These circuits, using the precision elements of Chapter 3 as ratio standards, allow any of the previously discussed converters to be realized. Sections 5.1-5.3 examine source followers, and single and two stage transconductance amplifiers, while Section 5.4 discusses various design considerations in integrators. Section 5.5 covers output stages, while comparators are examined in Section 5.6. Noise in analog circuits is discussed in Section 5.7.

For the designs in this chapter, a simple MOS model has been adopted for ease of calculation [84]. More rigorous modeling information should be included in computer circuit simulations for optimization of results (for example [85,86]). It is assumed here, that for an NMOS device,

$$I_D = \frac{\beta(V_{GS} - V_T)^2}{2(1 - \lambda V_{DS})}, \quad V_{DS} > V_{GS} - V_T \quad (5.1)$$

and

$$I_D = \frac{\beta[(V_{GS} - V_T)V_{DS} - V_{DS}^2/2]}{(1 - \lambda V_{DS})}, \quad V_{DS} < V_{GS} - V_T \quad (5.2)$$

where $\beta = \mu C_{ox} W/L$, λ is the channel length modulation parameter and V_{gs} , V_{ds} and V_T indicate the gate to source, drain to source and threshold voltages respectively. For a PMOS device the same equations apply, but where now λ is negative, and a positive current indicates current flow from (rather than to) the MOSFET.

5.1 Source Followers

Source followers are commonly employed as a level shifter between high gain stages, or as a buffer between a high impedance gain stage and a low impedance load. With slight modifications to reduce distortion, they are also suitable for use as moderate precision voltage followers, which are useful in pipeline converters and input sample and holds.

5.1.1 Frequency Response and Gain

In its simplest configuration, the source follower takes the form shown in Fig. 5.1a. For this circuit, the first order small signal gain is given by

$$\frac{V_{out}(s)}{V_{in}(s)} = \frac{(sC_1 + g_{m_1})}{(g_{m_1} + g_{mb_1} + 1/R_L + s(C_1 + C_L))} \quad (5.3)$$

where g_{mb_1} is the back-bias transconductance of M_1 and R_L is the parallel impedance of the nonideal current source and the output load. At low frequencies, this expression reduces

$$\frac{V_{out}}{V_{in}} = \frac{g_{m_1}}{(g_{m_1} + g_{mb_1} + 1/R_L)} \quad (5.4)$$

which for small values of g_{mb_1} and $1/R_L$ can be made very near unity.

As is evident from Eq. (5.3), the source follower exhibits a single-pole single-zero frequency response. By placing the pole and zero at the same frequency, this circuit may be made very broadband. This is seen to occur, where

$$\frac{C_L}{C_1} = \frac{g_{mb} + 1/R_L}{g_{m_1}} \quad (5.5)$$

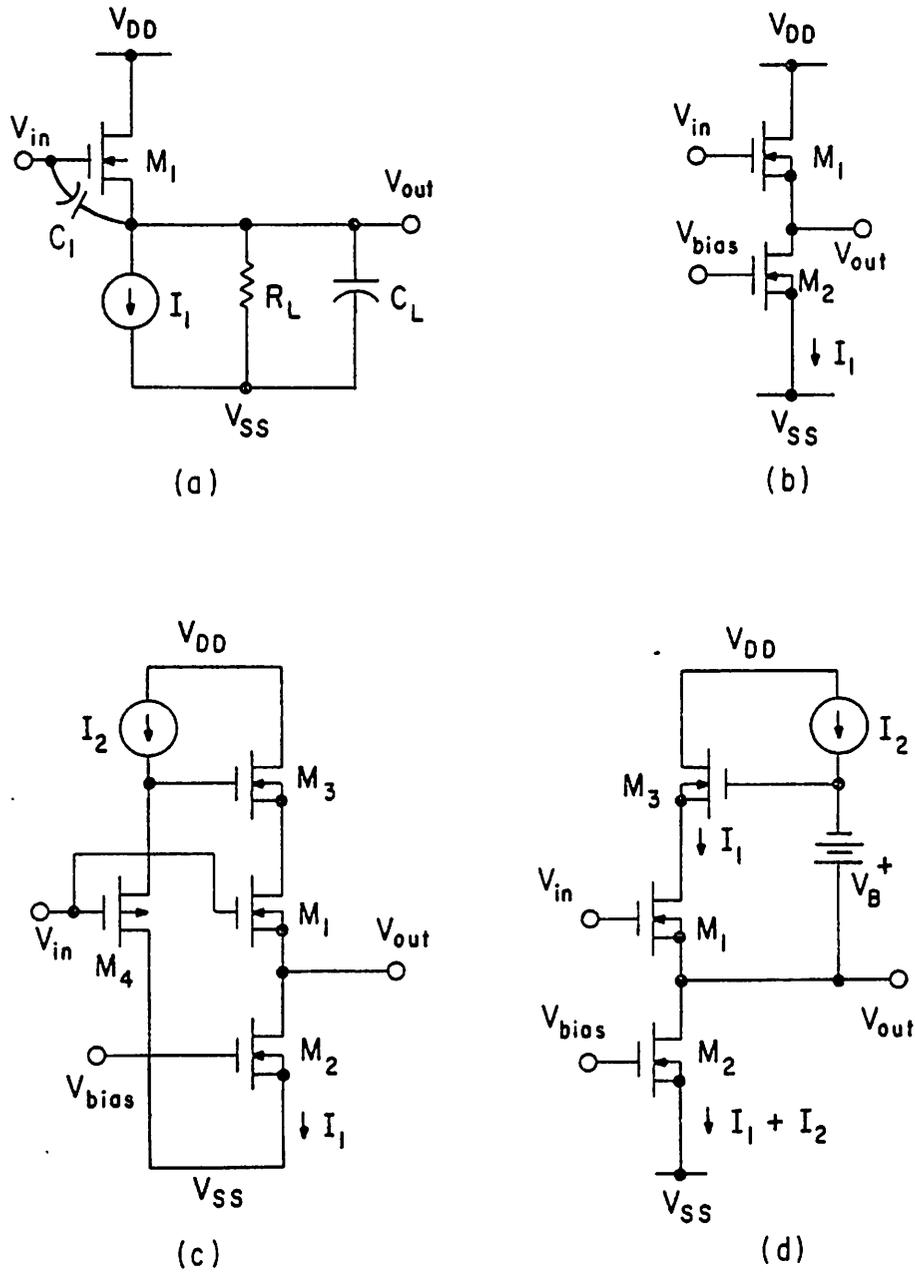


Fig. 5.1 Source follower configurations in CMOS. (a) Basic circuit. (b) Input device in an isolated well. (c) Input derived variable cascode for reduced distortion. (d) Output derived variable cascode.

For source followers which must exhibit a gain approaching unity, this condition can rarely be satisfied. Broadband circuits which drive low impedance loads (on the order of $1/g_{m_1}$) are easily designed, however, for operation over at least a small dynamic range.

5.1.2 Transfer Characteristic Nonlinearity

A common problem with the source follower circuit as previously shown, is nonlinearity of the transfer function over large output swings. This problem occurs because of nonzero drain output conductances in each of the MOS devices and backbias modulation of the threshold voltage of M_1 . The output conductance of the current source may be made high by employing a long geometry device and/or by using series feedback. These techniques are not generally applicable to M_1 , however, as they will directly reduce circuit bandwidth. The nonlinearity produced by the device output conductances and the threshold modulation of M_1 in a typical circuit, is demonstrated in Fig. 5.2 (curve a). This nonlinearity may be reduced by placing M_1 in an isolated well which is tied to its source, as shown in Fig. 5.1b. This procedure, which is only possible in a CMOS or epitaxial process, eliminates first order threshold modulation effects in M_1 and reduces circuit nonlinearity to that shown in Fig. 5.2 curve b. As is evident from the slope of this curve, the drain conductance of M_1 contributes heavily to the overall circuit nonlinearity. Required is a circuit technique which will reduce this dependence, as M_1 should ideally have a short channel length in order to provide good bandwidth. A circuit which at least reduces this problem is shown in Fig. 5.1c. Here, a cascode device (M_3) has been added above M_1 to maintain it at an approximately constant V_{ds} value. This is accomplished by biasing M_3 at a voltage which itself tracks the input voltage (via the source follower

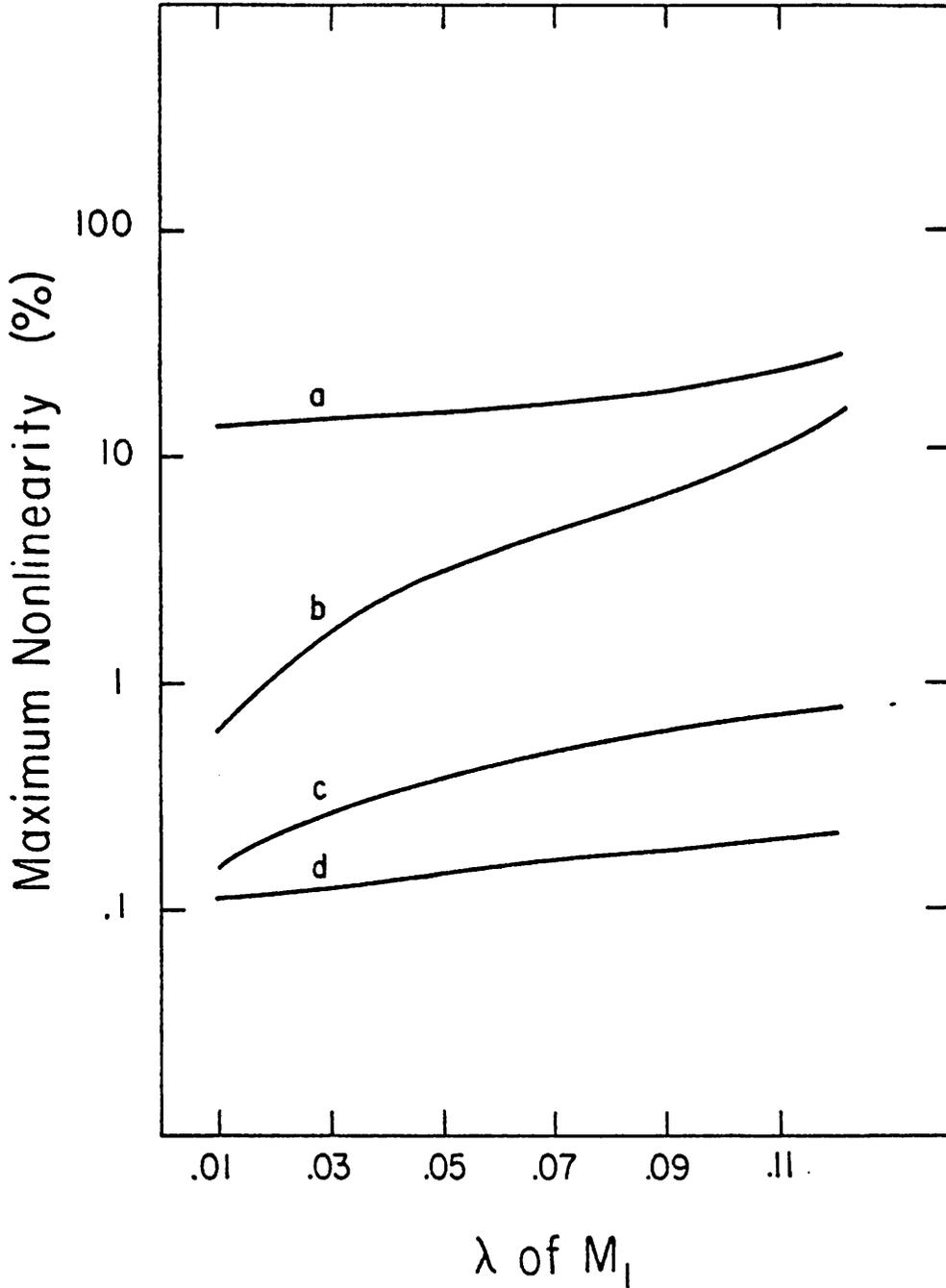


Fig. 5.2 Maximum nonlinearity for each of the source followers of Fig. 5.1. Assumes V_{out} may vary between $\pm 3V$, $V_{DD} = 5V$, $V_{SS} = -5V$, $\sqrt{2\epsilon q N_A}/C_{ox} = .7$ (for NMOS), $\sqrt{2\epsilon q N_D}/C_{ox} = .35$ (for PMOS), $\lambda_2 = .002$, $\lambda_3 = .04$, $\lambda_4 = -.04$, $(V_{gs} - V_T)_{ALL} \approx 1V$.

M_4). As demonstrated by curve c in Fig. 5.2, this technique greatly reduces the contributions of M_1 to the overall nonlinearity, although back-bias effects in M_4 now become the limiting factor. A further modification of the circuit eliminates this problem. By biasing the cascode circuit from the output node, rather than from the input, the use of the PMOS source follower may be avoided.

This circuit, shown in Fig. 5.2d provides the most linear performance of any of the circuits considered, as demonstrated by curve d of Fig. 5.2. In this configuration, the voltage source V_B may be any output independent impedance or voltage drop, such as a pair of diode connected NMOS devices placed in M_1 's well. A practical problem with this circuit, lies in the realization of the current source I_2 . As this source is in parallel with the output node, it may contribute a gain roll-off and nonlinearity if its output impedance is not kept sufficiently high. This almost necessitates the use of a series feedback circuit, which regrettably, will further limit the allowable circuit output swing. Thus, low distortion has ultimately been achieved at the expense of dynamic range.

5.1.3 Elimination of Intrinsic Offset

The usefulness of a source follower circuit is limited in many applications by the intrinsic voltage difference (offset voltage) between input and output. Although this is not normally important in AC coupled circuits, it is a severe problem in sample and hold and voltage follower circuits of the type discussed in Chapter 3. In these circuits an intrinsic offset may directly contribute to the creation of an effective converter offset, or worse yet, a nonlinearity (in pipeline converters).

Two circuit configurations which may be used to reduce this problem are shown in Fig. 5.3. Both of these circuits employ depletion input

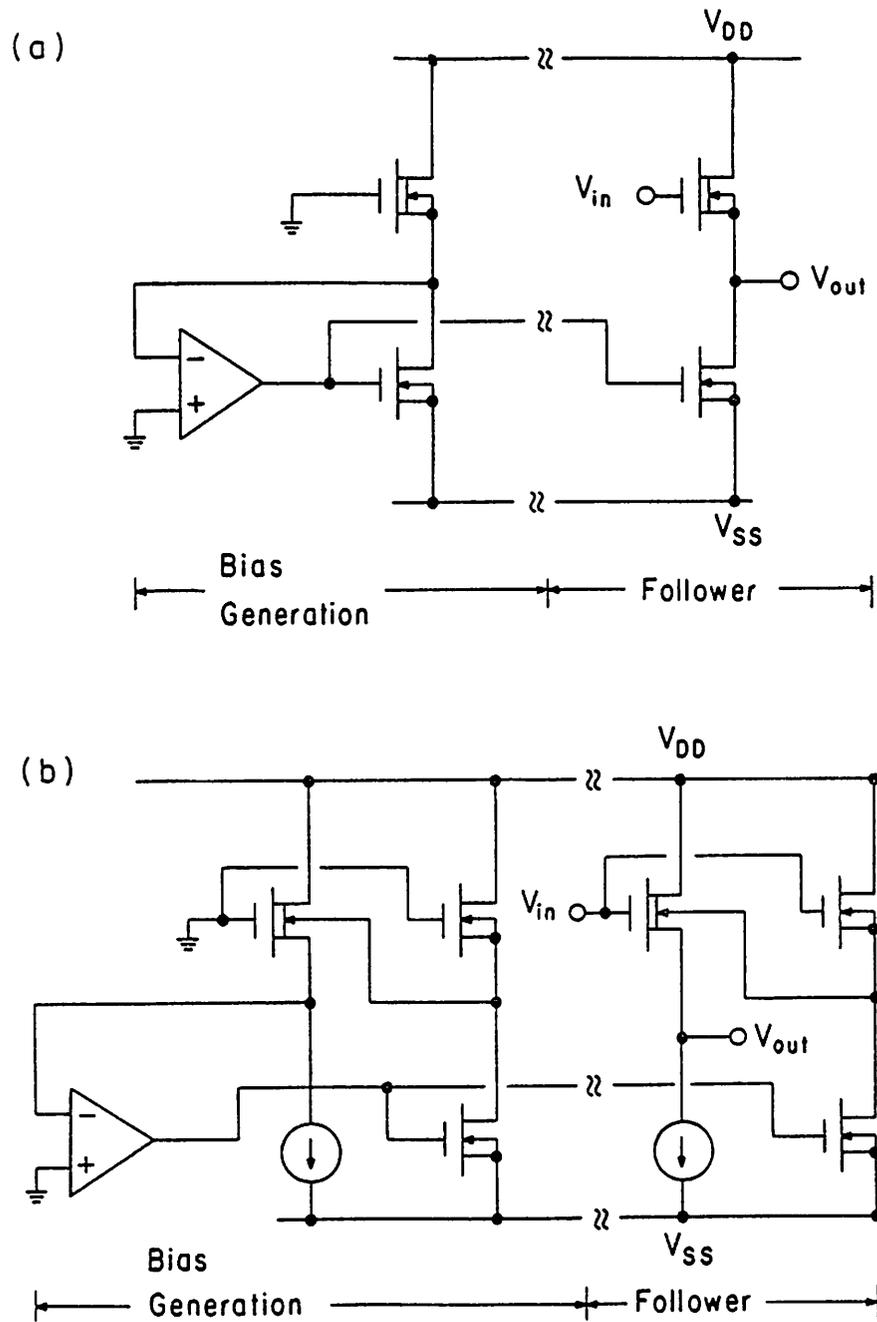


Fig. 5.3 Use of depletion source followers with variable bias conditions allows elimination of intrinsic offset. (a) Variable current. (b) Variable backbias.

follower devices which are biased to a V_{gs} of approximately zero volts. The first circuit, that shown in Fig. 5.3a, employs an adjustable follower bias current, while the second, in Fig. 5.3b, uses an adjustable input device back bias technique. The bias generators for each of these circuits monitor the behavior of a "dummy follower" enabling good offset cancellation over large variations in process parameters and operating conditions. Further, one bias generator may be shared over several followers, making this technique fairly area efficient. Note that care must be exercised in the choice of linear amplifiers for these circuits to ensure that the bias generators are stable over all operating conditions. For optimal performance, this technique may be combined with the low distortion methods discussed above.

5.2 Single Gain Stage Linear Amplifiers

In many circuit applications, the source followers described in the previous section may not be used, as the application may require a gain above unity, very low distortion or differential operation. In these cases some form of linear amplifier must be used. In many instances the amplifier may only have to drive capacitive loads, while in others (usually those driving off the chip), a low impedance output is required. In this section single-high gain stage transconductance amplifiers are described, which are only suitable for use with high impedance loads. Subsequent sections will discuss two-stage transconductance amplifiers and output buffering techniques.

5.2.1 Class-A Amplifiers

A simple common source CMOS Class-A amplifier circuit, is shown in Fig. 5.4. In this figure, r_0 represents the total effective output loading

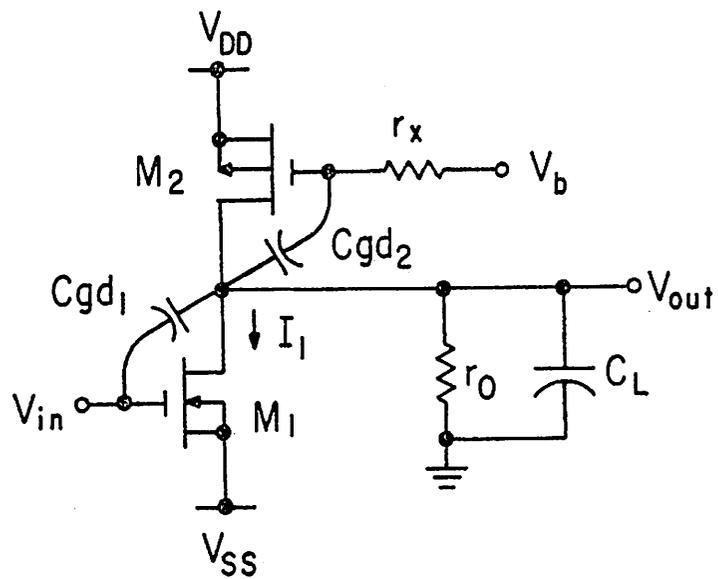


Fig. 5.4 CMOS Class-A gain stage.

from the parallel output impedances of M_1 and M_2 and the output load (if any), while C_L represents the total output node capacitance. The impedance of the bias generator for current source M_2 , is represented by r_x . It is straightforward to show for this circuit, that the small signal transfer function may be given by

$$\frac{V_{out}(s)}{V_{in}(s)} = \frac{-g_{m1} r_0 (1 - sC_{gd1}/g_{m1})(1 + sr_x C_{gd2})}{1 + Bs + As^2} \quad (5.6)$$

where

$$A = (C_{gd1} + C_L) C_{gd2} r_0 r_x$$

and

$$B = (C_{gd1} + C_L) r_0 + C_{gd2} (r_x + r_0 - r_0 r_x g_{m2}) \quad (5.7)$$

The poles and zeros of Eq. (5.6) may be given by

$$z_1 = +g_{m1}/C_{gd1}$$

$$z_2 = -1/(r_x C_{gd2})$$

$$p_1 \approx -1/B$$

and

$$p_2 \approx -B/A \quad (5.8)$$

while the gain at low frequencies is simply,

$$A_V = \frac{V_{out}}{V_{in}} = -g_{m1} r_0 \quad (5.9)$$

By way of the previously presented MOS models, the values for r_0 , g_{m1} and g_{m2} may be determined from the device sizes and bias conditions. Specifically, as

$$I_1 \approx \frac{B_1}{2(1 - \lambda_1 V_{DS1})} (V_{in} - V_{T1})^2 \quad (5.10)$$

then

$$r_{O1} \approx \left[\frac{\partial I_1}{\partial V_{DS1}} \right]^{-1} \approx \frac{(1 - \lambda_1 V_{DS1})}{I_1 \lambda_1} \quad (5.11)$$

Similarly,

$$r_{O2} \approx \frac{(1 - \lambda_2 V_{DS2})}{I_1 \lambda_2} \quad (5.12)$$

where $r_0 = r_{O1} \parallel r_{O2}$. The device transconductance values may be given by

$$g_{m1} \approx \frac{\partial I_1}{\partial V_{in}} \approx \sqrt{\frac{2B_1 I_1}{(1 - \lambda_1 V_{DS1})}} \quad (5.13)$$

and

$$g_{m2} \approx \frac{\partial I_1}{\partial V_{bias}} \approx \sqrt{\frac{2B_2 I_2}{(1 - \lambda_2 V_{DS2})}} \quad (5.14)$$

Note that the transconductance is proportional to $I_1^{1/2}$ (in the first order model) while r_0 is proportional to I^{-1} . This implies that the low frequency gain goes as roughly $I_1^{-1/2}$. Hence, the gain is usually highest when operating at the lowest possible current levels.¹ As the bandwidth of this circuit is also dependent on the current levels, however, adjustment of the current to achieve a required gain is often not possible. In fact, it is often the case that this simple circuit will display insufficient gain for many applications. In these situations

¹At low current densities and high drain voltages, channel avalanching may dominate output impedance behavior [87,88]. This may substantially reduce amplifier gain.

the gain may be improved (at the expense of output swing), by employing a cascode circuit above one or both MOS devices, as shown in Fig. 5.5a. This configuration modifies the effective output impedances to

$$r_{O\downarrow} \approx r_{O3} (1 + g_{m3} r_{O1}) \quad (5.15)$$

and

$$r_{O\uparrow} \approx r_{O4} (1 + g_{m4} r_{O2}) \quad (5.16)$$

which raises the stage gain by a factor of $r_{O\downarrow} \parallel r_{O\uparrow} / r_{O\downarrow}$. The frequency response of the cascode is usually improved over that of the simple common source amplifier, in that the Miller capacitance around M_1 (C_{gd1}) now looks into a low impedance node. This greatly increases the frequency of the right half plane zero. The doublet created by overlap capacitance in the current source also creates less problems in the cascode configuration, in that it is normally much more compressed.

The circuit of Fig. 5.5a, is still not useful in many applications, however, in that the circuit displays a large intrinsic offset. This problem may usually be eliminated by incorporating an offset correcting input stage as shown in Fig. 5.5b. By adjusting the reference current, a known offset may be developed across the input device, which may be used to cancel any offset in the gain stage. This single ended circuit is generally sufficient for most capacitively coupled circuits, if an offset cancellation cycle may periodically be used to ensure that the amplifier operates in its linear region. The low frequency gain of this circuit is simply

$$A_V = \frac{V_{out}}{V_{in}} = \frac{-g_{m5} g_{m1}}{(g_{m5} + g_{m6})} (r_{O\uparrow} \parallel r_{O\downarrow}) \quad (5.17)$$

while the bandwidth (if $|p_1 A_V| < |p_2|$) is given by

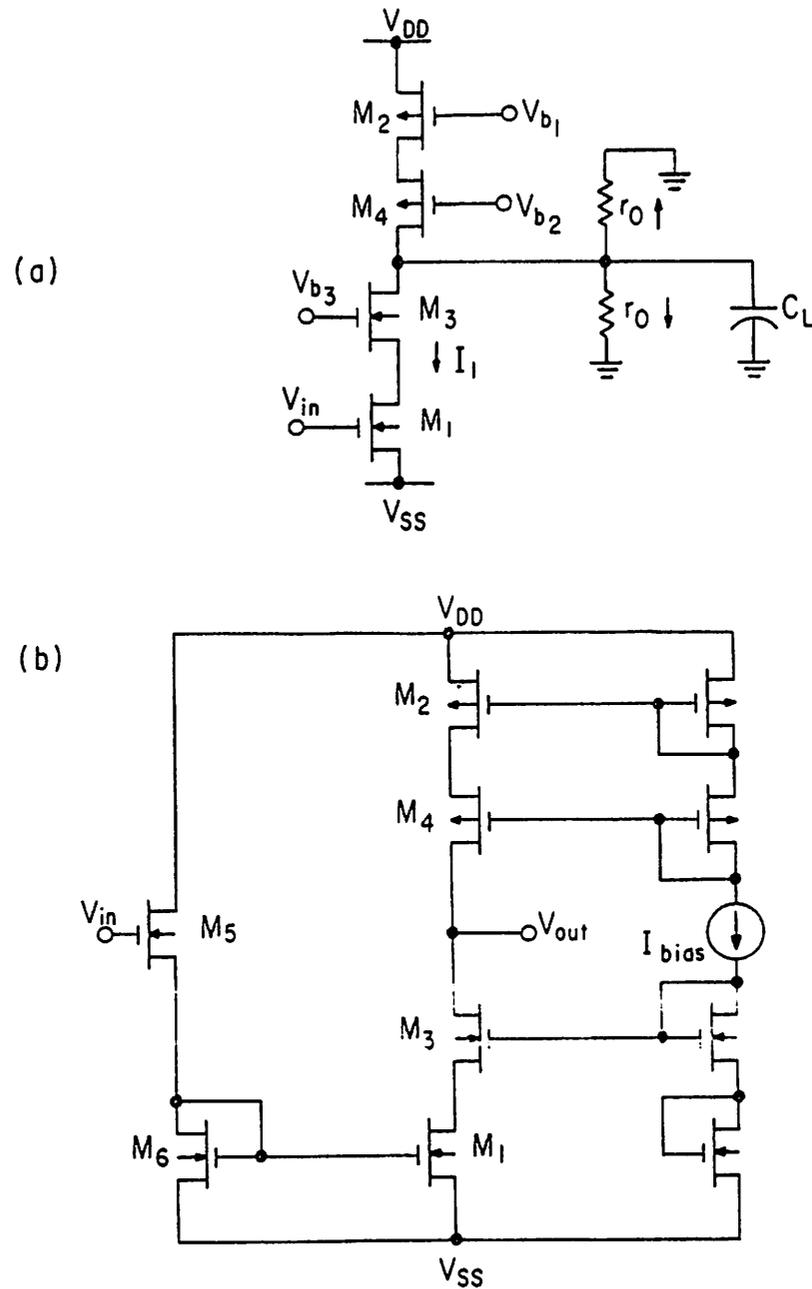


Fig. 5.5 Class-A gain stages using cascode devices. (a) Simple circuit. (b) Circuit with level shift at input for reduction of intrinsic offset.

$$\omega_0 \approx \frac{g_{m5} g_{m7}}{(g_{m5} + g_{m6}) C_L} \quad (5.18)$$

Note that this bandwidth may be reduced if any other poles are near $p_1 |A_V|$. As a rule, the load capacitance C_L should be made sufficiently large that an approximate single-pole response is obtained even in the presence of the nondominant poles.

To allow use of this high-gain circuit in applications requiring differential inputs, as well as in circuits which cannot tolerate use of a variable bias current for coarse offset cancellation, a differential input stage may be used as shown in Fig. 5.6. Note that in this circuit, a Wilson current mirror has been substituted for the cascode/current-mirror circuit used previously.

5.2.2 Class-B Operation

In high-speed amplifiers, and in amplifiers which must use power sparingly, a Class-B design is often favored over a Class-A approach. In circuits where the useful bandwidth is limited by the maximum slew rate, and in applications where the output current requirements vary widely, a Class-B circuit may provide substantially improved performance. An example of a Class-B single-ended amplifier is shown in Fig. 5.7. In this circuit the bias sources V_B^+ and V_B^- are included to allow adjustment for zero intrinsic offset, as no constant current sources are now available. The modifications shown in this circuit result in an effective amplifier transconductance, of

$$G_m = \frac{g_{m5} g_{m6}}{g_{m5} + g_{m6}} \cdot \frac{I_1}{I_2} + \frac{g_{m7} g_{m8}}{g_{m7} + g_{m8}} \cdot \frac{I_1}{I_2} \quad (5.19)$$

and a bandwidth (assuming a single pole response) of simply

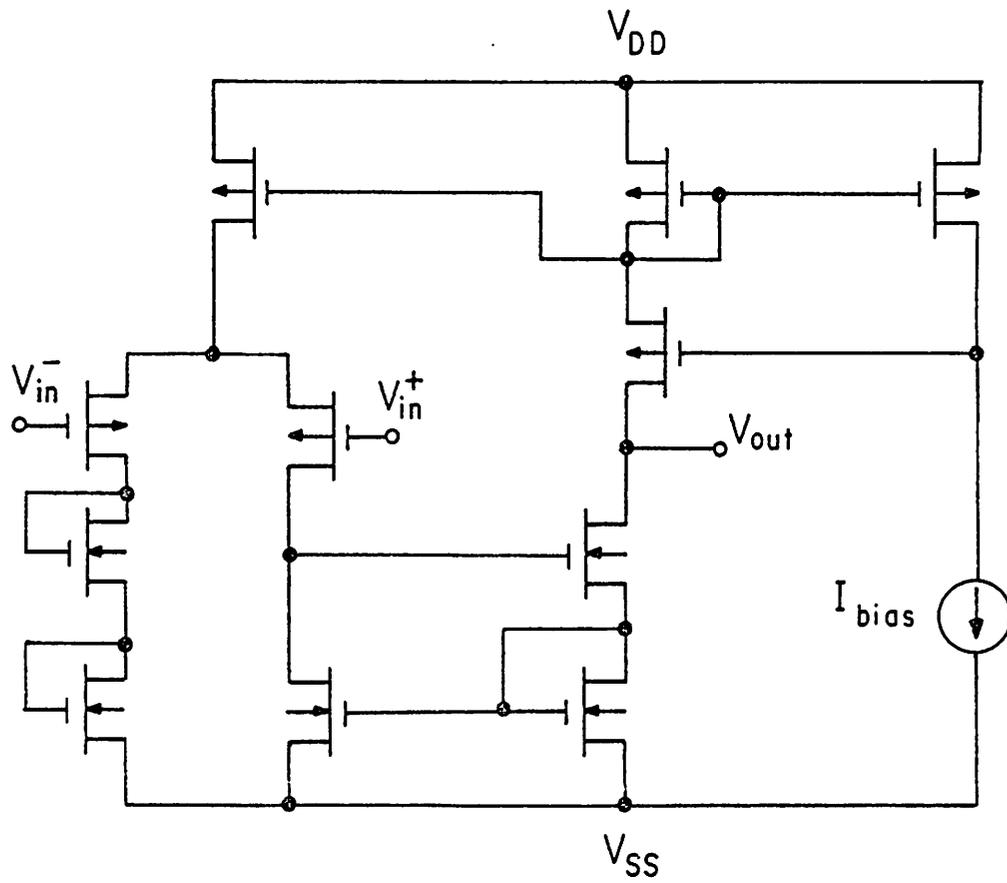


Fig. 5.6 Differential input high gain amplifier.

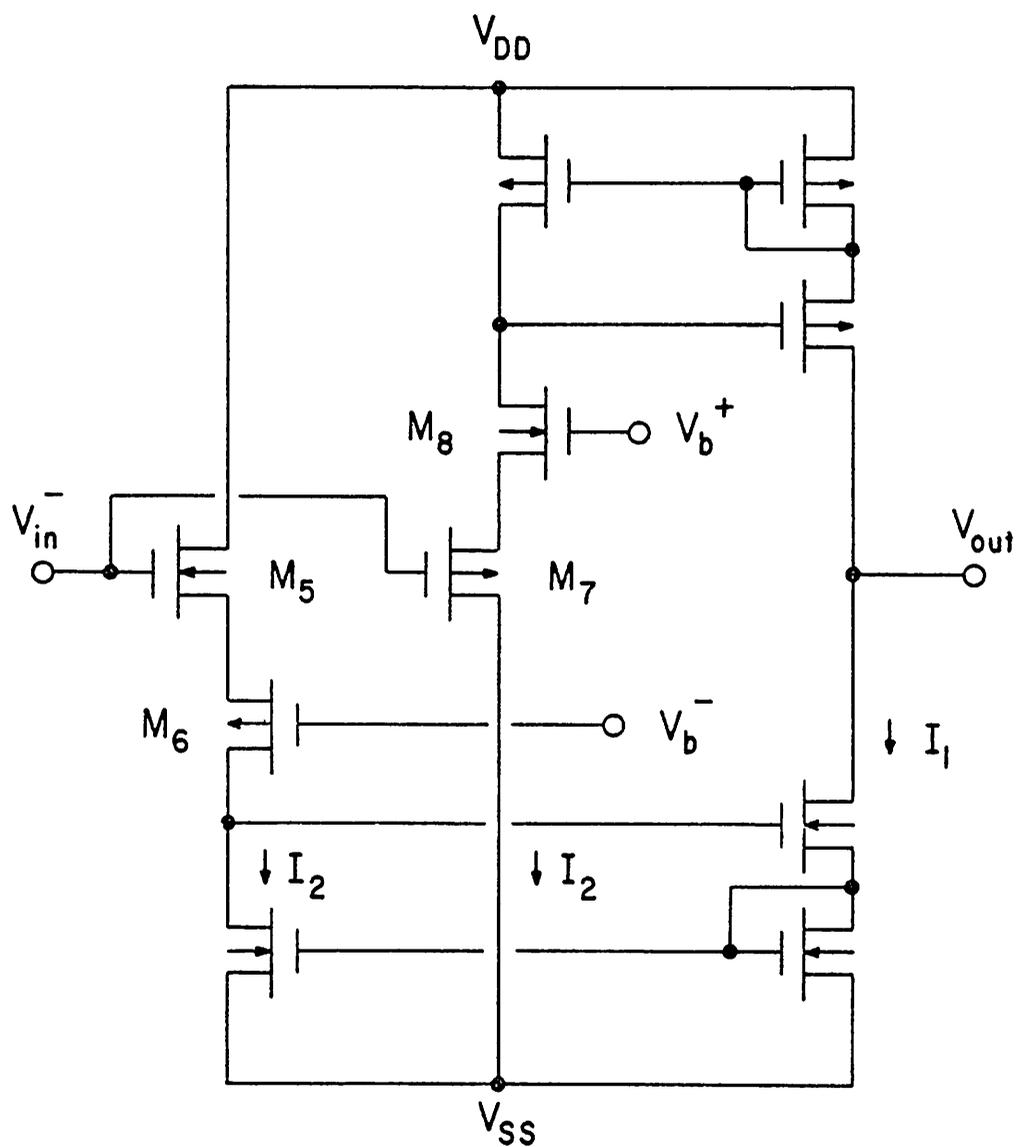


Fig. 5.7 Class-B single-ended amplifier.

$$\omega_0 = G_m / C_L \quad (5.20)$$

This circuit may be made with differential inputs by modifying the circuit as shown in Fig. 5.8. While this circuit allows full Class-B operation with differential inputs, it will in some applications present difficulties due to the fairly small common-mode input range. This problem may be reduced by employing the circuit configuration shown in Fig. 5.9. In this circuit a Class-A differential stage is used to provide an increase in common-mode range, while a level shifter is used to maintain push-pull operation at the output. As opposed to the previously described fully Class-B circuit, the circuit in Fig. 5.9, will still be subject to slew rate limiting, but at twice the value of a Class-A circuit with the same gain-stage current.

5.2.3 Methods for Increasing Bandwidth

For a specified load capacitance, circuits similar to those just described will exhibit an improved bandwidth if the overlap and gate capacitances are minimized, and the circuit is operated at relatively high current levels. Reducing the thickness of the gate oxide will help to a point, although at a constant current level this will actually reduce the frequency of some nondominant poles. An alternative broadbanding technique, which is only possible in some CMOS processes, is to use high transconductance bipolar devices at critical nodes. This technique is illustrated in Fig. 5.10. In this circuit, triple-diffused isolated NPN transistors are used to improve the frequency response of the current mirror on the negative supply, as well as to increase the frequency of the current source doublet. This technique is perhaps most useful in moderate or large linewidth processes ($L_{\min} > 6\mu$) where the useful MOSFET

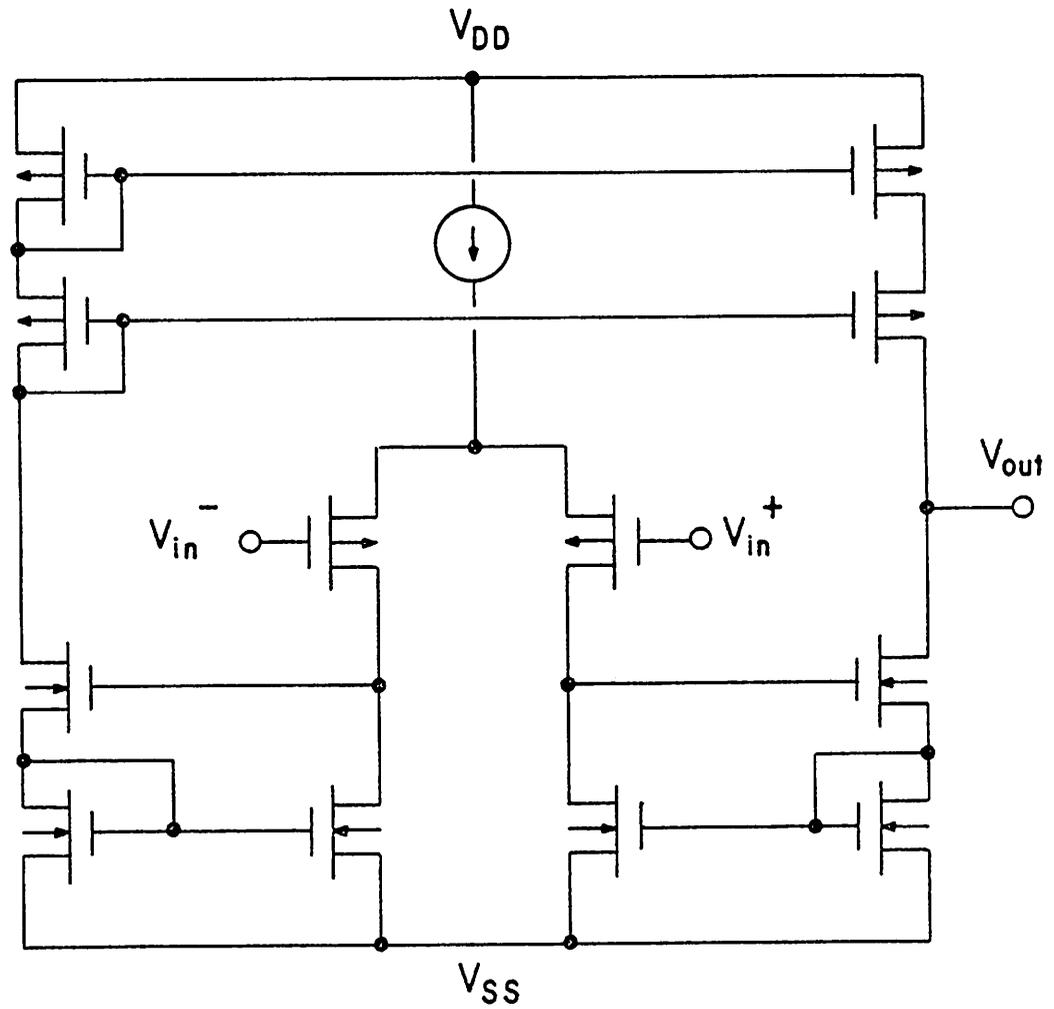


Fig. 5.9 Differential amplifier with improved common-mode range but reduced slew rate.

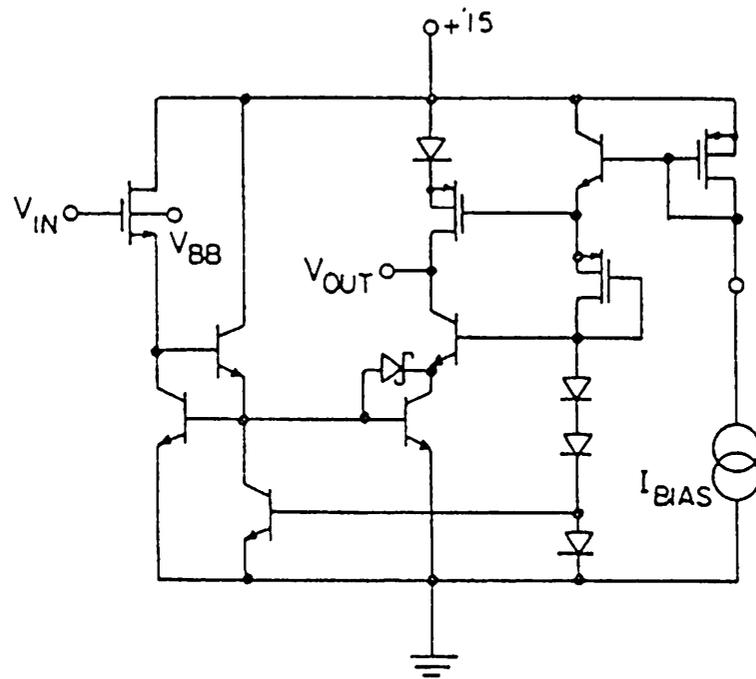


Fig. 5.10 Use of isolated bipolar devices at critical nodes allows for improved bandwidth. This technique is only possible in specially tailored CMOS processes.

frequency range is low relative to that of a bipolar transistor. In advanced processes, however, the utility of this technique is reduced as the collector resistance of the triple-diffused bipolar devices becomes a speed-limiting factor. This amplifier and the CMOS/bipolar process discussed above, have been described elsewhere by the author [89,90] and will not be expanded upon here.

5.3 Two Gain-Stage Linear Amplifiers

In the single gain-stage circuits just presented, it is usually necessary to employ cascode devices (or the equivalent) in order to achieve high gain. Although these additional devices do improve the stage gain, they also reduce circuit output swing. Further, as will be shown in a subsequent section, the level shifters needed to provide a low offset with a single-stage circuit will greatly increase amplifier equivalent input noise. These problems may often be eliminated by employing an amplifier with two high-gain stages. In these circuits the gain requirements of each stage are greatly reduced, as only the product of the individual gains need exceed the total gain requirement. Further, the equivalent input noise may usually be kept small by employing a small number of devices in a high gain input stage.

An example of a two gain-stage differential input Class-A amplifier is shown in Fig. 5.11. In this circuit, the block labeled "compensation circuitry" is necessary if stable amplifier operation is to be achieved while employing feedback. Several types of compensation circuits are discussed below.

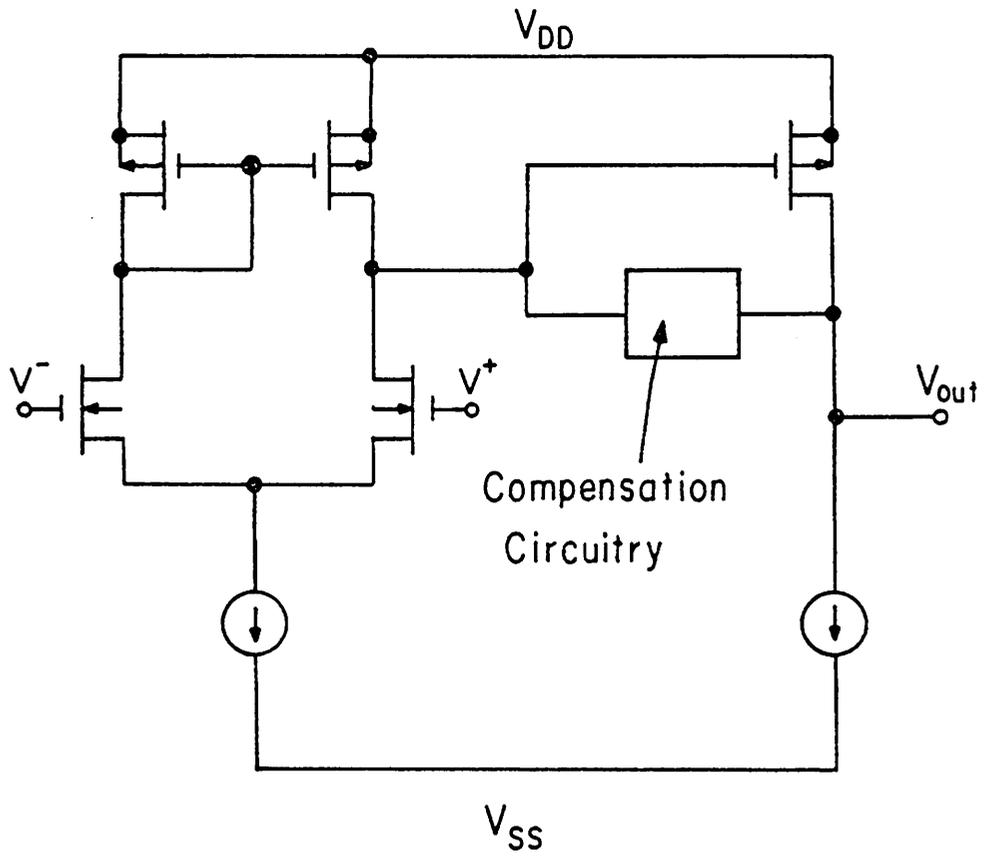


Fig. 5.11 Two gain-stage differential input amplifier.

5.3.1 Compensation Techniques

A. Simple Pole-Splitting

A problem with two-stage amplifiers, which is typically more complicated than with single gain-stage types, is that of frequency compensation. While a number of techniques exist for achieving stable feedback operation, the most useful involve some form of pole-splitting. This method in simple form, is shown in Fig. 5.12a. In this figure, the effective output impedance of each stage is represented by resistors r_1 and r_2 , while the effective capacitive loads are represented by C_1 and C_L . As has been shown for this circuit [91], the small signal transfer function may be given by

$$\frac{v_{out}(s)}{v_{in}(s)} = \frac{-g_{m1} g_{m2} r_1 r_2 (1 - sC_f/g_{m2})}{1 + Bs + As^2} \quad (5.21)$$

where

$$A = r_1 r_2 (C_1 C_L + C_f C_L + C_f C_1)$$

and

$$B = g_{m2} r_1 r_2 C_f + r_1 (C_f + C_L) + r_2 (C_f + C_L) \quad (5.22)$$

the poles and zeros of Eq. (5.21) may be given by

$$\begin{aligned} z_1 &= +g_{m2}/C_f \\ p_1 &\approx -1/(g_{m2} r_1 r_2 C_f) \\ p_2 &\approx -g_{m2} C_f / (C_1 C_L + C_f (C_1 + C_L)) \end{aligned} \quad (5.23)$$

while the gain at low frequencies is simply,

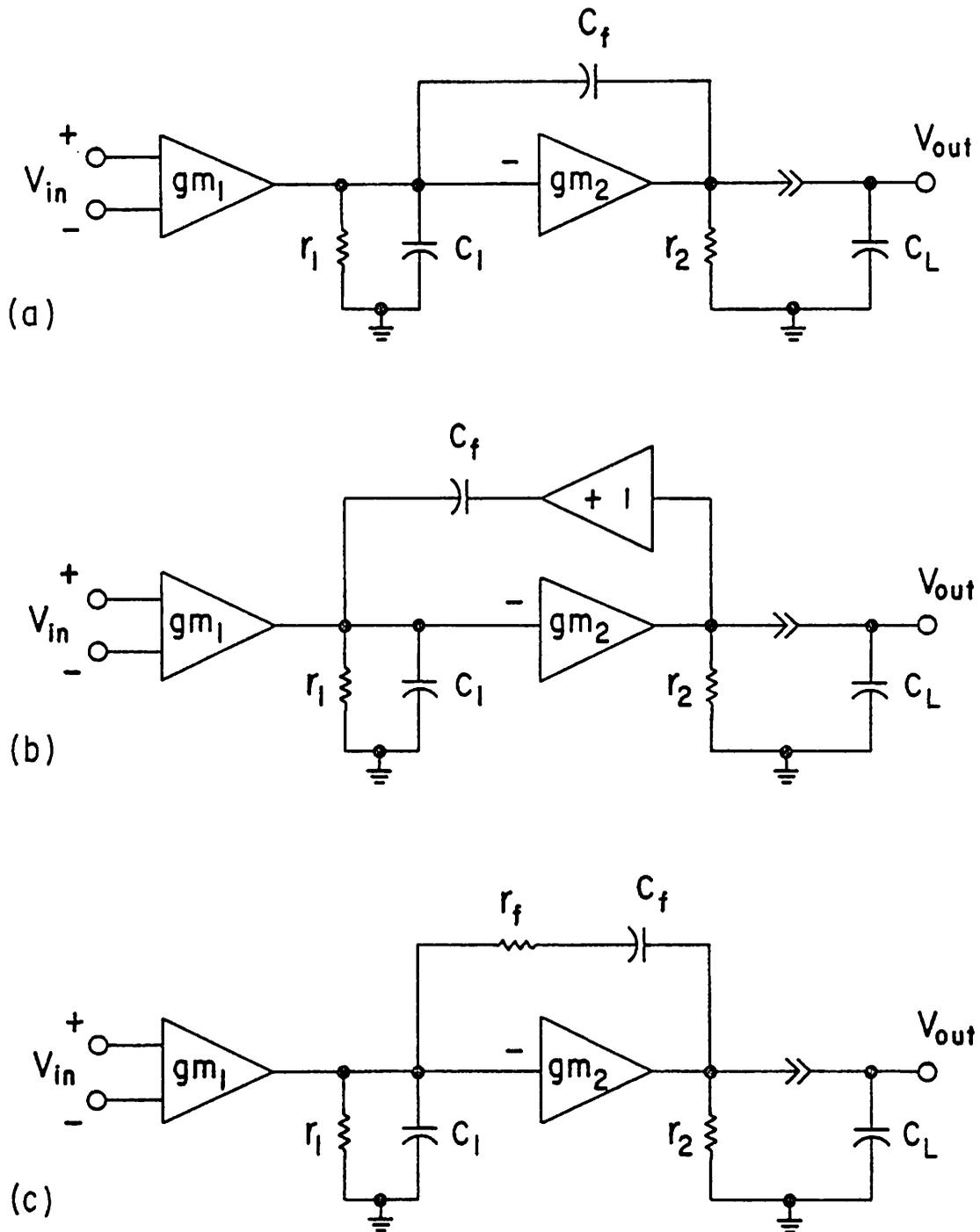


Fig. 5.12 Compensation techniques for two gain-stage amplifiers. (a) Simple pole-splitting. (b) Pole-splitting with feedforward blocking. (c) Pole-splitting with R-C control of zero.

$$A_V = \frac{V_{out}}{V_{in}} = -g_{m_1} g_{m_2} r_1 r_2 \quad (5.24)$$

In implementations of this circuit which employ bipolar transistors, the zero indicated above is usually at a high frequency, leaving a basically two pole response. For a 45° phase margin ($|p_2| \approx |A_V p_2|$), it is straightforward to show, that C_f must be sufficiently large, that

$$C_f \geq \frac{g_{m_1} (C_1 + C_L) + \sqrt{g_{m_1}^2 (C_1 + C_L)^2 + 4g_{m_1} g_{m_2} C_1 C_L}}{2g_{m_2}} \quad (5.25)$$

which for $C_L \gg C_1$, reduces to

$$C_f \geq \frac{g_{m_1} C_L}{g_{m_2}} \quad (5.26)$$

In MOS circuits, however, the zero at g_{m_2}/C_f is rarely above p_2 . In fact, in many circuits it is at a sufficiently low frequency that use of this compensation technique will lead to marginal closed loop stability. In these situations, modifications of this basic scheme must be considered. Several of these are described below.

B. Pole-Splitting with Feedforward Blocking

A commonly used technique for eliminating the right-half plane zero of Eq. (5.21), is to prevent feedforward through the compensation capacitor, as shown in Fig. 5.12b. In this circuit, a voltage follower is used to block signal injection forward, but still allow the pole-splitting feedback. This modifies the small signal transfer function, to [92]

$$\frac{v_{\text{out}}(s)}{v_{\text{in}}(s)} = \frac{-g_{m_1} g_{m_2} r_1 r_2}{1 + Bs + As^2} \quad (5.27)$$

where

$$A = r_1 r_2 (C_1 C_L + C_f C_L)$$

and

$$B = g_{m_2} r_1 r_2 C_f + r_1 (C_f + C_L) + r_2 C_L \quad (5.28)$$

The poles of Eq. (5.27) are given by

$$p_1 \approx -1/(g_{m_2} r_1 r_2 C_f)$$

and

$$p_2 \approx -g_{m_2} C_f / (C_1 C_L + C_f C_L) \quad (5.29)$$

and are quite similar to those presented previously. Most importantly, however, note that the right half plane zero has been eliminated.

This technique is routinely employed with a source follower acting as a buffer circuit. It should be noted that although a source follower will eliminate the right half plane zero, it will contribute a left half plane doublet. Specifically, the pole and zero which normally occur in a source follower circuit will appear as a zero and a pole in the amplifier transfer function. This doublet is potentially useful in obtaining a broadband amplifier, if the zero may be made to cancel the second amplifier pole. In other cases, however, it may lead to a poor settling response as the doublet may be at low frequency and not well compressed.

C. R-C Compensation Method

A third technique for achieving pole-split amplifier compensation, is shown in Fig. 5.12c [93]. It may be shown for this circuit, (given

enough time), that the poles and zeros are given by

$$\begin{aligned} p_1 &\approx -1/(g_{m_2} r_2 r_1 C_f) \\ p_2 &\approx -g_{m_2} C_f / (C_1 C_L + C_f (C_1 + C_L)) \\ p_3 &\approx -1/r_f C_1 \end{aligned}$$

and

$$z_1 = -1/(C_f (r_f - 1/g_{m_2})) \quad (5.30)$$

Note that when $r_f = 1/g_{m_2}$, the zero is at infinity, leaving a basically two pole response (assuming $C_L \gg C_1$). In this case the requirement upon the compensation capacitance is, as given previously,

$$C_f \geq \frac{g_{m_1}}{g_{m_2}} C_L \quad (5.31)$$

Further, the maximum amplifier bandwidth is given by

$$\omega_0 \approx |p_2| \approx g_{m_2} / C_L \quad (5.32)$$

It would be desirable if a technique could be developed which would allow the bandwidth to be increased and/or the compensation capacitance to be made smaller without modifying the basic amplifier and load parameters. This is seen to occur, if the zero is brought in from infinity and placed atop the second pole.² This zero frequency will result when

$$r_f \approx (C_L + C_f) / (g_{m_2} C_f) \quad (5.33)$$

providing again a two pole response, but where now the stability

²Because of the presence of p_3 and other high frequency poles, the optimal settling condition is actually where z_1 is at a slightly lower frequency than p_2 .

requirement is given by

$$|p_3| \geq |A_V p_1| \quad (5.34)$$

or, in terms of the compensation capacitance

$$C_f \geq \sqrt{g_{m1}/g_{m2} \cdot C_1 C_L} \quad (5.35)$$

This is a substantially smaller capacitor than that required previously, allowing a considerable increase in bandwidth. In fact, the maximum amplifier bandwidth (neglecting additional high frequency poles) is now given by

$$\omega_0 = |p_3| = \frac{g_{m2} C_f}{(C_L + C_f) C_1} \quad (5.36)$$

The difficulty with this technique is in maintaining the proper resistance r_f over large process, temperature and supply variations. If this is not done, the $p_2 z_1$ doublet will separate, resulting in an increased settling time. A circuit which alleviates this problem is shown in Fig. 5.13, where a tracking resistance scheme is used. In this circuit, the required series resistance is generated by the drain-to-source impedance of device M_x , whose gate (and hence, resistance) is biased by device M_z , voltage source v_{os2} , and current source I . The second gain stage is represented by driver device M_y and current source $K \cdot I$, while the input voltage of this stage is represented by v_{in2} . The input voltage of the second stage during quiescent periods is represented by V_{os2} , while capacitances C_f and C_L are the compensation and load capacitances as before.

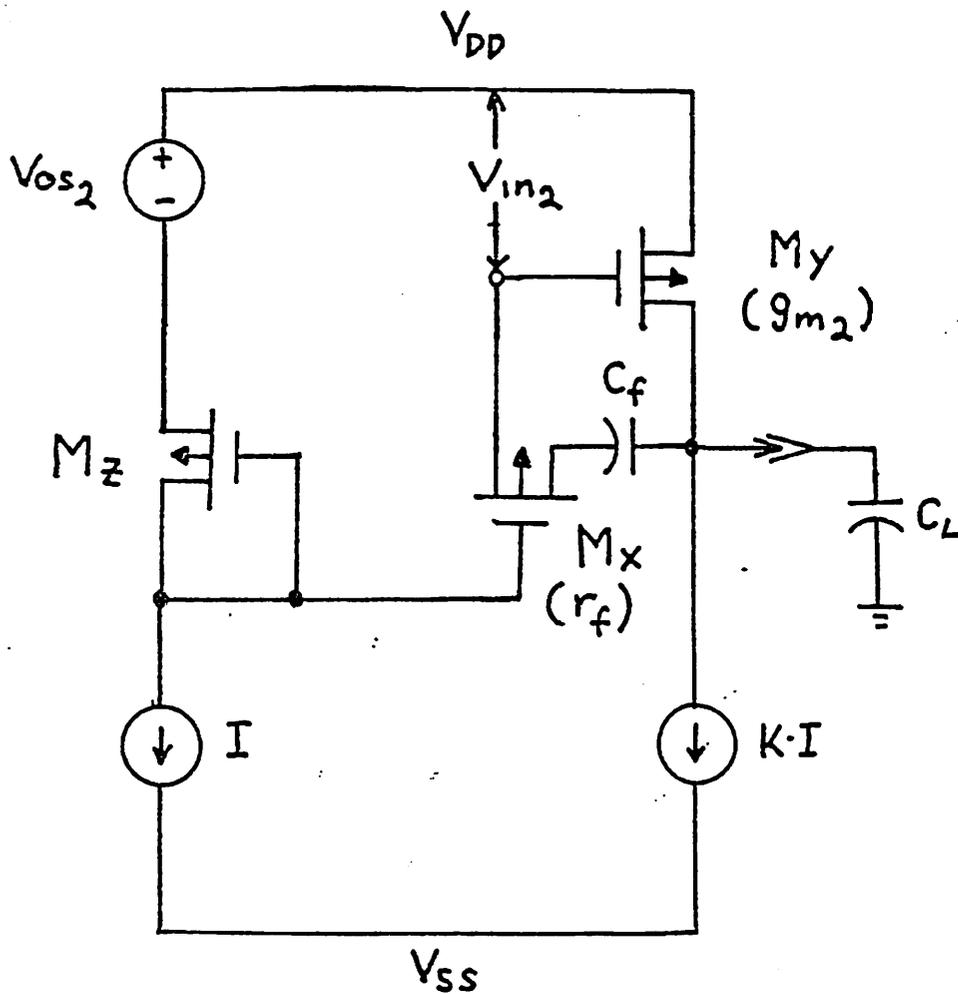


Fig. 5.13 Tracking R-C compensation circuit. R_{ds} value of M_x tracks the transconductance of M_y , enabling optimal compensation to be maintained over large process, temperature and current variations.

Using the MOS models presented previously, it is easy to show that if

$$\left(\frac{W}{L}\right)_x \approx \left[\left(\frac{W}{L}\right)_y \cdot \left(\frac{W}{L}\right)_z \cdot K \right]^{1/2} \cdot \frac{C_f}{(C_f + C_L)} \quad (5.37)$$

then the resistance condition is satisfied, or

$$R_{ds_x} \approx \frac{(C_f + C_L)}{g_{m2} C_f} \approx R_f(\text{nom}) \quad (5.38)$$

Note that Eq. (5.37) is totally independent of process, temperature and supply variations, and is only a function of relative device sizes, which may be well-matched and easily specified. This technique has been employed in the amplifiers in a high order monolithic filter [94], allowing the use of a relatively small compensation capacitor with even a large input stage transconductance and large capacitive load (which were required for reasons of noise and power supply rejection). Similarly, this technique may be employed in applications with small capacitive loads to achieve a very high bandwidth and a fast settling time. This technique is generally most useful in circuits employing poly/poly or metal/poly capacitors, as the optimal resistance value for r_f is dependent upon the output capacitive loading. Implementations in a process employing a diffused bottom plate will typically exhibit a sufficiently large voltage dependence of capacitance, that very good pole-zero cancellation may be difficult to achieve over the entire output voltage range.

5.4 Integrators

Nearly all of the amplifier applications discussed in Chapter 3 use some form of integrator configuration. As such, it is important that

the amplifier requirements for this type of circuit be considered. In this section, amplifier bandwidth and compensation requirements are discussed, with regard to integrator frequency and step response. Design techniques for minimizing power supply coupling are also considered in detail.

5.4.1 Dynamic Characteristics

A simple integrator employing a single-pole transconductance amplifier, is shown in Fig. 5.14. In this figure, r_0 indicates the total resistive output loading, while C_I , C_L and C_p indicate the integrating, output load and input parasitic capacitances respectively. It is straightforward to show for this circuit, that

$$\frac{V_{out}}{i_{in}} = \frac{-g_m r_0 (1 - sC_I/g_m)}{sC_I(g_m r_0 + 1 + C_p/C_I) \left(1 + \frac{s r_0 (C_L (1 + C_p/C_I) + C_p)}{(g_m r_0 + 1 + C_p/C_I)}\right)} \quad (5.39)$$

which at low frequencies reduces to

$$\frac{V_{out}}{i_{in}} \approx \frac{-g_m r_0}{sC_I(g_m r_0 + 1 + C_p/C_I)} \quad (5.40)$$

The poles and zeros of Eq. (5.39) are given by

$$z_1 = +g_m/C_I$$

$$p_1 = 0$$

and

$$p_2 = \frac{-(g_m r_0 + 1 + C_p/C_I)}{r_0 (C_L (1 + C_p/C_I) + C_p)} \quad (5.41)$$

As is evident from these equations, this circuit functions as a lossy integrator at low frequencies, but exhibits non-ideal gain and phase

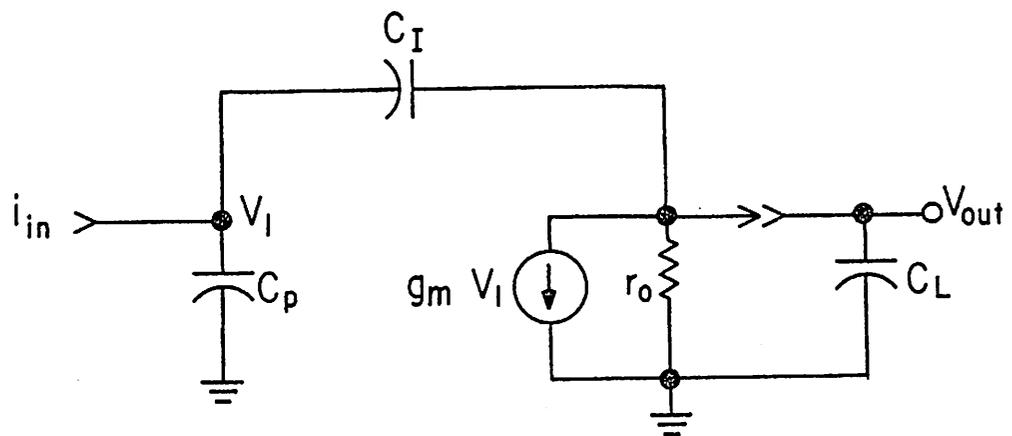


Fig. 5.14 Small-signal model of a simple integrator.

behavior at high frequencies. The right half plane zero occurs because of feedforward through C_I , while P_2 is due to amplifier gain rolloff.

In most of the amplifier applications considered here, the step response of this circuit is of critical importance. In pipeline converters, and in sample and holds, the integrator's settling time will directly affect the maximum circuit operating rate. Applying the method of partial fractions to Eq. (5.39), and by performing an impulse analysis via the Laplace transform method, the step response of this circuit may be seen to be,

$$V_{\text{out}}(t) = V_{\text{step}} [1 + e^{-p_2 t} (p_2/z_2 - 1)] \quad (5.42)$$

where V_{step} is the nominal output step size, $V_2(0)$ is assumed to be zero, and p_2 and z_1 are as indicated above. For a maximum permissible error, a total settling time is seen to be required of the amount

$$t_{\text{min}} = -\frac{1}{p_2} \ln \left[\frac{-|\epsilon_{\text{max}}|}{\left(\frac{p_2}{z_1} - 1\right)} \right] \quad (5.43)$$

where

$$\epsilon = \frac{V_{\text{out}}(t) - V_{\text{step}}}{V_{\text{step}}} \quad (5.44)$$

For the integrators employed in a pipeline converter or in a sample and hold, it will almost always be the case that the circuit is used as an AC coupled voltage multiplier. In this configuration, the input voltage is fed through the bottom plate of C_p , providing a forward voltage gain at low frequencies of

$$A_x \approx \frac{C_P}{C_I} \frac{g_m r_O}{(g_m r_O + 1 + C_P/C_I)} \quad (5.45)$$

This places the high frequency pole at

$$P_2 \approx \frac{-g_m}{(C_L(1 + A_x) + C_P)} \quad (5.46)$$

and the ratio of the pole and zero locations at

$$\frac{P_2}{z_1} \approx \frac{C_P}{A_x(C_L(1 + A_x) + C_P)} \quad (5.47)$$

In most practical situations, the zero will not seriously affect the settling response unless the integrator output is fed back to its input through additional circuitry. While this is not the case here, this problem may be alleviated in general, by placing a resistor in series with the integrating capacitance. This is an identical situation to that discussed in Section 5.3 for the R-C compensation technique, which is applied to the integrating second gain stage of an amplifier.

While the preceding analysis has assumed the use of a single stage transconductance amplifier, it is often the case that a multiple gain-stage circuit will be used. If it is assumed that a multiple stage amplifier exhibits a low frequency gain of a , and a dominant pole at P , it is easy to show that when used as an integrator, it will display a new pole frequency at

$$P' \approx \frac{2aP}{1 + A_x} \quad (5.48)$$

where A_x is as previously defined. Note that in both single and multiple stage circuits, care must be exercised to avoid poor settling response due to the presence of non-dominant poles. These poles may be

due to capacitive loading on the amplifier output (in multiple stage amplifiers), or due to internal poles from current mirrors or level shifters. As a general rule, all nondominant poles should be kept at least as high as the dominant pole position after the application of feedback.

Note that because the dominant pole frequency decreases as the forward gain increases, the amplifier need not be heavily compensated for cases where A_x is large. Advantage can be taken of this fact in AC-coupled multipliers with gain greater than one, by reducing the amplifier compensation capacitance, and hence increasing the circuit bandwidth. This is not always practical, however, as most AC-coupled multipliers of this type must periodically be offset cancelled to eliminate drift from leakage currents. This offset cancellation cycle usually entails shorting the inverting input and output nodes together, requiring that the amplifier be unity gain stable. In applications which need only be reset occasionally, such as between lines in a video system, it would seem plausible to electrically switch between several compensation capacitors depending upon whether the circuit was in a reset mode or in a mode requiring gain. Further, if different gains are programmable, optimal bandwidth performance may be obtained by programming the total compensation capacitance as well as the components which determine the forward gain.

5.4.2 Power Supply Rejection

A. General Case

The sources of power-supply coupling in integrators are considerably different than those which affect circuits with DC feedback. In most

monolithic implementations of this circuit, supply coupling via parasitic capacitance on the integrating node will dominate PSRR behavior. This capacitance may be switch or layout related, or associated with the amplifier used in the integrator. The parasitic capacitances which are most important in terms of supply rejection, are shown in Fig. 5.15. In this figure, device M_1 represents the input device of the amplifier, while impedance r_1 represents the impedance from any input stage load ($1/g_{m1load}$ if a current mirror). The gain block represents the remainder of the amplifier, while capacitances C_{P1} and C_{P2} are switch or layout parasitics. Note that this circuit may be used to model either single or two stage amplifiers with single-ended or differential inputs, by modifying the various circuit component values. It is easy to show for this circuit, that the low frequency supply rejection may be given by

$$\frac{\partial v_{out}}{\partial v_{SS}} \approx \frac{1}{C_I} \left[C_{gs} \left(\frac{\partial I}{\partial v_{SS}} \cdot \frac{1}{g_{m1}} + \frac{\partial v_{T1}}{\partial v_{SS}} \right) + C_{P2} \right] \quad (5.49)$$

and

$$\frac{\partial v_{out}}{\partial v_{DD}} \approx \frac{-1}{C_I} \left[C_{gd} \left(1 - \frac{\partial I}{\partial v_{DD}} \cdot r_1 \right) + C_{P1} \right] \quad (5.50)$$

The first of these two equations is typically the more troublesome of the two, in that C_{gs} is typically fairly large. This is particularly true in amplifiers which must use a high transconductance or a low noise input device. The first of the terms in Eq. (5.49) may be eliminated by employing a supply independent current source. Note, however, that this is not possible in some of the single-ended amplifiers presented in Section 5.2. The second term in Eq. (5.49), may be reduced by using

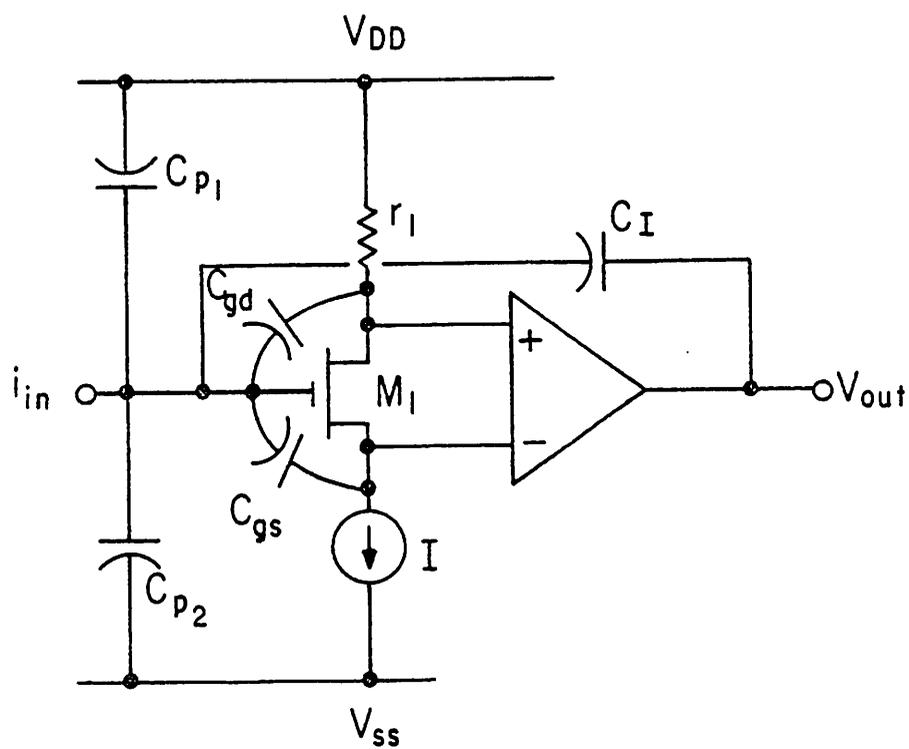


Fig. 5.15 Sources of power supply coupling in a real integrator.

input devices which are on a very lightly doped substrate, or eliminated by placing the input devices in an isolated well. In most CMOS processes, this latter technique is usually preferable. If a supply independent current source is used, the transfer function from the positive rail becomes simply $-(C_{gd} + C_{p1})/C_I$. This may result in poor supply rejection for large width input devices, even in a self-aligned process. To eliminate this problem altogether, a cascode circuit may be used to buffer the drain of the input devices from positive supply variations [95].

Although the above techniques will reduce the amplifiers contribution to the total supply coupling, the switch and layout parasitics have yet to be considered. The simplest means of reducing feedthrough from these sources, is to minimize the switch sizes and use a large integrating capacitance. Regretfully, this will directly reduced the circuit speed as well. Alternately, driving the switch gates between supply-independent levels, and the use of an on-chip supply regulator may allow improved performance [96].

B. Effect of Offset Cancellation Cycle

In many integrator applications, an offset cancellation cycle is used to eliminate the effect of leakage currents and amplifier drift. In input sample and holds this is typically performed once per conversion, while in pipeline converters it is performed on alternate clock cycles. As power supply coupling may be viewed as an apparent drift in amplifier offset, it will tend to be reduced by the reset procedure. In fact, assuming delta function sampling, the supply transfer function to the output of an integrator becomes modified, to

$$H_s(f) \approx H_c(f) \left[1 - \frac{\sin(\pi T_s f)}{\pi T_s f} \right] \quad (5.51)$$

where $H_c(f)$ is the continuous (unlocked) transfer function, $H_s(f)$ is the new transfer function and T_s is the reset frequency.

As is apparent from this equation, the supply rejection will be substantially improved at low frequencies. This reset procedure will be of little help, however, in eliminating power supply noise at, or above the clock frequency. For good rejection at these frequencies, a differential-in, differential-out integrator must usually be used which is well matched on each side. As power supply coupling will appear as a common-mode signal, it will be attenuated (at even high frequencies) by differencing between sides. This technique has been used to improve supply rejection in a monolithic filter [97] and should be useful in converter applications as well.

5.5 Output Stages

The circuitry discussed to this point is useful in on-chip applications with capacitive loads, but is not generally applicable to off-chip drivers or on-chip circuitry which must drive resistors. In these situations, circuitry with a lower output impedance and a higher current drive is usually required. Typically, this can assume the form of a buffer circuit which is used in conjunction with the transconductance amplifiers previously discussed.

The simplest form of output circuitry for an MOS amplifier, is the source follower (Section 5.1). While this type of circuit does provide relatively good bandwidth, it is fundamentally limited in terms of output current capability, by the constant pull-down current source. Although this current may be designed to be as large as necessary, the efficiency of this circuit is extremely low. A preferred embodiment of

a buffer circuit is shown in Fig. 5.16. In this figure, the Class-A source follower has been replaced by a Class-B output driver which is typically merged with a Class-A gain stage. This circuit is usually preferable to the simple source follower, in that it is much more efficient under resistive loading, and is easily designed to offer symmetrical output drive.

A problem with this technique in MOS implementations, however, is that the dynamic range of the circuit is reduced because of body-effect related threshold increases at large output voltages. These threshold increases can substantially reduce the maximum output swing. In a conventional CMOS process, the swing towards the positive rail may be improved by placing each of the NMOS devices in a separate p-well, which is shorted to the respective device's source. This prevents any NMOS threshold increases from occurring, regardless of the output voltage. Maintaining a good swing in the negative direction, however, is more difficult. One approach is to replace the PMOS output source follower with an NMOS pull-down circuit which functionally resembles the PMOS device, but which is not affected by threshold changes at negative output values. A circuit of this type which is commonly employed in NMOS circuits, is shown in Fig. 5.17 [98,99]. In this stage, M_{24} acts as a simple source follower while M_{20} and M_{21} act as a differencing circuit between V_{in} and V_{out} . Depending upon the values of V_{in} and V_{out} , a voltage is sent to M_{25} which determines the magnitude of the total pull-down. The output swing of this circuit in the negative direction is a strong function of output loading and device geometries, but in its simplest form is just $V_{SS} + V_{gs21} + V_{gs25}$. Depending upon the specifics of the process, and the circuit, this may or may not be an improvement over the negative swing of the Class-B

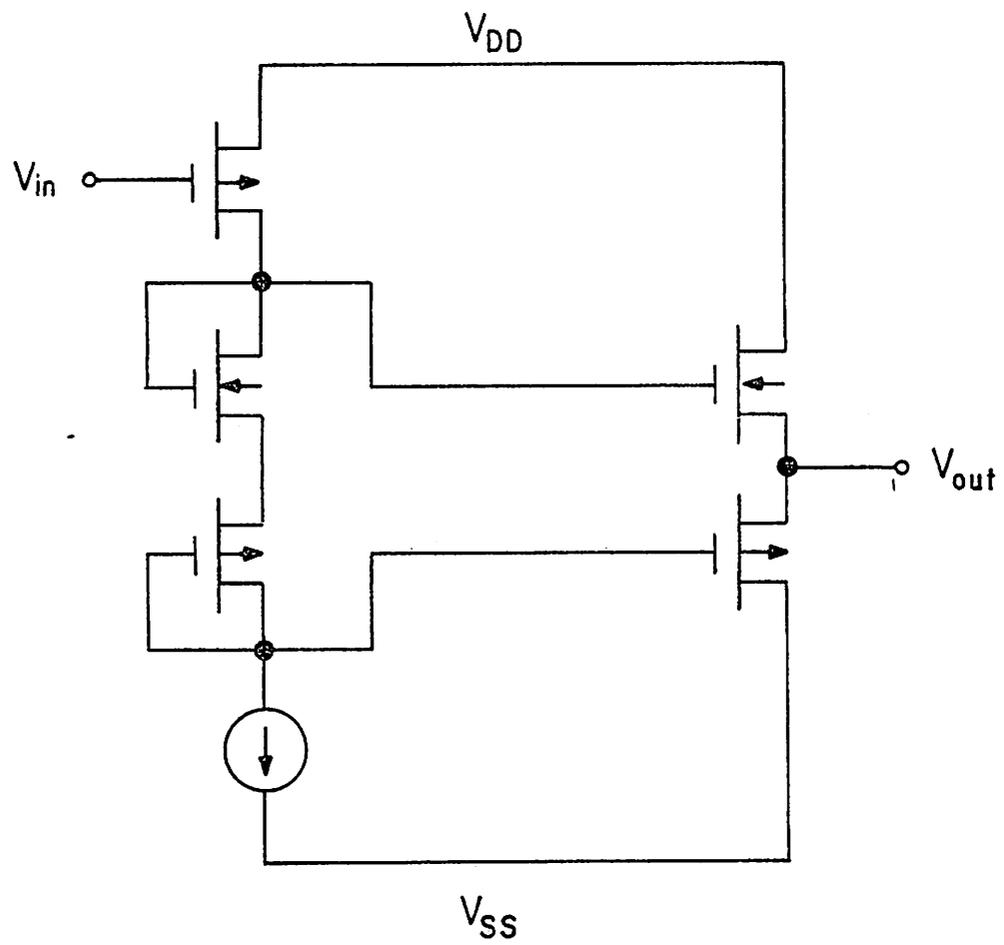


Fig. 5.16 Class A-B gain and output stage.

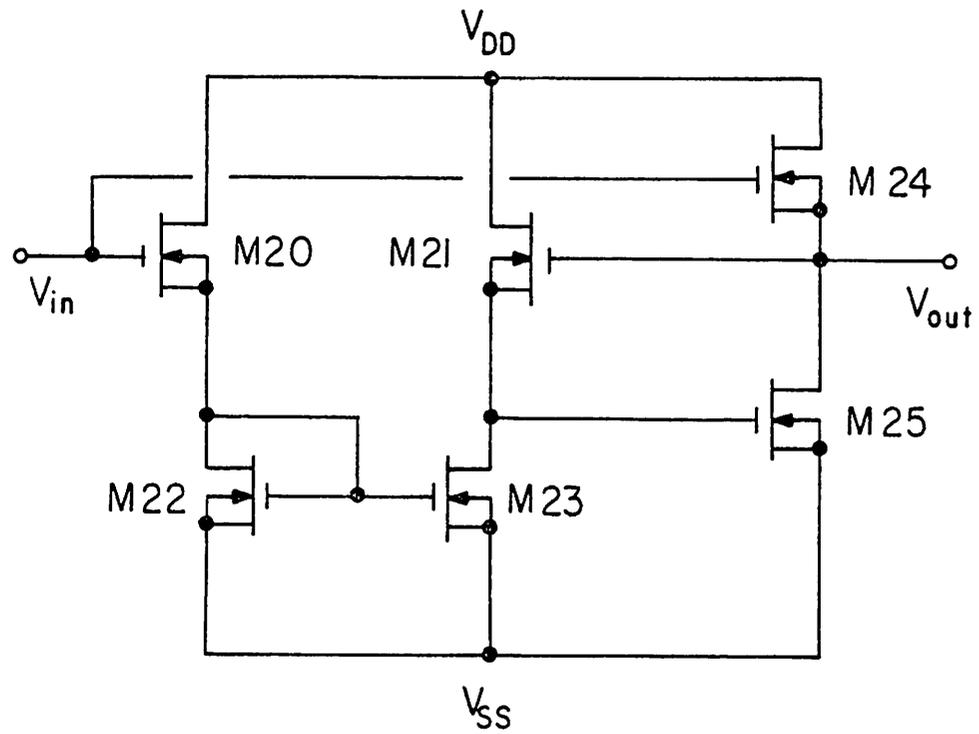


Fig. 5.17 Push-pull output stage commonly employed in NMOS circuits.

source follower. A problem with this circuit, however, is that the quiescent current through the output stage is a strong function of output voltage. This actually reduces the positive output swing, and makes the output pole frequency level dependent. This can lead to level dependent oscillatory behavior, or a poor settling response with certain output signal amplitudes. These problems may be reduced by employing the CMOS modification of this circuit, shown in Fig. 5.18. In this circuit, M_{20} and M_{24} are source followers while M_{25} is again the pull-down device. In this case, however, M_{25} is biased by a CMOS differential stage which maintains V_{out} at approximately the same potential as the source of M_{20} . Because of this feedback, the quiescent current through the output stage is kept at a constant level regardless of the output voltage ($I_3 \approx I_1 B_{24}/B_{20}$). Further, the swing in the negative direction is very good as no cascoding of NMOS devices occurs between the output node and the gate of M_{25} . Thus the major problems of the previous circuit have been alleviated.

Under certain circumstances, however, the circuit of Fig. 5.18 will still present some problems. Specifically, because local feedback is taken from the output, peaking will usually occur with even moderate capacitive loads. Note that this will also be the case, in the circuit of Fig. 5.17. In these situations, the most straightforward approach is to eliminate the local feedback from the output entirely and attempt to modify the simple Class-B follower for improved negative swing. One possible modification, is to use a depletion MOSFET for both the diode connected and source follower PMOS devices. As a depletion PMOS device will exhibit a low threshold even with a large body bias, the maximum output swing towards the negative rail will be greatly improved.

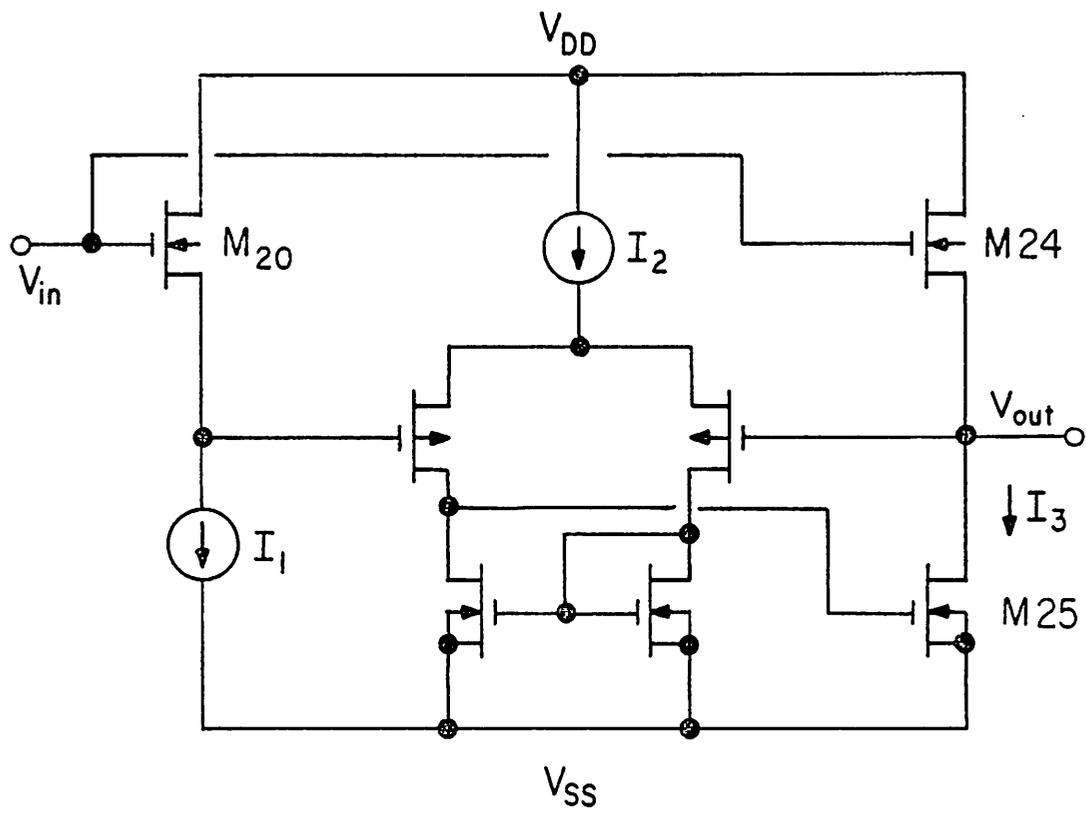


Fig. 5.18 CMOS push-pull output stage.

Depletion devices however, if used in this circuit, may allow widely varying currents to flow through the output devices. This will occur because the diode-connected PMOS device may not always be in the saturation region. This problem may be alleviated by using the output stage shown in Fig. 5.19. In this circuit, constant quiescent current is maintained, regardless of the output level, by sensing the current through the "dummy" output pair M_{20} and M_{21} by way of the PMOS current mirror M_{22} and M_{23} . The current being pulled by M_{23} forces the gates of depletion devices M_{21} and M_{25} to a point which will keep the current through M_{20} and M_{21} constant, regardless of the depletion PMOS threshold voltage. As M_{24} and M_{25} are in parallel with M_{20} and M_{21} , the current through these devices is also stabilized, although at a potentially higher level depending upon the device geometries. This output stage has been used in a high order monolithic filter circuit, to obtain excellent stability, output swing and drive efficiency in the presence of even 1000 pF and/or 300 Ω loads [100]. Further, this performance has been obtained with a quiescent power of only a few milliwatts.

5.6 Voltage Comparators

Every type of quantization circuit requires some means of thresholding or comparing. In the analog/digital conversion circuits described here, this typically entails performing a number of rapid voltage comparisons between an unknown input voltage and a number of known reference standards. As demonstrated in Chapter 3, these comparisons may be performed: in parallel by a number of distinct comparators, sequentially by a single comparator; or both. In many of these techniques, the ultimate conversion rate is often determined by the comparator's speed, while

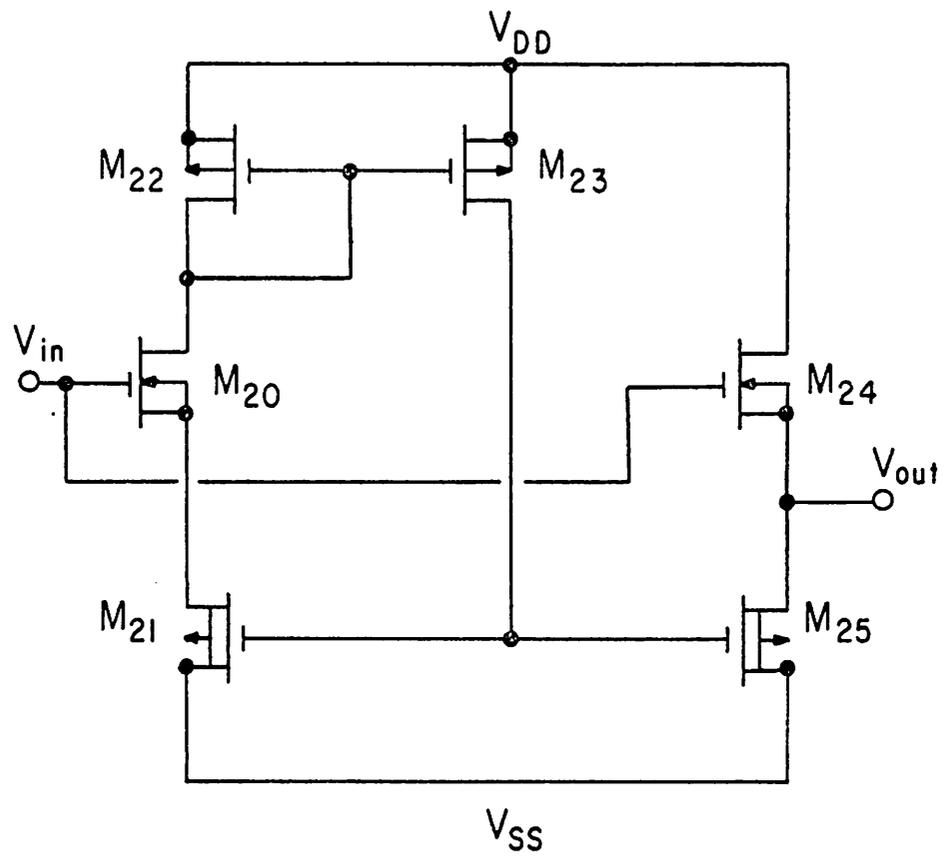


Fig. 5.19 Wide dynamic range Class-B output stage.

in some cases the comparators limit the conversion accuracy as well.

5.6.1 Single Ended Comparators

The simplest form of comparator, is just a single-ended amplifier with very high gain. In this type of circuit, the voltage comparison is between the present input voltage (and it's recent history) and the amplifier offset point. In order to eliminate potential output ambiguities, the amplifier is typically used in conjunction with a regenerative stage as shown in Fig. 5.20a. In this configuration, the effective comparator offset may be given by (ignoring gain nonlinearities and regenerator hysteresis)

$$V_{os_T} \approx V_{os_1} + V_{os_2}/a \quad (5.52)$$

where V_{os_1} is the offset of the gain stage, V_{os_2} is the 1-0 trip point of the regenerator, and a is the gain of the amplifier. As discussed in previous sections, the intrinsic offset of the amplifier may usually be designed to be small. Invariably, however, process and device variabilities will tend to make offsets distributed about their mean (intrinsic) value. As a result, a converter which uses these circuits will also display an offset, or in some cases (most notably in pipelines) a comparator induced nonlinearity. Although the offset distribution may be made small by making the initial gain contributing devices of the amplifier large and interdigitized, this invariably impacts the overall die size and may effect the comparator's speed. A preferable method is shown in Fig. 5.20b. In this circuit the comparator is capacitively coupled to the input voltage, allowing the comparators offset to be stored on the input capacitor. This is accomplished by shorting the input and

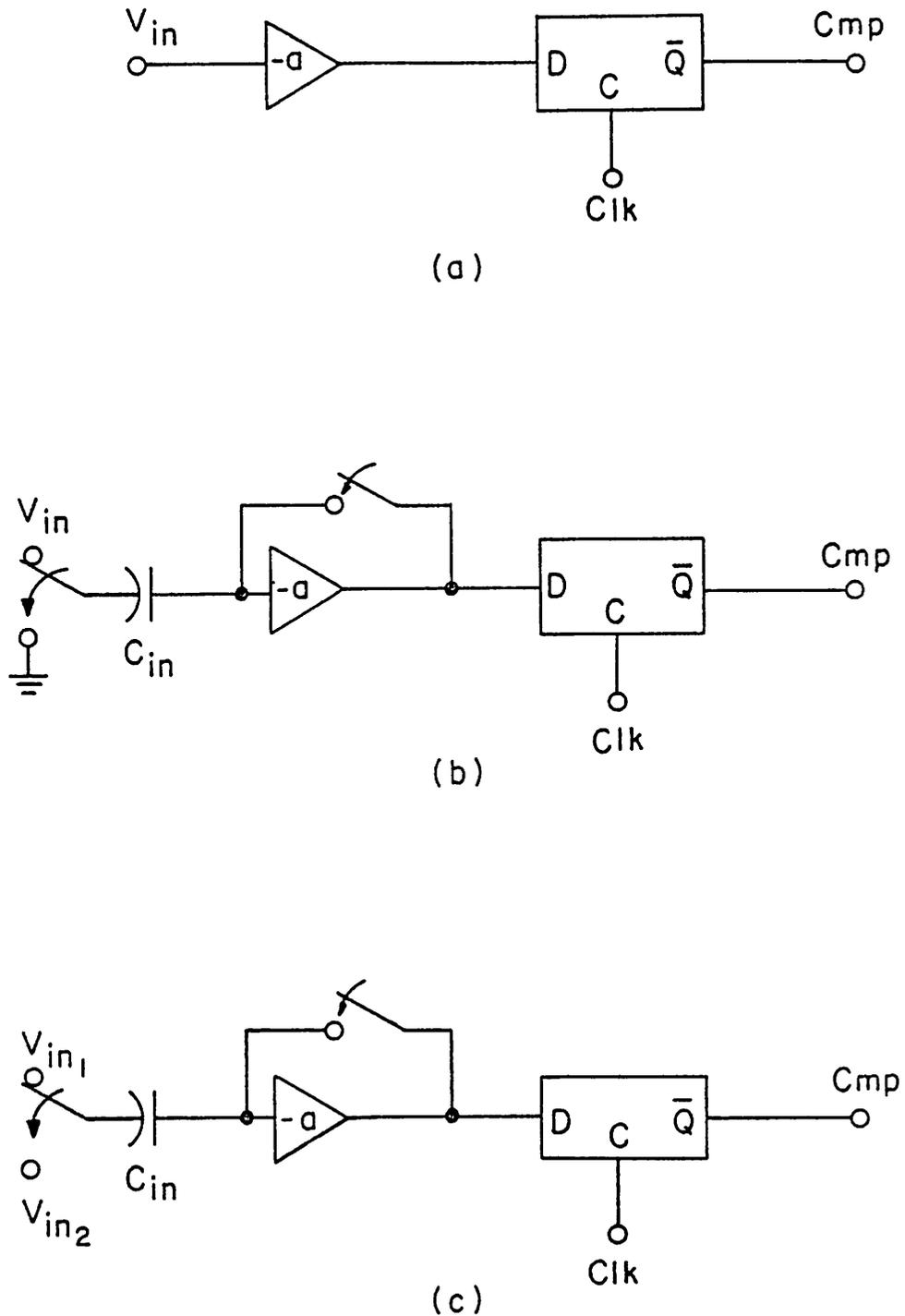


Fig. 5.20 Single-ended comparators. (a) Simple gain stage and regenerator cascade. (b) A-C coupled comparator with first-order offset cancellation. (c) Same as in "b" but with input differencing.

output nodes of the amplifier together, and at the same time grounding the input side of the coupling capacitor. After this reset cycle, the input capacitor is again switched to the input voltage, allowing comparisons to be performed as before. This has the effect of modifying the total input offset voltage, to

$$V_{os_T} \approx (V_{os_1} + V_{os_2})/a + V_{ft_1} \quad (5.53)$$

where V_{ft_1} is the feedthrough resulting from the amplifier shorting switch. This feedthrough results from both gate overlap capacitance and MOSFET channel charge pumping [10] as the shorting switch turns off. For circuits employing a small shorting switch and a large input capacitor this latter term may be kept small, although in most CMOS comparators which use a high gain input amplifier, the feedthrough term will usually dominate the total circuit offset.

Although the above circuit must be periodically reset for offset cancellation to occur, this need not happen before every comparison. Typically, the reset frequency will be the same as the converter sampling rate, as leakage currents will (usually) create only a sub-LSB drift over one conversion cycle. In fact, in converters employing a binary weighted capacitor technique, the comparator during its reset cycle will usually function as the input sample and hold. Note that this offset-cancelled comparator may be operated in a differential mode between two independent input signals, as shown in Fig. 5.20c.

In high precision converters, this simple offset reduction method may not provide a sufficiently small offset. This will be particularly true in circuits which have employed large reset switches for increased speed. Although a feedthrough cancellation device or capacitor may be

used to provide some reduction in switch related offset, the degree to which cancellation occurs will typically be process dependent. Fortunately, the offset reduction technique described above can be modified to reduce first order switch feedthrough effects [102]. This technique is shown in Fig. 5.21. In this circuit, several offset cancelled stages are cascaded together, but with slightly different reset phasing. By continuing to offset cancel the second stage even after the first stage reset switch turns off, the feedthrough from the first stage is effectively eliminated. Further, the effective switch feedthrough from the second stage is attenuated by the gain of the first. In fact, the total offset is now given by

$$V_{os_T} \approx ((V_{os_2} + V_{os_3})/a_2 + V_{ft_2})/a_1 \quad (5.54)$$

This cascoding procedure may be continued indefinitely (at least to the thermal noise limit), until the required offset is obtained. Note that care must be exercised in circuits of this type, to ensure that the feedthrough of any stage when multiplied by the stage gain does not result in gain stage clipping. This may require that individual stage gains be kept fairly low, or that feedthrough compensating devices or capacitors be used to keep the overall feedthrough at a tolerable level.

5.6.2 Differential Comparators

In many situations, the single-ended comparators just presented will be unsatisfactory. This will certainly be the case in any application which requires a differential comparison between two rapidly changing signal levels, or in applications which require good common mode rejection at high frequencies. In these situations, the comparator

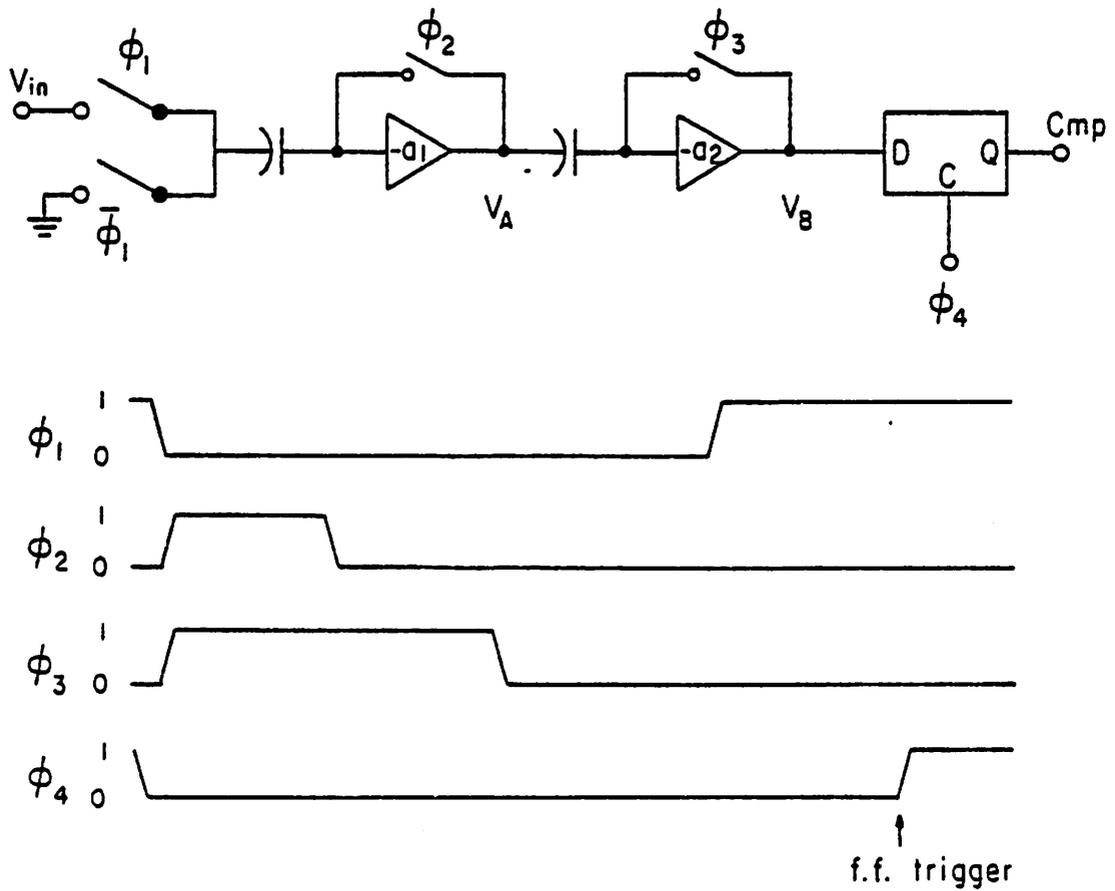


Fig. 5.21 Very low offset comparator employs a cascade of offset cancelled amplifier stages.

configurations are modified, as shown in Fig. 5.22. The first of these figures (Fig. 5.22a) shows a direct coupled differential amplifier used with (again) a regenerator. Note that in this case, however, the flip-flop is used to differentially discriminate between two signal levels, rather than between a single input and the trip-point. As before, both stages will contribute to a net comparator offset, as given by Eq. (5.52). By capacitively coupling the inputs of the gain stage to the differential input lines, an offset cancellation may again be made to occur, as indicated by Eq. (5.53) and as shown in Fig. 5.22b. As before, this technique may be expanded to include multiple AC-coupled gain stages, as shown in Fig. 5.22c.

5.6.3 Bandwidth Considerations

In the circuits must presented, there are two distinct types of comparator operations. The first occurs only in circuits employing offset cancellation cycles, and is when one or more stages are being reset, by shorting their input(s) and output(s) together. The second type of operation is that normally associated with a comparator; open loop differencing between an input voltage (absolute or differential) and the comparator offset point. These two cases are analyzed below.

A. Reset Cycle

As the reset cycle usually only occurs once per conversion, and possibly less, the effect of this operation upon the maximum conversion rate is usually small.³ This operation is extremely important, however,

³Even in flash converters which employ a reset cycle fully half of the time, the comparator is still not usually the speed limiting factor. Typically ladder R-C charging times will determine the maximum useful conversion rate.

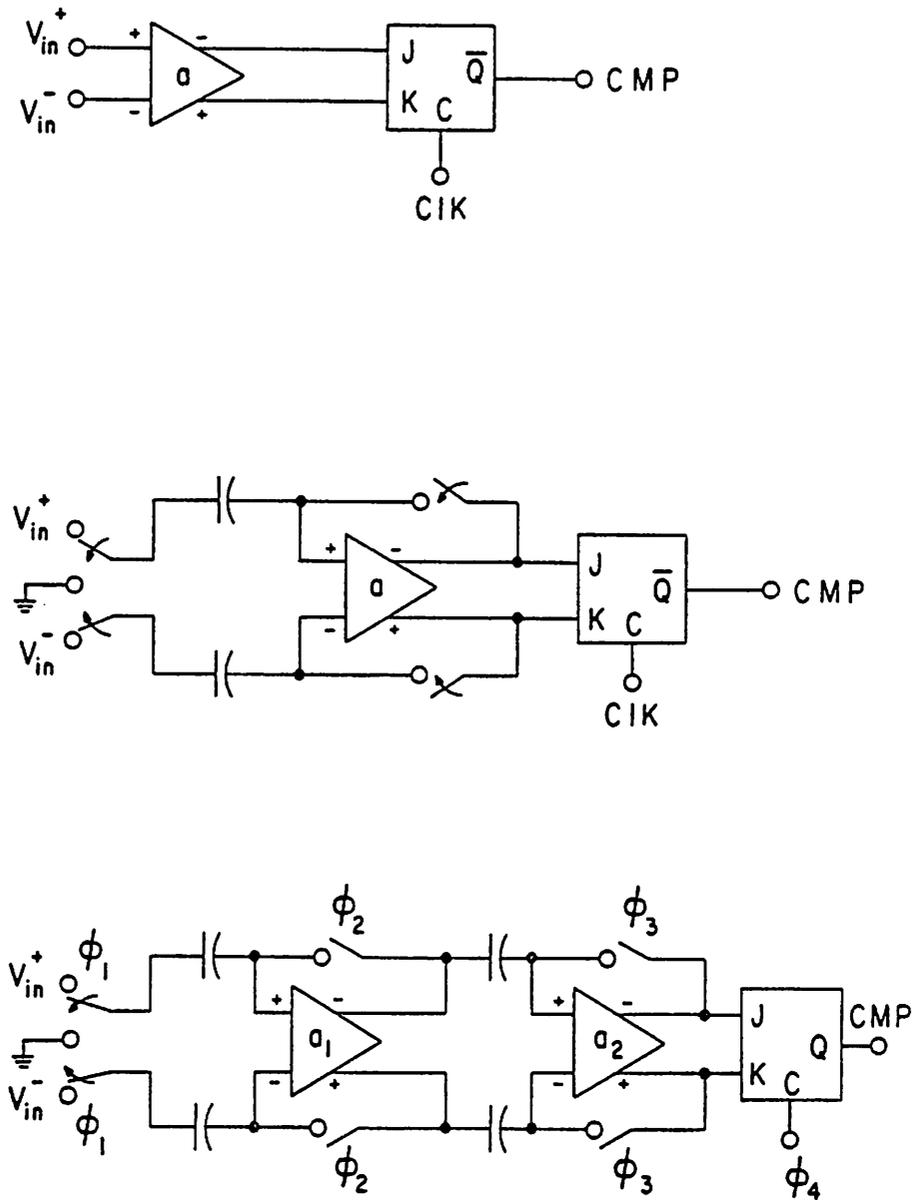


Fig. 5.22 Differential comparators. (a) Direct-coupled gain stage and regenerator. (b) A-C coupled comparator with one stage of offset cancellation. (c) Multiple stage offset cancellation.

if the input stage of the comparator is used as the input sample and hold. In these situations, the first comparator stage will generally determine the converters distortion characteristics for high frequency input signals.

A small signal model of an inverting gain stage during a reset cycle, is shown in Fig. 5.23. In this figure, C_L and r_0 represent the output capacitance and equivalent load impedance, while r_f is the feedback switch impedance, C_{in} is the input capacitance and R_{in} is any series input switch resistance. It may be shown for this circuit, that

$$\frac{V_1}{V_{in}} \approx \frac{s(C_{in}/g_m)(1 + sr_f C_L)}{1 + As + Bs^2 + Cs^3} \quad (5.55)$$

where

$$\begin{aligned} A &= C_{in}(1/g_m + r_{in}) + C_L/g_m \\ B &= \frac{C_{in}C_L}{g_m} (1/g_m + r_{in}) \\ C &= \frac{C_{in}^2 C_L r_{in}}{g_m} (1/g_m - r_f) \end{aligned} \quad (5.56)$$

For cases where $1/g_m \approx r_f$ and $C_{in} > C_L$, this reduces to

$$\frac{V_1}{V_{in}} \approx \frac{s(C_{in}/g_m)}{(1 + sC_{in}(1/g_m + r_{in}))} \quad (5.57)$$

This equation indicates that the input to the comparator may not be properly reset for high frequency input signals. In fact, at the end of the sampling period, the comparator may exhibit an input dependent offset, which will substantially reduce the signal to noise ratio of high frequency input signals. In practice, this will require that the comparator -3dB frequency be 50-100 times higher than the maximum full-

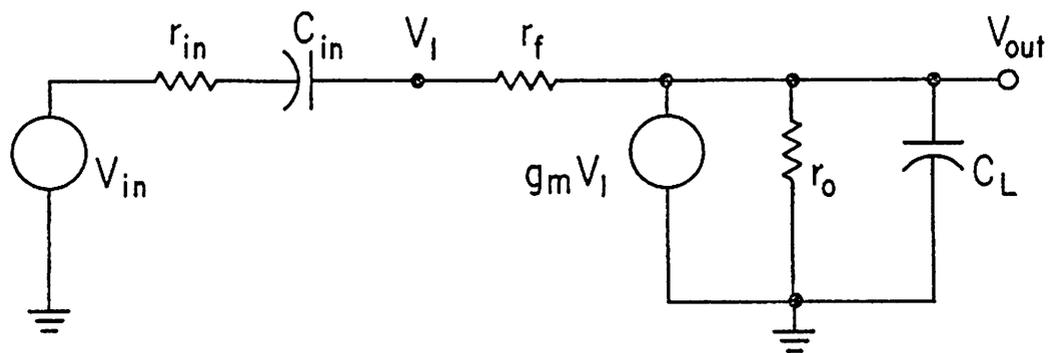


Fig. 5.23 Small signal model of the input gain stage of a comparator during a reset cycle.

scale input frequency for a 40dB signal to noise ratio (see Chapter 6).

In addition to introducing distortion, the comparator will also attenuate the sampled value. The voltage across the sampling capacitor may be seen to be

$$V_{\text{sam}} \approx V_x - V_1 \approx \frac{V_{\text{in}}}{(1 + sC_{\text{in}}r_{\text{in}})} \left[1 - \frac{sC_{\text{in}}/g_m(1 + sr_f C_L)}{1 + As + Bs^2 + Cs^3} \right] \quad (5.58)$$

which, for $g_m \approx 1/r_f$ and $C_{\text{in}} > C_L$, may be approximated by

$$V_{\text{sam}} \approx \frac{V_{\text{in}} sC_{\text{in}} (r_{\text{in}} + 1/g_m)}{(1 + sC_{\text{in}}r_{\text{in}})(1 + sC_{\text{in}}(r_{\text{in}} + 1/g_m))} \quad (5.59)$$

In cases where $1/g_m \approx r_{\text{in}}$, V_{sam} will only be attenuated from V_{in} by -9.5dB at a frequency of approximately g_m/C_{in} . In general, this is not as severe a problem as the distortion case considered above.

B. Compare Cycle

For comparators which always operate in the linear region, the relevant time constant for each comparator stage is simply g_m/C_L . Thus for an n-stage cascaded comparator, a minimum comparison time will be required of the amount

$$\tau_{\text{cmp}} \approx \sum_{i=1}^n \frac{g_{m_i} \ln \epsilon_{\text{max}}}{C_{L_i}} \quad (5.60)$$

where ϵ_{max} is the allowable relative error. In practice, however, few comparators behave linearly over a large input range. This is particularly true for circuits which employ Class-A gain stages, where a significant slewing delay may be encountered. In these circuits an additional

time must be added to each term of Eq. (5.60), which indicates this slewing delay at each stage.

In comparators employing a two stage amplifier, the slew rate may well be determined by stability requirements for the comparator reset period (if one is used). Specifically, the requirement that the amplifier be stable under unity gain conditions may force the use of a large compensation capacitor, which will considerably degrade amplifier slew rate. In these situations, the compensation capacitor may actually be switched out during the compare interval, as no feedback is used.

5.7 Noise in Analog Circuits

Nonideal behavior in A/D converters is most often due to a mismatch in precision components, or a limited circuit bandwidth. In high precision converters, it may also be due to device noise. According to most MOS models [103,104], the drain noise current of a MOSFET in saturation over a bandwidth Δf , is given by

$$\overline{i_d^2} = 4kT \left(\frac{2}{3} g_m \right) \Delta f + \frac{KI_d^a}{f^b} \Delta f \quad (5.61)$$

where k is Boltzman's constant, T is the absolute temperature and K , a and b are device constants. The first term in Eq. (5.61) is thermal noise resulting from a nonzero channel resistance, while the second term is flicker noise created by (apparently) surface carrier trapping. This drain noise current may be expressed in terms of an equivalent gate noise voltage, by dividing through by g_m^2 , hence

$$\begin{aligned} e_n^2 &= \frac{8kT}{3g_m} \Delta f + \frac{KI_d^a}{g_m^2 f^b} \Delta f \\ &= e_t^2 + e_f^2 \end{aligned} \quad (5.62)$$

where e_f may be expressed as [105]

$$e_f \approx \sqrt{\frac{a_f}{WL}} \quad (5.63)$$

In this equation, W is the device channel width and L is the length, while a_f is a single parameter which incorporates all of the previous parameters above. Using Eq. (5.62), an apparent input noise voltage may be found for any circuit, by rms summing the noise contributions of each device at the circuit output, and dividing by the circuit's transfer function. Specifically, if H_i is the transfer function from the gate of the i th device to the output, and H_p is the principal transfer function, then the equivalent circuit input noise is given by,

$$e_n^{in} \approx \frac{\left(\sum_{\text{all } i} e_{n_i}^2 H_i^2 \right)^{1/2}}{H_p} \quad (5.64)$$

If circuit frequency dependence is included in this equation, it becomes fairly involved for even simple circuits. For these situations computer analysis will usually be necessary. At low frequencies, however, where flicker noise dominates and circuit behavior is similar to that at DC, most circuits are easily analyzed. The results for three simple circuits are given below [106];

a) Source follower (Fig. 5.24a)

$$e_n^{in} \approx \left[\frac{a_f}{W_1 L_1} \left[1 + \left(\frac{L_1}{L_2} \right)^2 \right] \right]^{1/2} \quad (5.65)$$

b) Common source amplifier (Fig. 5.24b)

$$e_n^{in} \approx \left[\frac{a_{f1}}{W_1 L_1} \left[1 + \frac{\mu_n a_{f2}}{\mu_p a_{f1}} \left(\frac{L_2}{L_1} \right)^2 \right] \right]^{1/2} \quad (5.66)$$

c) Differential input amplifier (Fig. 5.24c)

$$e_n^{in} \approx \left[\frac{2a_{f1}}{W_1 L_1} \left[1 + \frac{\mu_n a_{f2}}{\mu_p a_{f1}} \left(\frac{L_1}{L_2} \right)^2 \right] \right]^{1/2} \quad (5.67)$$

As is evident from all of these equations, noise performance at low frequency is improved by lowering the load transconductance relative to that of the driver. In practice, this is easier to achieve with a PMOS load and an NMOS driver (in gain stages), where advantage may be taken of the ratio of PMOS to NMOS mobilities.

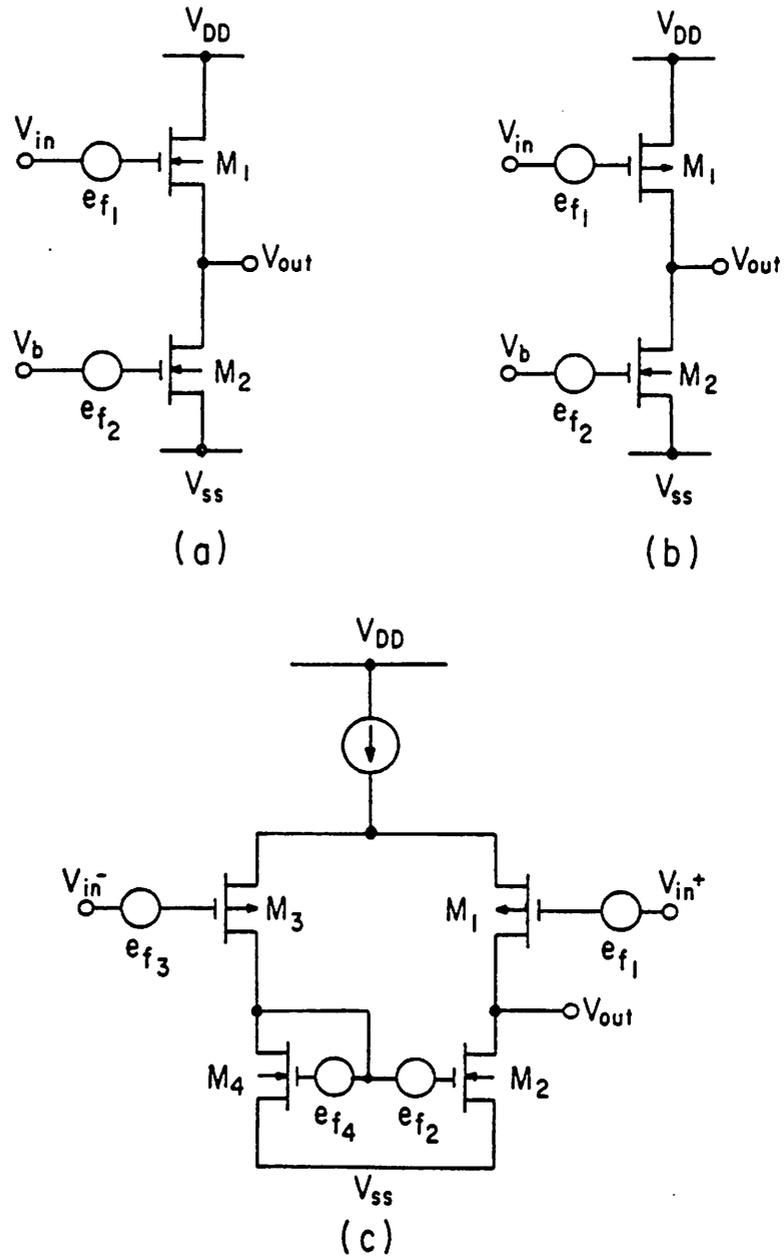


Fig. 5.24 Noise models for Eqs. (5.65)-(5.67).

CHAPTER 6

TIME INTERLEAVED CONVERTER ARRAYS

As discussed in Chapters 1 and 2, there is considerable motivation for reducing the production costs of high-speed A/D converters. Present day high-speed converters use techniques which require large die sizes and/or fairly exotic fabrication processes, which makes them quite expensive. Furthermore, these A/D techniques will be exceedingly difficult to integrate along with a VLSI digital signal processor.

In the method presented here, the ultimate cost of a high-speed converter may be considerably reduced by employing the converter-array technique shown in Fig. 6.1. As will be shown, this technique may be used to achieve a substantial reduction in die size and power dissipation over other circuits, yet without sacrificing bandwidth or signal-to-noise ratio. This type of converter configuration is easily implemented in a MOS technology, allowing on-chip compatibility with dense digital signal processors. Furthermore, this technique may be used to achieve faster sampling rates than are possible with any single converter.

6.1 Comparison with Other Methods

The various characteristics of several common high-speed converter types have been discussed at length in the previous chapters. A summary of the more relevant characteristics of these methods is presented in Table 6.1. In this table, it has been assumed that all converters are implemented in the same MOS process. Relative values may vary by perhaps 30% depending upon specific assumptions. Note that the conversion rate

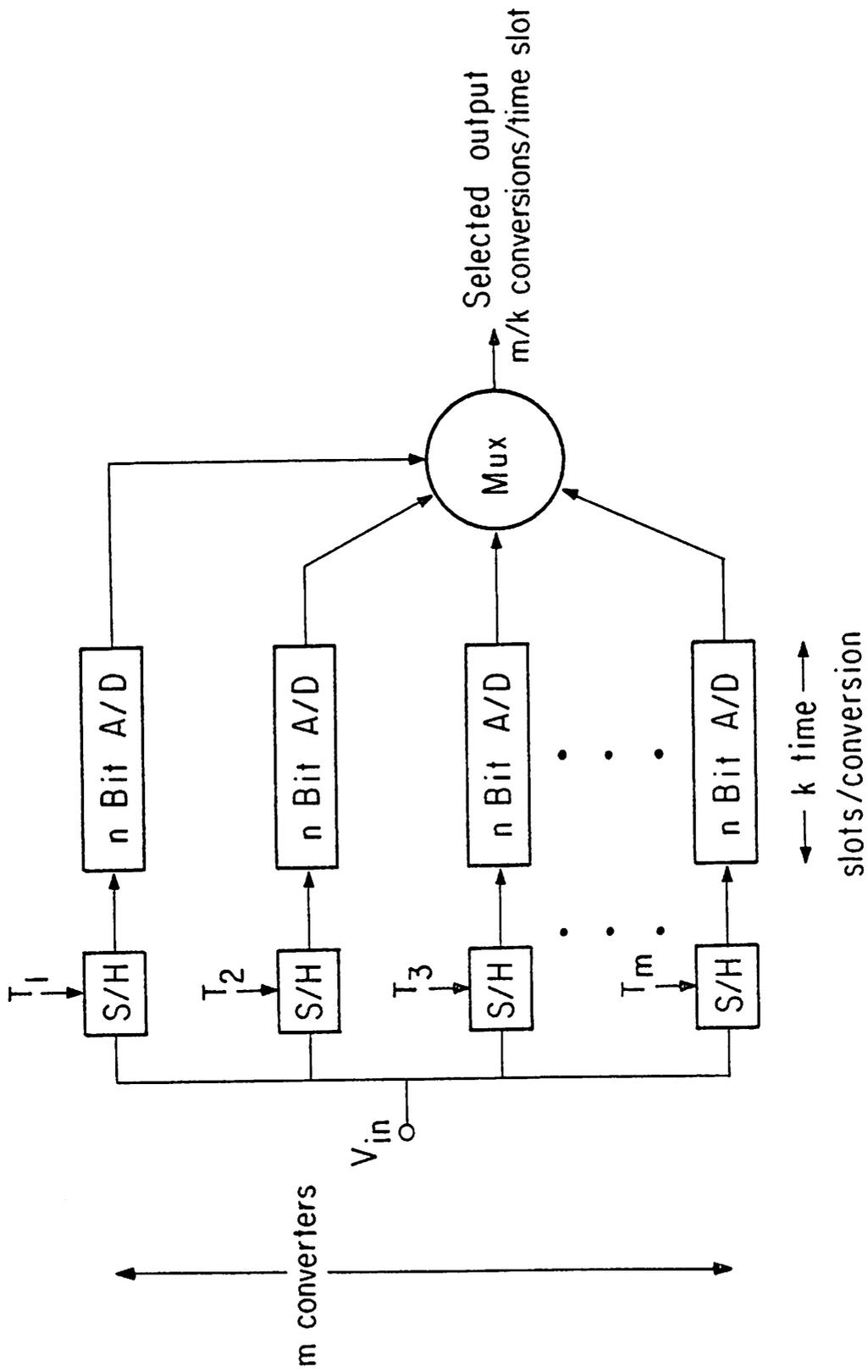


Fig. 6.1 Converter array technique uses a number of converters with interleaved sampling times to achieve a fast effective conversion rate. An n -bit, m -way array is shown.

TABLE 6.1

Method	Relative Die Size D	Time Slots per Conversion T (n=7)	Throughput per Area $1/(D \cdot T), n=7$ $\times 10^3$
• 1-Step Parallel (Flash)	2^n	2	4
• 2-Step Parallel (1/2-Flash)	$3 \cdot 2^{n/2}$	3-4	7-9
• n-Step Pipeline (Succ. Approx.)	4.5n	4-6 *	5-8 *
• n-Step Sequential (Succ. Approx.)	1.5n	n+1, (8)	12
• 4-Way Successive Approx. Array	4(1.5n)	(n+1)/4, (2)	12

*See text

Summary of A/D converter characteristics.

per time slot for the pipeline converter is actually half to a third of that indicated in Chapter 3 where converters of this type were discussed. This is because the minimum time slot for a pipeline converter is 2 to 3 times longer than that of the other converters listed (in a given process), due to amplifier settling requirements. It has been assumed for the successive approximation converters, that binary-weighted capacitors, or their equivalent, are used in the respective designs.

A figure of merit which is rarely cited in converter literature, is the relative conversion rate (samples per second) per unit of integrated circuit chip area. This is listed for each of the converter types in the last column of Table 6.1. As is evident, the converters which display the best area efficiency are in fact those with the lowest conversion rate. Thus, although the successive approximation technique is slower than any of the other converters listed, it usually will use its area more efficiently. It would be desirable, if a technique could be developed which would allow the throughput of parallel circuits to be obtained, yet with the area efficiency of the successive approximation method. As is indicated in the last row of Table 6.1, an array of interleaved successive approximation converters may be used to achieve this desired end. Note that for the 7-bit case, 4 parallel converters are needed to achieve a throughput equal to that of the flash approach, but with only a third of the required die-size. Further, this area savings becomes even more pronounced at higher resolution levels, as shown in Fig. 6.2.

This array technique may also be used to achieve significantly higher bandwidths than are possible with any single converter. By interleaving more converters, conversion rates limited only by the sample and hold aperture times may be obtained. Thus, at the expense of die area,

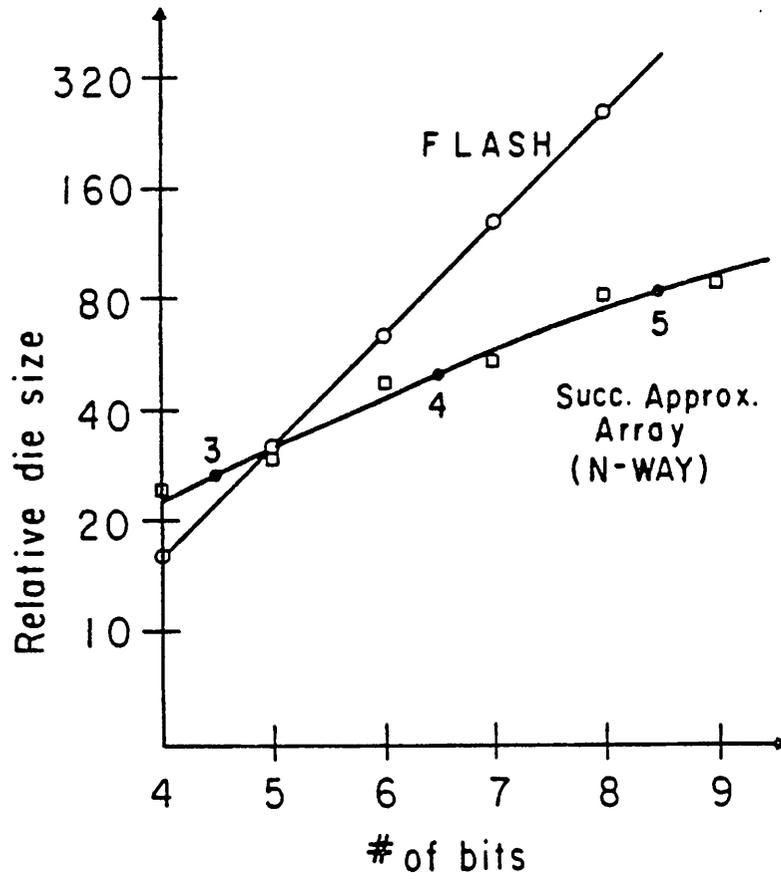


Fig. 6.2 Relative area comparison between flash and array converter methods as a function of resolution level. Small numbers along graph indicate size of array needed to equal flash conversion rate.

even a slow technology may be used to achieve fast conversion rates. Similarly, a fast technology may be used to achieve bandwidths that were previously impossible to obtain.

As was shown in Table 6.1, a total of $n+1$ time slots are needed for an n -bit successive approximation converter. Thus 9 different time states will be needed for an 8-bit converter. In order to reduce reconstruction or signal processing difficulties, it is generally desirable to stagger the sampling periods by constant time intervals, as shown in Fig. 6.3a. If the smallest unit of time displacement between converters is one time slot, then only certain array sizes may be realized for a given resolution converter. This is demonstrated in Fig. 6.3b. For example, in the 8-bit case just considered, only a three or nine way array is possible. If more time slots are added, however, new configurations may be realized. These additional time slots may be used to provide more charging time in the sample mode or MSB bit times¹ or even thrown away. Thus in the eight bit case, adding one time slot would allow 2, 5 or 10 way arrays to be used. Note that in most cases, allowing fractional time slot displacements will increase the number of allowable arrays. In particular, delays of $1/2$ period are quite simple to generate if the input clock duty cycle is exactly 50%. Extreme care must be taken if this is done, however, as an undesirable sampling time skew may develop between converters which trigger on the positive and negative clock edges. As discussed in the next section, this may lead to a considerable degradation in performance.

¹In some converters, making the higher-order bit times longer than others will lead to increased nonlinearity. This is apparently due to the comparator displaying set-up time dependent behavior [107].

T_1									T_9	
S	1	2	3	4	5	6	7	8	A/D	#1
6	7	8	S	1	2	3	4	5		#2
3	4	5	6	7	8	S	1	2		#3

Fig. 6.3a Staggering of time slot assignments maintains a uniform time interval between sampling periods. A 3-way 8-bit array is shown.

Total time slots	7	8	9	10	11	12
1 time slot min. delay	7	2 4 8	3 9	2 5 10	11	2 3 4 6 12
1/2 time slot min. delay (plus all above)	2		2	4	2	

Fig. 6.3b Allowable number of converters in an array for a constant period between sampling times.

6.2 Error Analysis

This time interleaved array technique may be used with any type of A/D converter. As will be shown, however, a variation of converter quantizing characteristics within an array may introduce additional noise or distortion. This will tend to make some converter types more suitable for use in arrays than others, and may result in a slightly increased component matching requirement for a specified performance level.

6.2.1 Noise and Distortion Error Power

As discussed in Chapter 3, the quantizing error of an A/D converter at an input voltage x , is given by

$$\epsilon(x) = \delta(x) \quad (6.1)$$

where

$$\delta(x) = \hat{x} - (ax + b) \quad (6.2)$$

in which \hat{x} is the quantized value, and a and b are best fit gain and offset terms which minimize the total error power,

$$E_p = \langle \delta^2 \rangle \quad (6.3)$$

If instead of a single converter, an array of interleaved converters is employed, the error analysis becomes more involved. This situation is depicted in Fig. 6.4, where the transfer functions of two converters in an array are shown, each of which exhibits different gain, offset and nonlinearity. For this case, the error in the i th converter is given by

$$\epsilon_i(x) = \delta_i(x) + \Delta_i(x) \quad (6.4)$$

where

$$\delta_i(x) = \hat{x}_i - (a_i x + b_i) \quad (6.5)$$

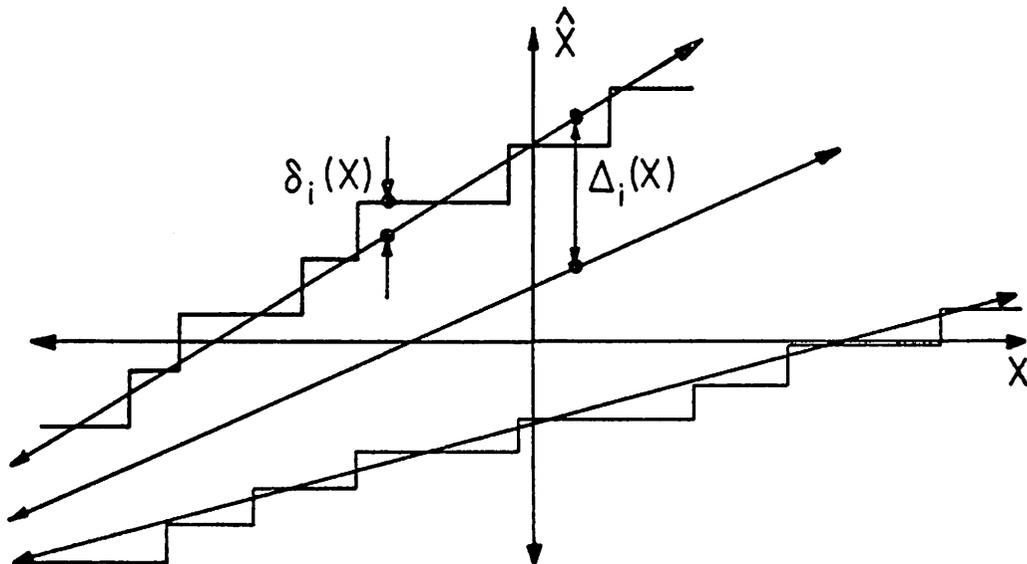


Fig. 6.4 Quantization and nonlinearity error in a converter array.

and

$$\Delta_i(x) = (a_i - a_{\min})x + b_i - b_{\min} \quad (6.6)$$

In these equations, $\delta_i(x)$ indicates the error that would exist if the i th converter were used alone, while $\Delta_i(x)$ indicates an additional error which exists when the i th converter is used as part of an array. The best fit gain and offset of the i th converter is represented by a_i and b_i , while the best fit gain and offset of the array as a whole is given by a_{\min} and b_{\min} . This results in an average error power for an m -way array, of

$$E_p = \frac{1}{m} \sum_{i=1}^m \langle (\delta_i + \Delta_i)^2 \rangle \quad (6.7)$$

or, expanding in terms of the above definitions,

$$\begin{aligned} E_p = \frac{1}{m} \sum_{i=1}^m [& (a_i - a_{\min})^2 \langle x^2 \rangle + (b_i - b_{\min})^2 \\ & + \langle \delta_i^2 \rangle + 2(a_i - a_{\min})(b_i - b_{\min}) \langle x \rangle \\ & + 2(a_i - a_{\min}) \langle \delta_i x \rangle + 2(b_i - b_{\min}) \langle \delta_i \rangle] \end{aligned} \quad (6.8)$$

For input signals which are asynchronous with the sampling clock, the last two terms of this equation are near zero, even for badly matched arrays. This results in the approximation,

$$E_p \approx \frac{1}{m} \sum_{i=1}^m [\langle \delta_i^2 \rangle + \langle \Delta_i^2 \rangle] \quad (6.9)$$

The minimizing values of a_{\min} and b_{\min} are obtained by differentiating Eq. (6.8) with respect to both variables and solving for the zero of each derivative. This occurs at simply $a_{\min} = \bar{a}_i$ and $b_{\min} = \bar{b}_i$, resulting

in an error power of

$$E_p = \sigma_a^2 \langle x^2 \rangle + \sigma_b^2 + \langle \delta_i^2 \rangle \quad (6.10)$$

in signals for which $\langle x \rangle = 0$, where

$$\sigma_x^2 = \overline{x^2} - \bar{x}^2 \quad (6.11)$$

These equations imply that converter gain and offset variations within an array will result in increased error power. These variations may be due to mismatches in gain or offset establishing components, or from differences in converter nonlinearities, which may result in a slightly different best fit gain and offset for each converter. In converters where these variations are well controlled, however, the array error power will be approximately the average of the individual error powers due to quantization and nonlinearity.

In both single converters and converter arrays, there is one additional source of error power; phase skew or jitter of converter sampling. Specifically, if the aperture timing of a comparator's sample and hold circuit exhibits any variance in period, the overall error power will increase. Phase skew is particularly troublesome in converter arrays, as it may result from even a slight difference in timing line layout or loading between converters.

The average error power for an m -way array of interleaved samplers may be given by

$$E_p = \frac{1}{m} \sum_{i=1}^m \langle (\hat{x}_i - x)^2 \rangle \quad (6.12)$$

where \hat{x}_i is the sampled value of x for the i th sample and hold. For a sinusoidal input waveform, and for distortionless samplers which exhibit

a timing skew,

$$x = A \sin(\omega t) \quad (6.13)$$

and

$$\hat{x} = A \sin(\omega(t + t_i - t_{\min})) \quad (6.14)$$

where t_i is the absolute delay in the i th sampler and t_{\min} is a minimizing sampling delay of the array as a whole. This results in an average error power given by

$$E_p = \frac{A^2}{m} \sum_{i=1}^m \langle (\sin(\omega(t + t_i - t_{\min})) - \sin(\omega t))^2 \rangle \quad (6.15)$$

in which t_{\min} is defined to be the value which minimizes E_p . The above expression may be expanded to give

$$E_p = \frac{A^2}{m} \sum_{i=1}^m [\langle \sin^2(\omega t + \phi_i) \rangle - \langle 2 \sin(\omega t + \phi_i) \sin(\omega t) \rangle + \langle \sin^2(\omega t) \rangle] \quad (6.16)$$

$$= \frac{A^2}{m} \sum_{i=1}^m (1 - \cos(\omega(t_{\min} - t_i))) \quad (6.17)$$

The value of t_{\min} may be obtained by differentiating the above and equating to zero, giving an error minimizing delay of simply $t_{\min} = \overline{t_i}$. The expression for E_p may now be expanded, resulting in

$$E_p = \frac{1}{m} \sum_{i=1}^m \left[\frac{\omega^2 (t_i - \overline{t_0})^2}{2!} - \frac{\omega^4 (t_i - \overline{t_0})^4}{4!} + \dots \right] \quad (6.18)$$

which for small values of time skew may be approximated by

$$E_p \approx \frac{A^2 \omega^2 \sigma_t^2}{2} \quad (6.19)$$

The amount of error power resulting from gain mismatch, offset mismatch and sampling skew is summarized in Table 6.2. To determine the overall signal-to-noise ratio of a converter array, the error powers listed in the second column need only be added together, along with the average error power due to individual nonlinearities, and compared with the original signal power ($V_{pp}^2/8$). The relative sensitivities of the error power to different non-idealities are shown in the third column. In this comparison, the indicated variance of each of the three parameters, is that which produces a degradation of S/N ratio to 40 dB when all other noise sources are zero. Note that the overall error power is least sensitive to gain variance and most sensitive to phase skew and jitter. As a point of reference, however, the gain matching requirement is actually twice as severe as the most significant bit component matching requirement of a single converter, for a nonlinearity limited S/N ratio of 40 dB. This implies that arrays of converters which have a specific gain determining element, such as R-2R ladders, will have a more severe component matching requirement when used in an array than when used individually (for a specified S/N ratio). Similarly, multiple flash circuits [108] may suffer from a reduced S/N ratio, as "bow" errors will modify the effective converter gains and offsets. All types of converter arrays, however, will display reduced performance levels in the presence of random component variations. The S/N ratio performance of 7-bit flash and weighted capacitor converter arrays which employ precision components with Gaussian error distributions, has been simulated via the program listed in Appendix A. The results are summarized in Figs. 6.5-6.7. As is apparent from these figures, both the flash and bipolar input weighted-capacitor converter methods display a degradation in S/N ratio

TABLE 6.2

Noise Source	Sinusoid Error Power	Requirement for Sinusoid S/N = 40 dB
Gain (a) Mismatch	$\frac{V_{p-p}^2 \sigma_a^2}{8}$	$\frac{\sigma_a}{\bar{a}} = 1\%$
Offset (b) Mismatch	σ_b^2	$\frac{\sigma_b}{V_{p-p}} = .71\%$
Phase (t) Skew	$\frac{V_{p-p}^2 \omega^2 \sigma_t^2}{8}$	$\frac{\omega \sigma_t}{2\pi} = .16\%$

$$\sigma_X^2 = \overline{X^2} - \bar{X}^2$$

Magnitude and sensitivity of error power to mismatch in a converter array. Values in table are for an input sinusoid centered about zero.

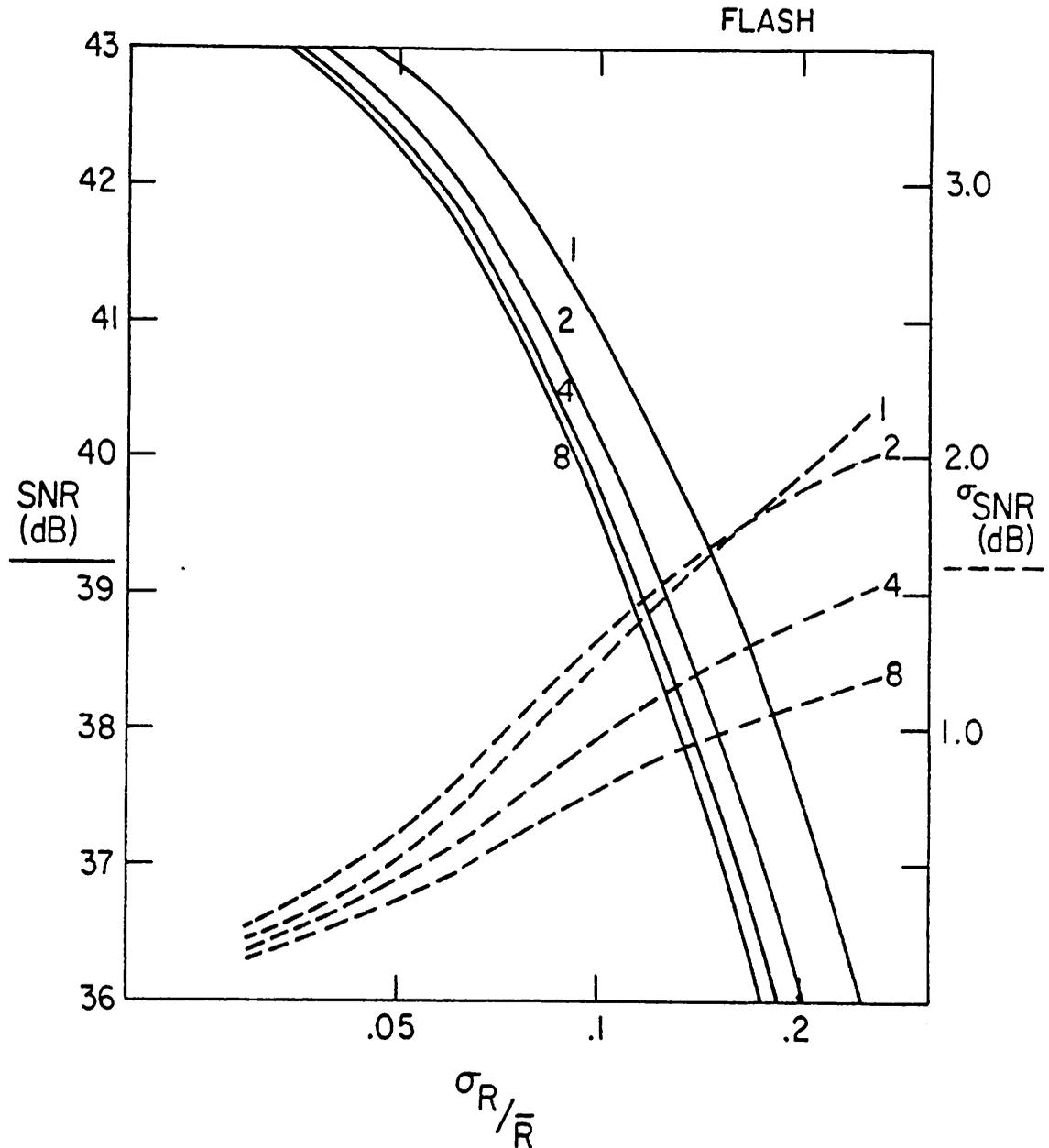


Fig. 6.5 Simulated SNR performance of 100 flash converters when configured as 1,2,4 and 8-way arrays, as a function of resistor mismatch. Ladder resistors are assumed normally distributed about their nominal value.

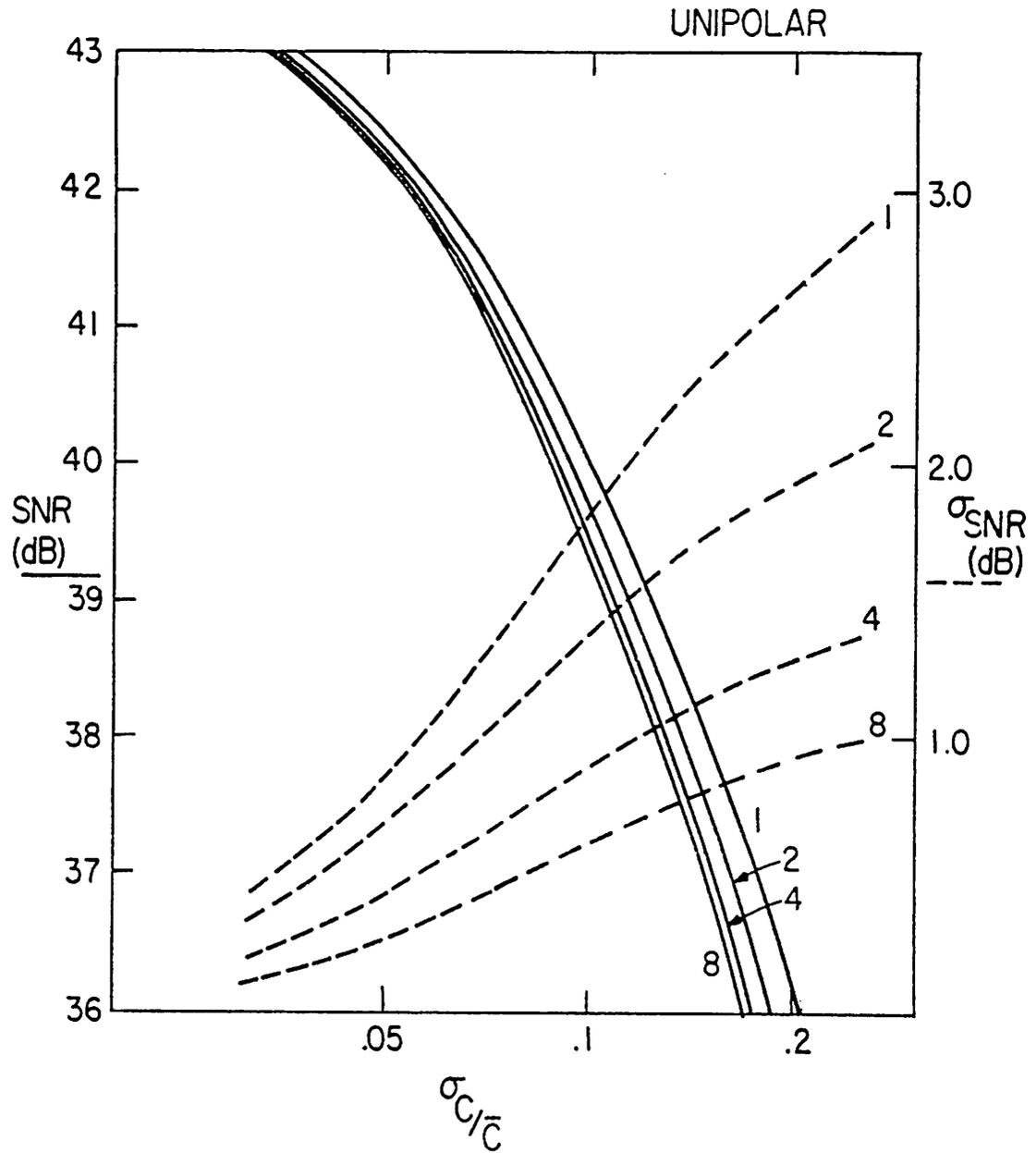


Fig. 6.6 Simulated SNR performance of 100 unipolar input weighted-capacitor A/D converters as a function of array size and unit capacitor matching.

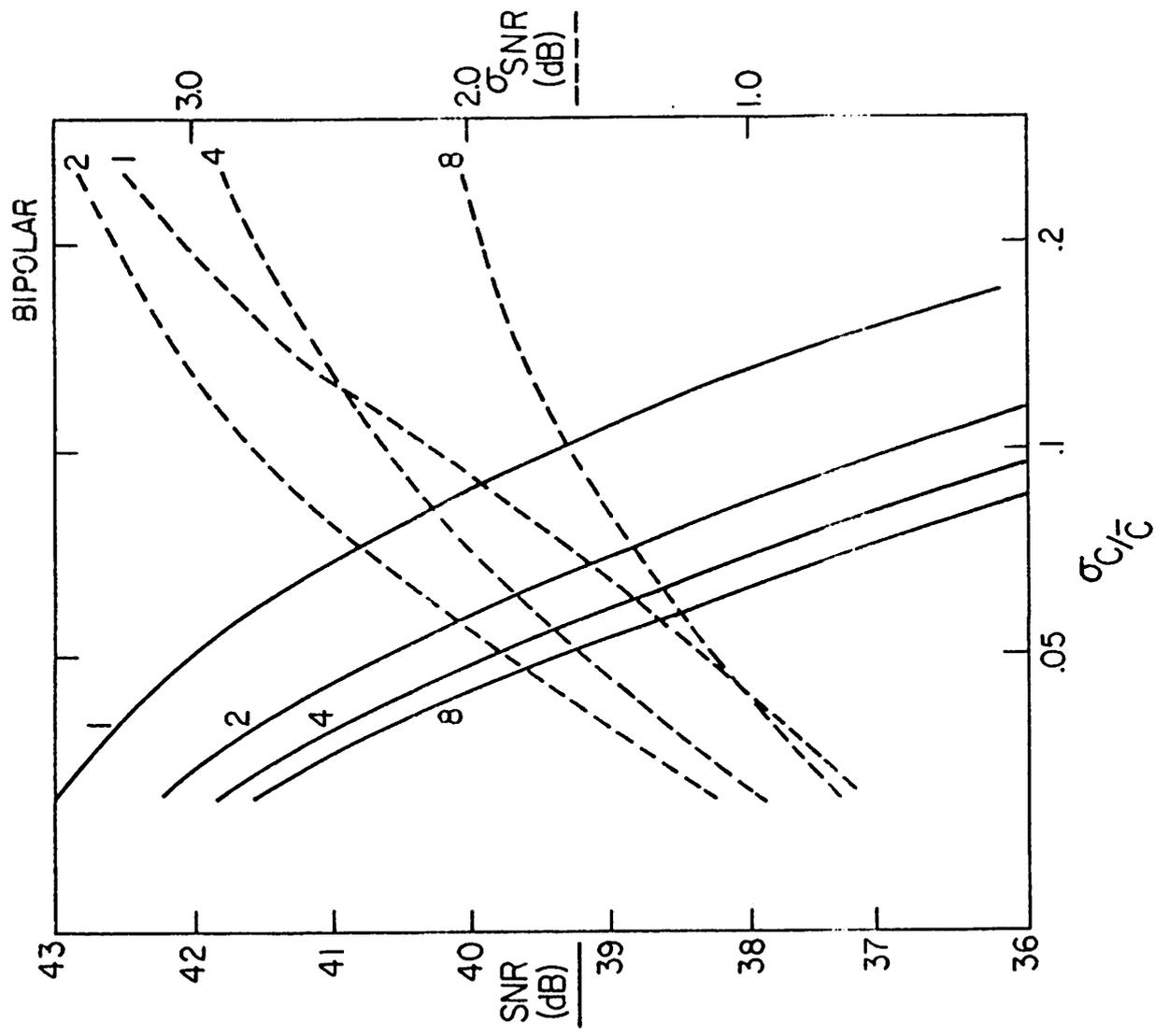


Fig. 6.7 Simulated SNR performance of 100 bipolar input weighted-capacitor A/D converters as a function of array size and unit capacitor mismatch.

The $k=0$ case in Eq. (6.20) corresponds to the original signal component, while the remaining terms are a form of aliasing distortion. This distortion is demonstrated for the case of a sinusoidal input signal and a 4-way array in Fig. 6.8b, as compared with the ideal case in Fig. 6.8a. As is evident from the figure, sidebands develop about fractions of the sampling rate, which are identical in spectral content to the original input signal. The relative magnitude of each sideband, which except for the $k=m/2$ case exist in pairs, is determined by the appropriate order of discrete transform as defined above. Note that for nonzero DC levels, gain errors will produce noise at each multiples of f_s/m .

Unlike the gain variations discussed above, a mismatch of converter offsets will produce a constant noise, which is independent of the input frequency spectrum. The reconstructed noise waveform for an offset mismatched converter array that is otherwise ideal, may be given by

$$\hat{x}(t) = \sum_{k=-\infty}^{\infty} \frac{1}{m} \sum_{i=1}^m b_i h(t - (mk + i - 1)T) \quad (6.22)$$

where b_i is the offset of the i th converter, T is the sampling period and $h(t)$ is the impulse response of the reconstruction filter. The above noise expression has a spectral distribution given by

$$Z\{\hat{x}(t)\} = \int_{-\infty}^{\infty} \hat{x}(t) e^{-j\omega t} dt \quad (6.23)$$

$$= \int_{-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \frac{1}{m} \sum_{i=1}^m b_i h(t - (mk + i - 1)T) e^{-j\omega t} dt \quad (6.24)$$

which may be expressed as

$$F\{\hat{x}(t)\} = \sum_{k=-\infty}^{\infty} e^{-j\omega T m k} \frac{1}{m} \sum_{i=1}^m b_i e^{-j\omega(i-1)T} H(\omega) \quad (6.25)$$

performance of typically 1 to 2 dB when used in arrays. Note that the unipolar input converter, however, maintains a tight performance distribution over all of the array sizes simulated. The reason for the improved performance of this converter over the bipolar input type, is principally due to the smaller effect of MSB component mismatch on best fit gain and offset. Note that even for the flash and bipolar input converters, however, a constant performance level may be obtained in large arrays by doubling the component matching requirement over that required for a single converter. This is equivalent (in matching requirement) to adding 1 bit of resolution to each of the converters.

6.2.2 Spectral Distribution of Noise and Distortion

Of equal importance to the error power magnitude is the noise and distortion spectral distribution. Analysis of the power spectrum resulting from a gain mismatched array is similar to that performed by Messerschmitt for the case of induced gain errors by "bit robbing" in digital channel banks [109]. For an m-way array with gain mismatch, an input signal of power spectrum $S_x(\omega)$ will result in a reconstructed output power spectrum (assuming impulse sampling) given by

$$S_{\hat{x}}(\omega) = |H(\omega)f_s|^2 \sum_{k=-\infty}^{\infty} |G_k|^2 S_x(\omega - k2\pi f_s/m) \quad (6.20)$$

where $H(\omega)$ is the transfer function of the reconstruction filter, f_s is the sampling frequency and G_k is the discrete Fourier transform of gain in the array

$$G_k = \frac{1}{m} \sum_{i=1}^m a_i e^{-j2\pi k(i-1)/m} \quad (6.21)$$

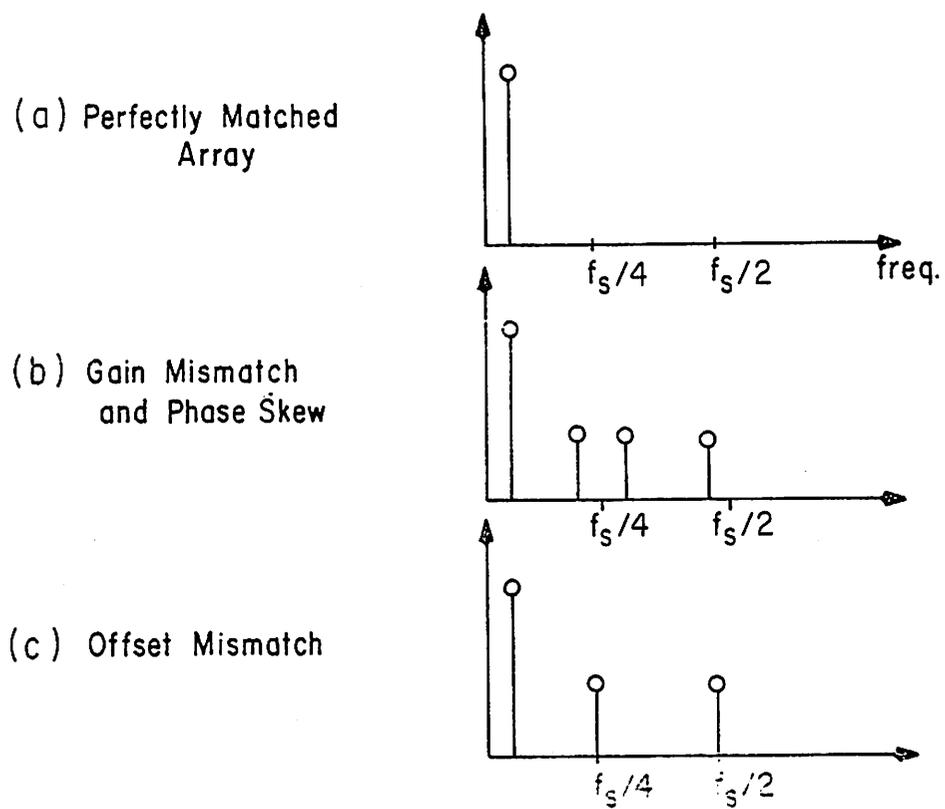


Fig. 6.8 Spectrum of a reconstructed sinusoid for a 4-way converter array.

This reduces to

$$F\{\hat{X}(t)\} = \sum_{k=-\infty}^{\infty} \delta(\omega - 2\pi k/mT) \frac{1}{m} \sum_{i=1}^m b_i e^{-j2(i-1)k/m} H(\omega) \quad (6.26)$$

which leads to an output noise power spectrum of

$$N_{\hat{X}}(\omega) = |H(\omega)f_s|^2 \sum_{k=-\infty}^{\infty} |B_k|^2 \delta(\omega - kf_s/m) \quad (6.27)$$

where B_k is the discrete Fourier transform of best fit offsets in the array. This noise is illustrated in Fig. 6.8c for a 4-way array, along with the original signal spectrum. As is evident from the above analysis, offset related noise appears at multiples of f_s/m , just as gain errors do for nonzero DC input levels.

The noise due to sampling time skew in a converter array is most easily analyzed with respect to an input sinusoid, rather than for an arbitrary distribution as was done in the gain error case. If the i th converter is assumed to exhibit a sampling time error of δt_i and is otherwise assumed to be ideal, the sampled value may be given by

$$\hat{x}_i(t) = A \sin(\xi(t + \delta t_i)) \quad (6.28)$$

This may be expanded, to give

$$\begin{aligned} \hat{x}_i(t) = A[& \sin(\xi t) + \xi \delta t \cos(\xi t) \\ & - \xi^2 \delta t_i^2 \frac{\sin(\xi t)}{2!} - \xi^3 \delta t_i^3 \frac{\cos(\xi t)}{3!} + \dots] \end{aligned} \quad (6.29)$$

which allows the converter error

$$\epsilon_i(t) = A \sin(\xi(t + \delta t_i)) - A \sin(\xi t) \quad (6.30)$$

to be approximated by.

$$\epsilon_i(t) = A\xi\delta t_i \cos(\xi t) \quad (6.31)$$

for small values of δt_i . For an array of interleaved converters, this results in a reconstructed output error waveform given by

$$\hat{x}(t) = \sum_{k=-\infty}^{\infty} \frac{1}{m} h(t - (mk + i - 1)T)\xi\delta t_i \cos(\xi t) \quad (6.32)$$

Note that this function is identical to that generated from a gain mismatched array for an input signal of $\cos(\xi t)$ and with effective gains of $a_i = \xi\delta t_i$. Thus the output distortion power spectrum for the phase skewed situation is approximately the same as the total output power spectrum for this latter case. Thus

$$N_x(\omega) \approx |AH(\omega)f_s|^2 \sum_{k=-\infty}^{\infty} |\phi_k|^2 \Pi\left(\frac{\omega - kf_s/m}{2\xi}\right) \quad (6.33)$$

where ϕ_k is the discrete transform of $\xi\delta t_i$ and Π is given by

$$\Pi(x) = \frac{1}{2} \delta(x + 1/2) + \frac{1}{2} \delta(x - 1/2) \quad (6.34)$$

Note that when δt_i is defined relative to the average sampling period, the $k=0$ term in the above is identically zero.

As is apparent from this analysis, and Fig. 6.8b, output distortion due to phase skew is similar in spectral content to that of converter gain error. The principal difference is that the magnitude of the distortion sidelobes in the phase skew case are frequency dependent, while in the gain error case they are not.

The previous spectral analysis has not specifically included the effects of quantization and nonlinearity. In converters with asynchronous sources, these effects usually appear as broadband noise and simple or aliased harmonic distortion. This is also the situation with converter arrays if the converter nonlinearities are similar. If the nonlinearities are different, however, a given input level may be encoded differently by different converters, resulting in a "dithered" output signal. This is ultimately equivalent to a gain and offset mismatch, as the different nonlinearities will result in different best fit converter parameters. In real converter arrays, which do not otherwise exhibit a gain or offset mismatch, this may result in an overall S/N degradation of perhaps a few dB in the reconstruction of a full scale input signal (see experimental results). This noise may be objectionable in some systems, particularly during low level or idle input conditions. In video systems it is often desirable, however, as it reduces the appearance of sharp quantization boundaries in picture areas where the luminance changes slowly []. If necessary, the dithering at a given voltage may be eliminated by altering the effective converter offsets (by either analog or digital means) so as to place a given voltage within a common quantization level.

Of final note in the overall error analysis is the effect of the previously considered noise and distortion sources on the S/N ratio of a bandlimited input signal. As significant portions of the noise and distortion products are clustered at, or near the Nyquist rate, the actual inband S/N ratio of an array, may be larger than what the power figures of Table 6.2 may indicate. More specifically, for a system whose signal bandwidth is limited to f_s/k ($k \geq 2$), up to $k/2$ badly

matched converters can be used in an array with no degradation of in-band signal-to-noise ratio.

6.3 A Monolithic Time-Interleaved Converter Array

6.3.1 Chip Description

To demonstrate the characteristics of this array-technique, a 4-channel CMOS test-chip has been fabricated in a 10 micron metal-gate CMOS process [11,112]. This chip consists of 4 time-interleaved 7-bit weighted-capacitor bipolar input A/D converters, associated timing and control logic and an implanted MOS voltage reference. Also included on this chip is a 4-way input multiplexor, which enables some, or all of the individual A/D converters to be connected with a specific input line. This allows each of the converters to be operated as part of an array (common inputs), or independently. A block diagram of this test chip is shown in Fig. 6.9. In this figure, the converters are seen to exist as a matrix of bit cells, where each row corresponds to a separate A/D converter, and where each column represents the bit-cells which are active in a given time-state. For example, during the first time-state, the first A/D converter is sampling, the second A/D converter is determining bit 6 of its conversion sequence and so on. A total of 8 time-states are required for a complete 7-bit conversion, which, because there are 4 converters per chip, results in an effective throughput of one conversion every two clock cycles. Note that this throughput may be doubled by interleaving two chips for a total of an 8-way converter-array.

This entire chip is defined in terms of general purpose cells, which were placed and interconnected via an automated layout program. Although

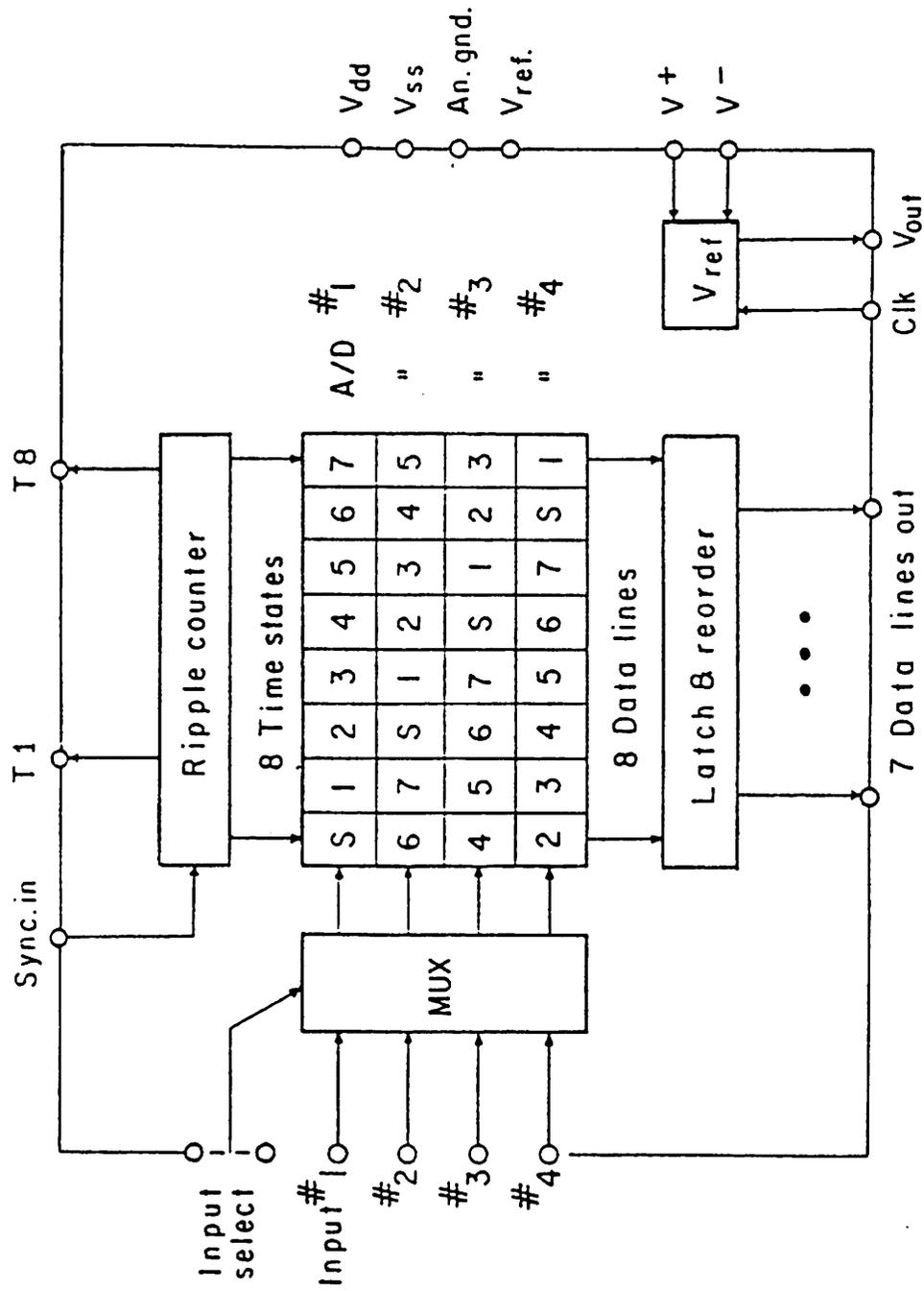


Fig. 6.9 Block diagram of converter array chip.

this approach resulted in a considerable increase in die size, it greatly simplified the required layout effort and provided a library cell-base from which other converter configurations can be designed. In theory, other converters may be realized by simply modifying the cell interconnect list, and rerunning the layout program.

An example of one of the cells used in this chip, a bit slice of a successive approximation register, is shown in cell form in Fig. 6.10. As is evident from the figure, the cells are defined as constant height, variable width units, where most of the interconnect occurs along one of the cell sides (bottom of cell in Fig. 6.10a). Typically, the cells exist in back to back rows, where the V_{SS} line can be shared by top and bottom cells (the V_{DD} connection is made through the substrate). Summing node and analog ground lines are shared by cells which are side by side, while some timing and data information may be passed vertically between cell rows where necessary. A layout of the complete chip, in cellular form, is shown in Fig. 6.11. Each of the four cell rows along the middle of the chip is a separate A/D converter, while timing logic exists in the cells at the left and control and buffer circuitry are placed at the right.

A metal-gate CMOS process was chosen for this circuit, for reasons of simplicity and ease of fabrication at Sandia Laboratories. A more advanced process would have been desirable from the standpoint of performance, but was unnecessary for the demonstration of the concept, and not immediately available. A layout of a small inverter in this process which demonstrates most of the minimum design rule dimensions, is shown in Fig. 6.12. As is apparent from this figure, this process required that diffused n+ or p+ guard-bands (which are not self-aligned)

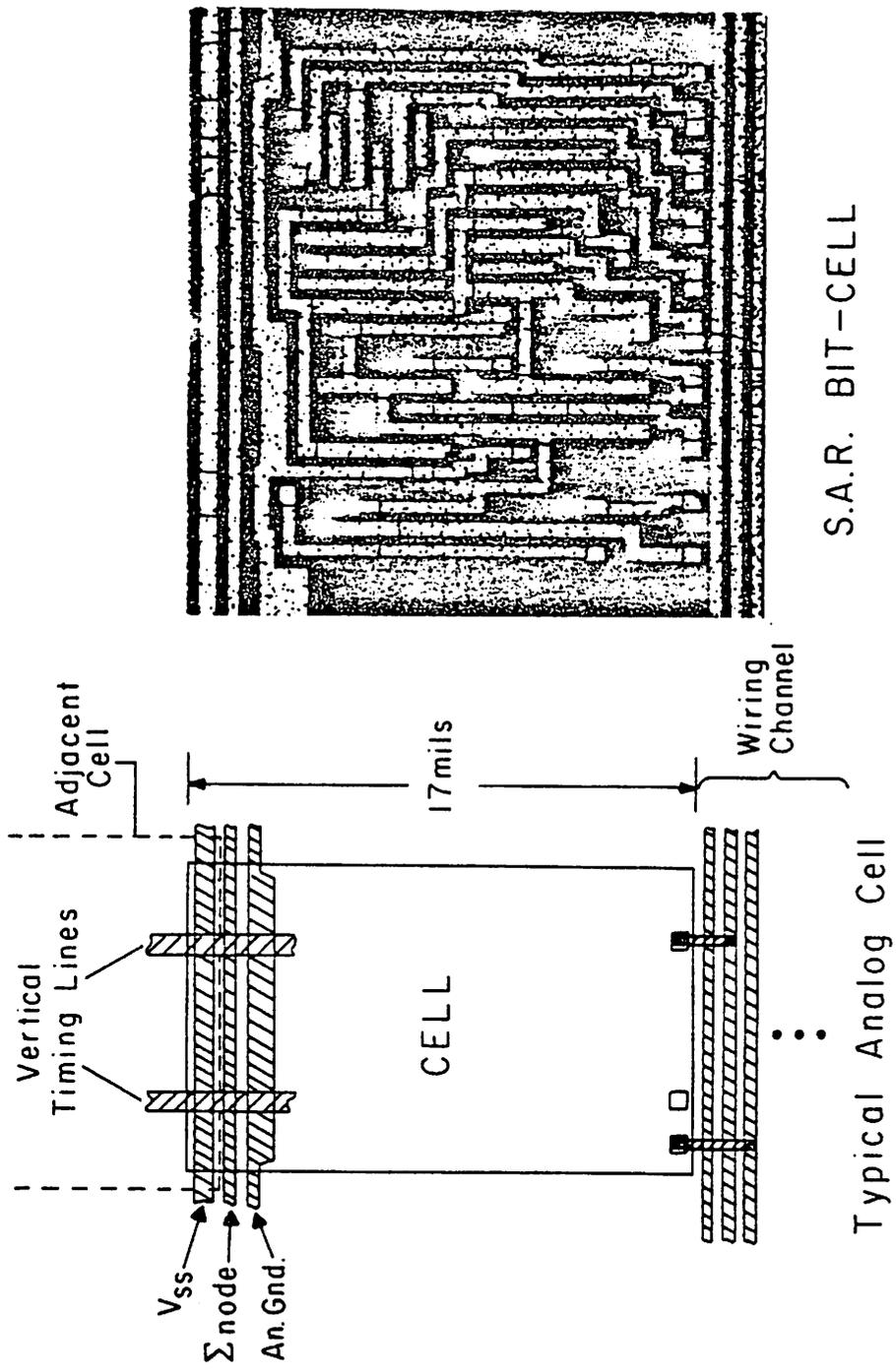


Fig. 6.10 Analog cell configuration.

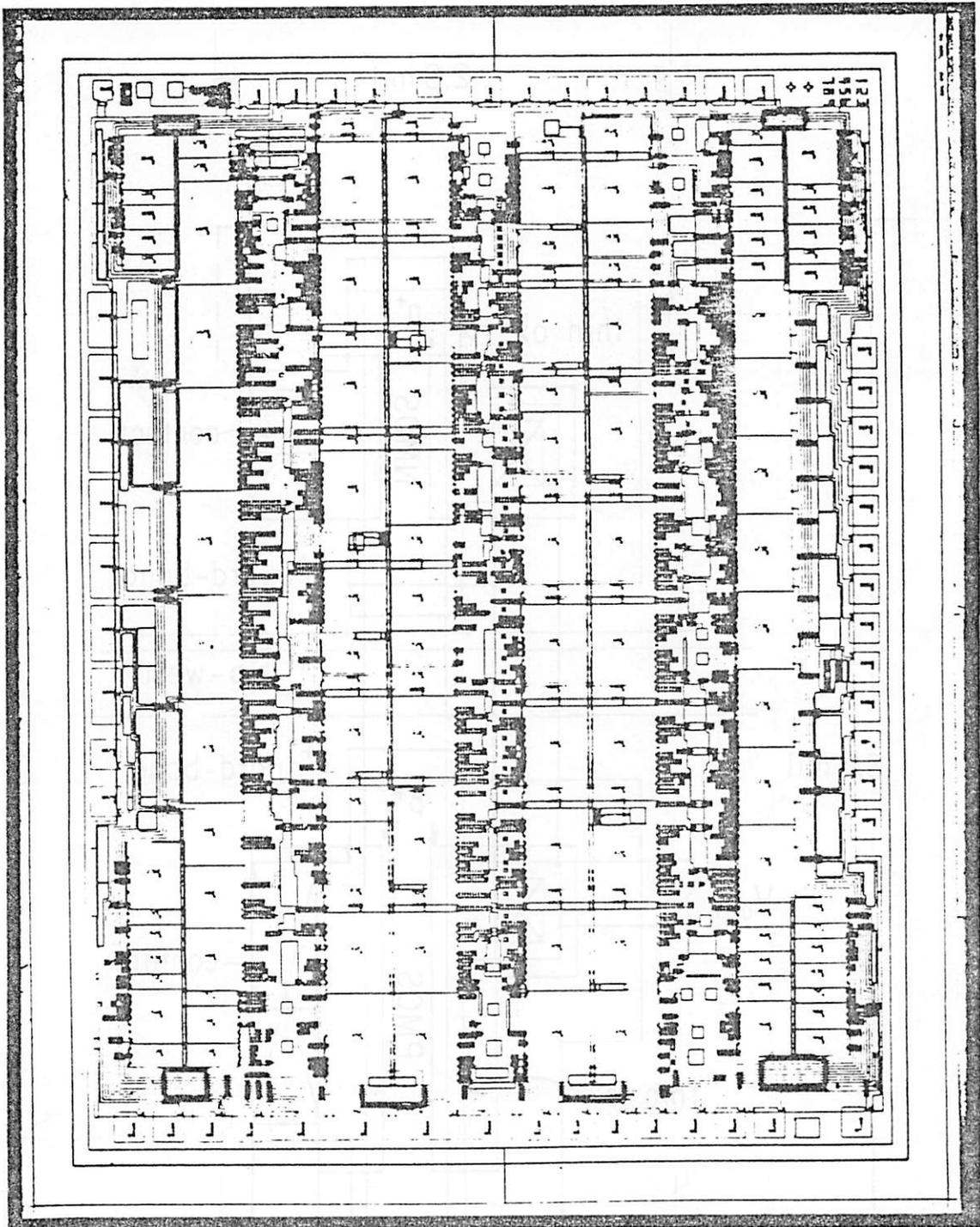


Fig. 6.11 Layout of converter array chip in cell form.

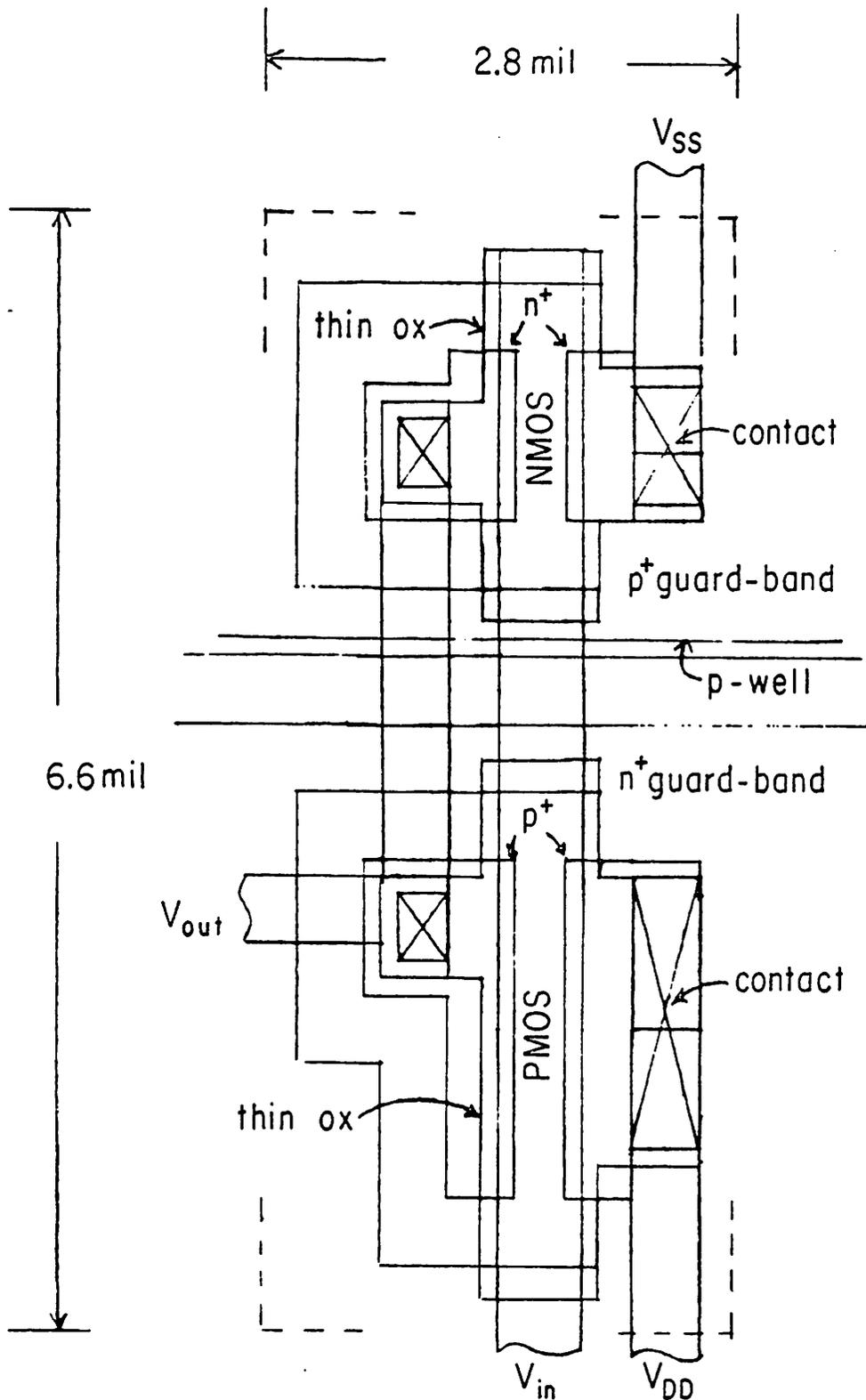


Fig. 6.12 Minimum geometry CMOS inverter in this process with symmetrical pull-up/pull-down. Die size is about 18.5 mil^2 .

be placed around each transistor. Although this ultimately reduced many of the various leakage currents, and certainly helped reduce latch-up (which was never observed), it greatly increased the circuit die size. As may be deduced from Fig. 6.12, a minimum geometry inverter with equal pull-up and pull-down capability occupies about 18.5 mil^2 .

A complete schematic of the SAR bit cell, which was previously shown in cell form, is presented in Fig. 6.13. This cell, or one quite similar to it, is used in each of the 28 bit cells present in the converter array. The state of this cell, is established by the voltage present upon the floating node " \bar{Q} ." This node is pulled high or low at the appropriate bit time, by the complementary devices shown on the right of the cell. For example, during the sampling interval it is pulled down to V_{SS} , while during the hold interval it is brought to V_{DD} . In turn, this node allows the associated precision capacitor to be charged to the appropriate value, by way of a pair of CMOS transmission gates. The state of the cell is read during the associated converters final bit time, by passing its value onto a tri-state data line which is common to all of the cells in a given time-state. Note that this cell is totally dynamic, in that no method for maintaining \bar{Q} at a constant potential in the presence of leakage currents exists.

Several of the bit positions use slight modifications of this cell, for example, the first bit position uses the cell shown in Fig. 6.14. This cell has been modified in order to allow bipolar input operation, as discussed in Chapter 3. The principle difference is that this bit position doesn't require either Sel_1 or \overline{Hold} timing lines. Similarly, the cells for the 6th and 7th bit positions are shown in Figs. 6.15 and 6.16. These cells have been simplified by eliminating the data read

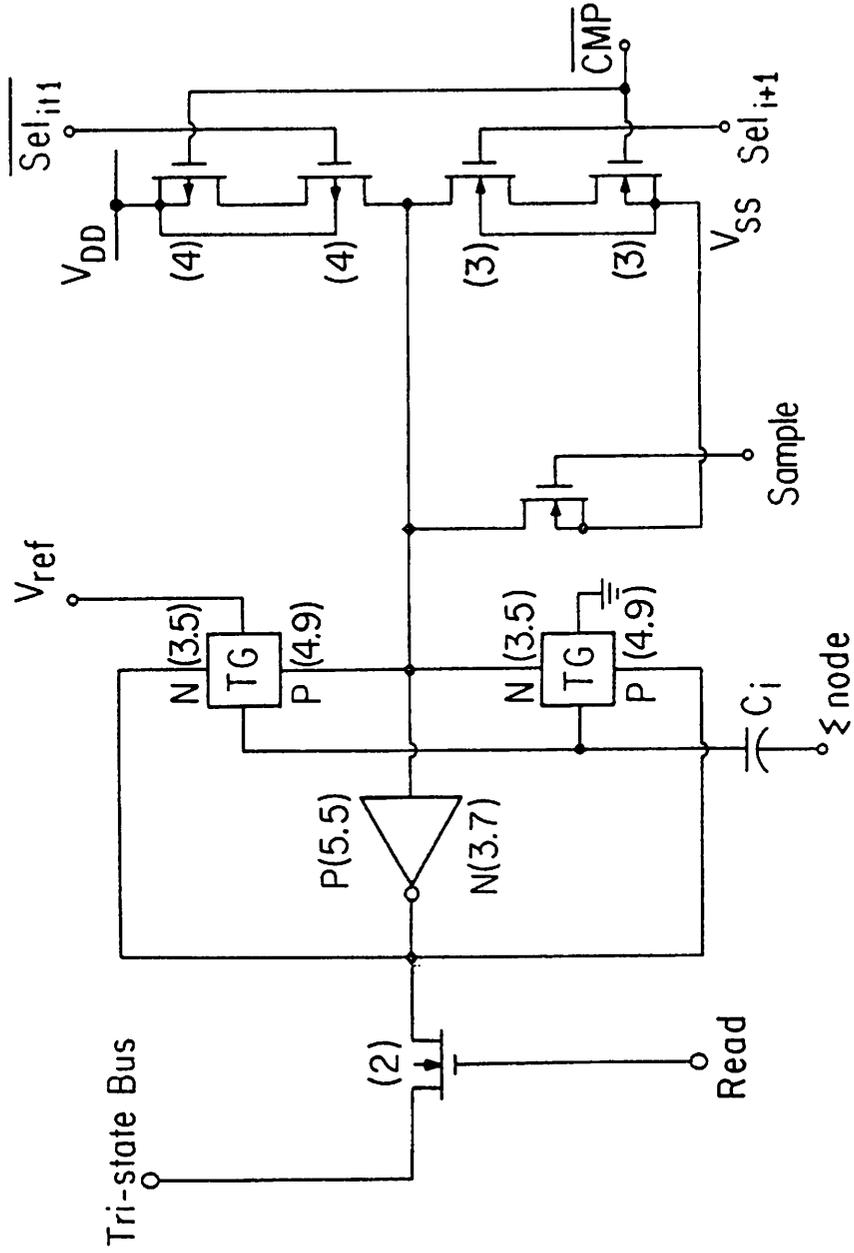


Fig. 6.14 Schematic of cell used in first bit position. Circuit has been modified from that shown previously to allow for bipolar input operation.

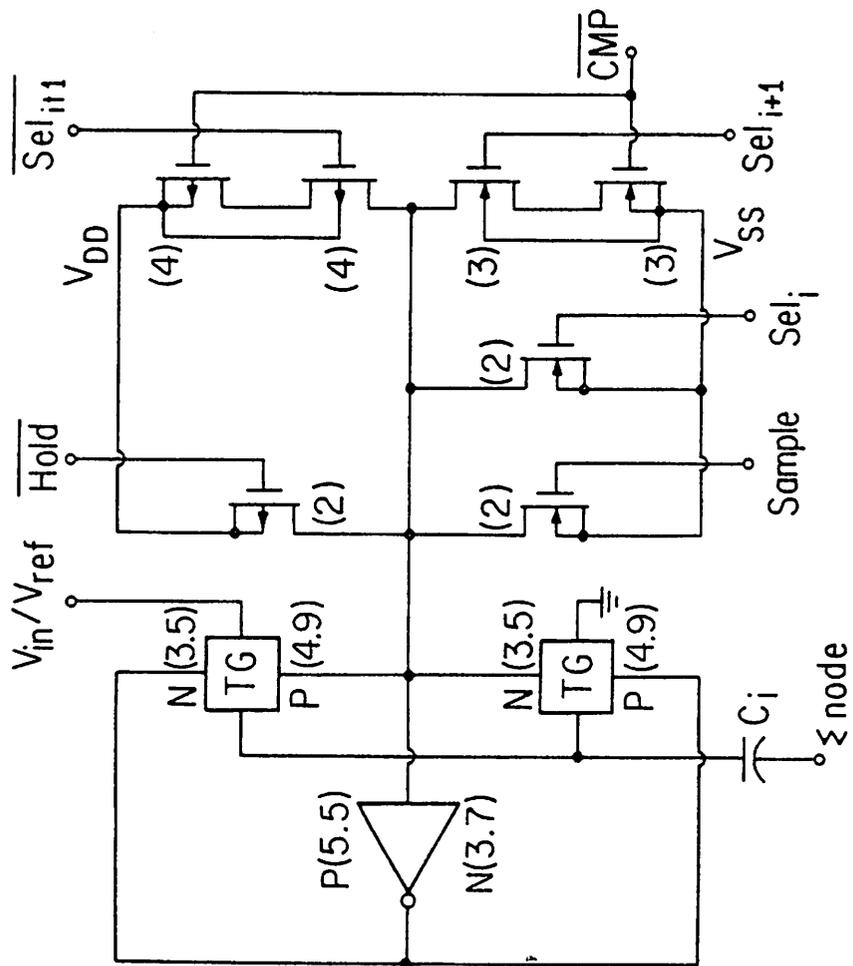


Fig. 6.15 Schematic of cell used in sixth bit position.

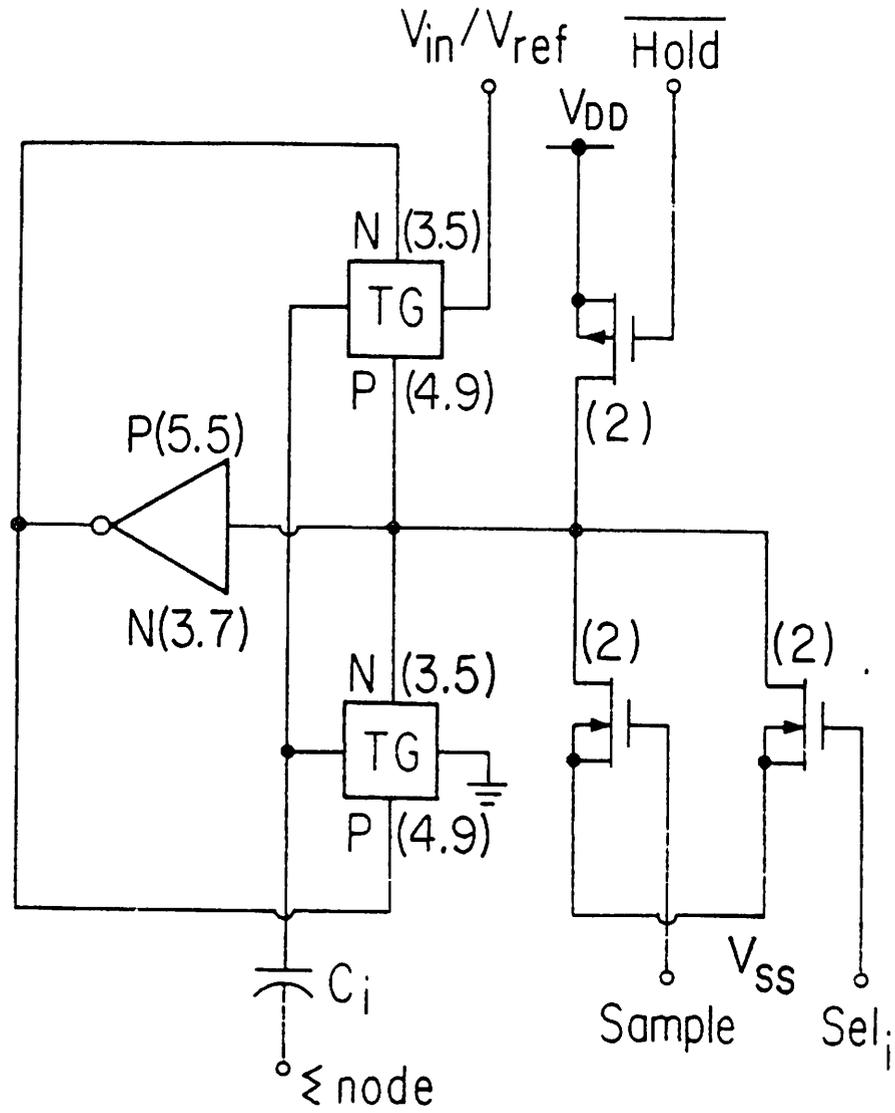


Fig. 6.16 Schematic of cell used in seventh bit position.

circuitry from each, and by eliminating the "next time state" logic from the 7th bit cell. For these bit locations, the state of each cell is transferred directly from the comparator to the output buffer (rather than through the cell), so that the data read circuitry wasn't required. Similarly, as the conversion ends after bit 7, there is no reason to store the result of bit 7 within the bit cell. If this were done, a timing conflict would actually occur as the next sample interval immediately follows the 7th bit time.

The various cells presented above provide the necessary circuitry for 7 of the binary-weighted capacitors. As discussed in Chapter 3, however, one additional capacitor exists which is equal in size to the smallest capacitor in the binary-weighted array. Circuitry for driving this capacitor is shown in Fig. 6.17a. This circuitry connects the capacitor's bottom plate to V_{in} during the sample mode and to ground thereafter while switches exist to line " V_y " which are phased in the reverse of this order. This extra line is used with two additional cells to provide a programmable offset voltage, as shown in Fig. 6.17b. Depending upon the state of two digital offset control pins, the offset of each A/D converter may be adjusted from +0 to +3/4 LSB in 1/4 LSB increments. Although many converters employ circuitry to provide +1/2 LSB of offset, the programmable feature used here was added to allow a simple evaluation of noise resulting from an offset mismatched array. This circuitry would be neither necessary, or particularly desirable in a chip designed for actual use in a system. The complete schematics for these two programmable offset cells are shown in Fig. 6.18. Note that when the programmable input pins are left floating, the first cell connects it's capacitor between V_y and the Σ node (providing 1/2 LSB offset)

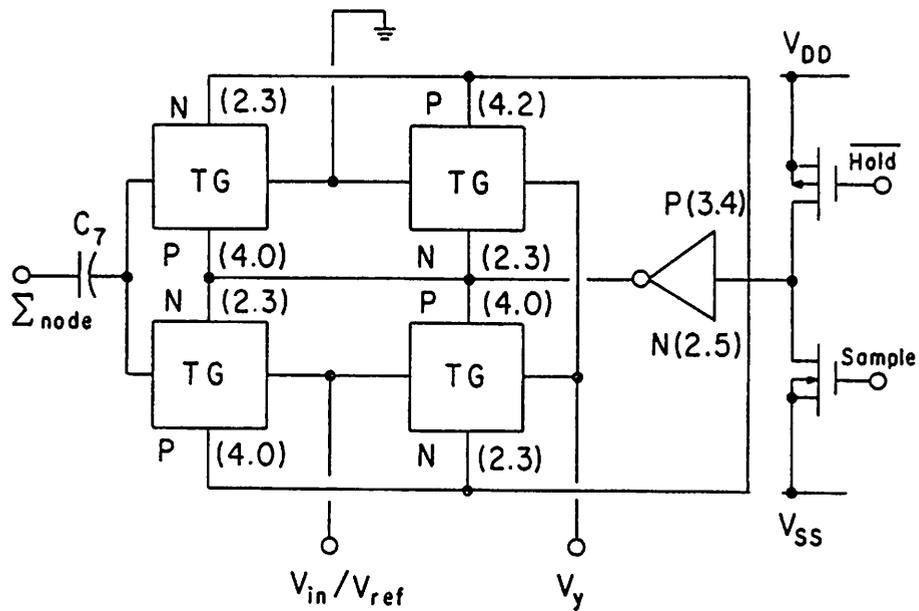


Fig. 6.17a Circuitry for "last" capacitor.

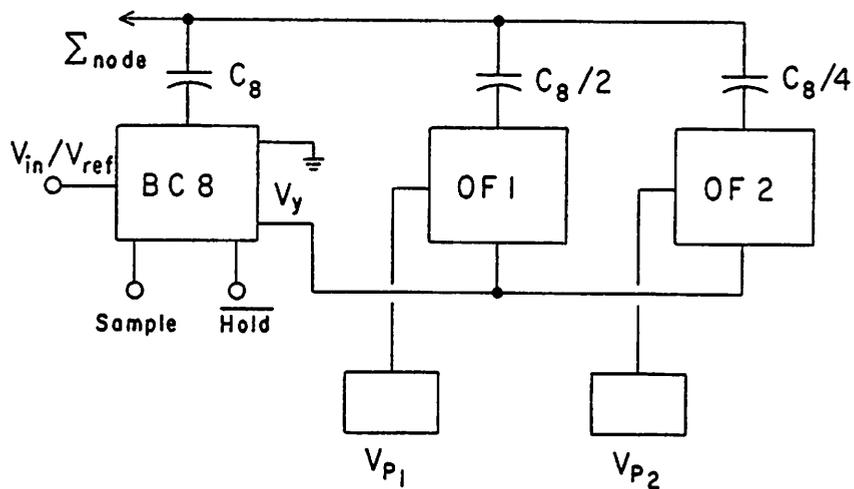


Fig. 6.17b Interconnect of cell shown in (a) with cells shown on next page for control of converter offset.

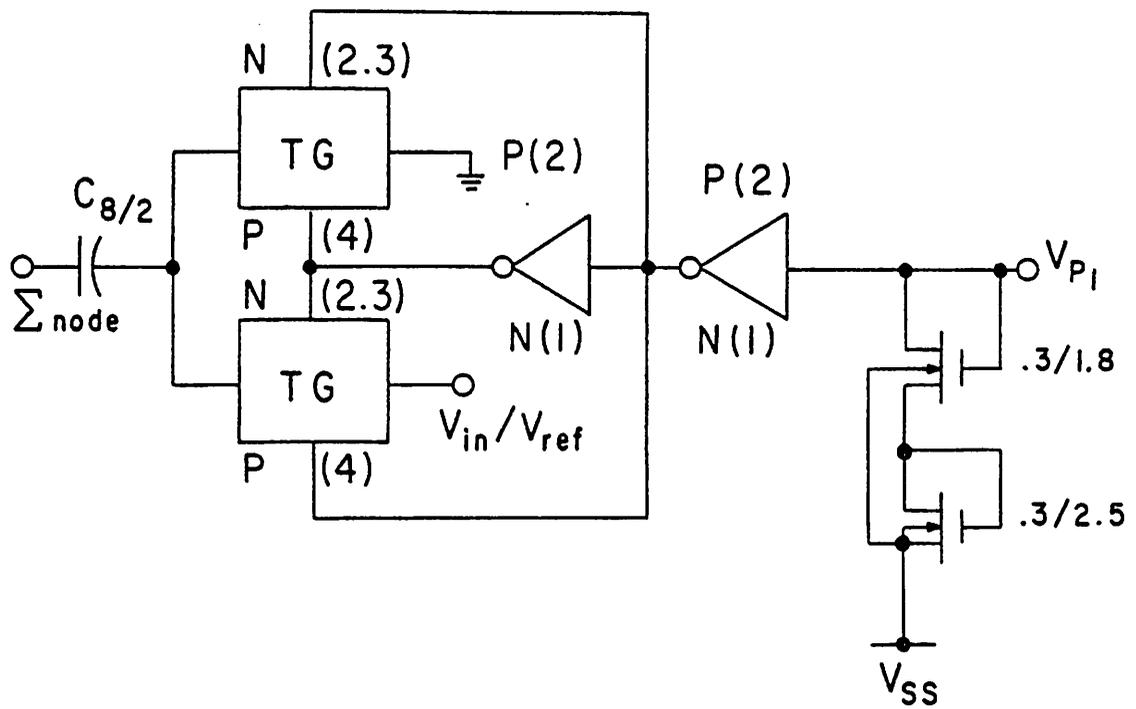


Fig. 6.18a 1/2 LSB offset control cell.

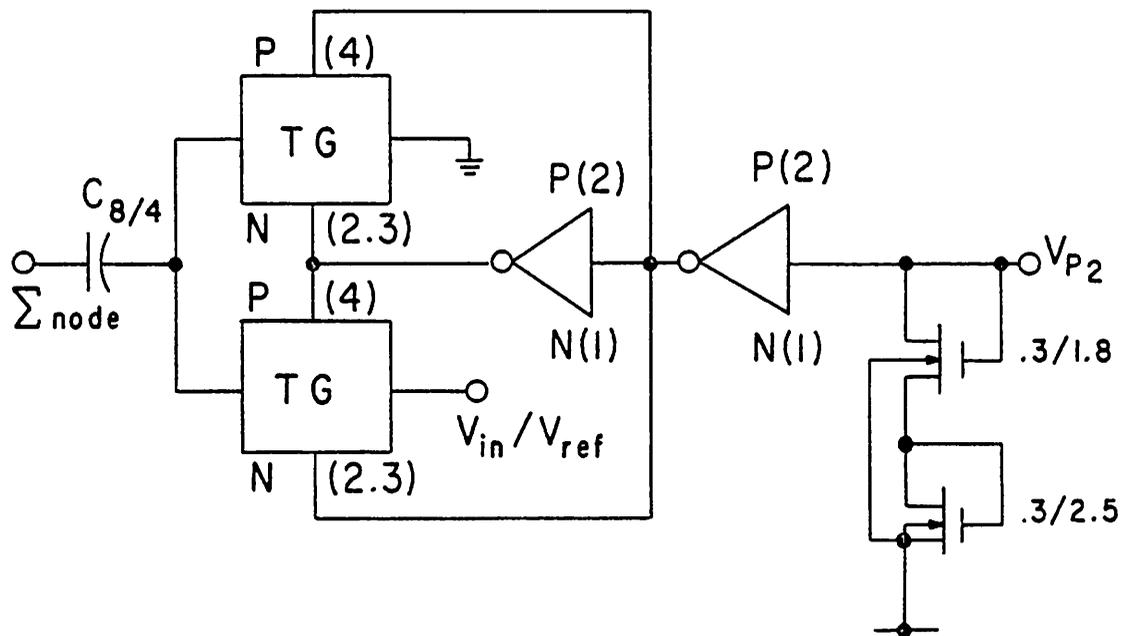


Fig. 6.18b 1/4 LSB offset control cell.

while the second cell does not.

Timing information for each of the various bit cells, is generated by a self starting dynamic ripple converter. This circuit, which is shown schematically in Fig. 6.19, consists of an 8-stage shift register and a 7-input dynamic "OR" gate. This "OR" gate uses a "wired OR" function from the first 7 stages of the shift register to determine whether a "1" or a "0" should be inserted into the input of the shift register. Both Q and \bar{Q} outputs from each stage are buffered by large geometry gates and sent to the bit cells as the lines "Sel_i" and " $\overline{\text{Sel}}_i$," where i indicates the corresponding time state (1-8).

Another cell, the voltage comparator, is shown in Fig. 6.20. This circuit is single-ended and is used with its input node tied directly to the top plate of a capacitor array. During a given converter's "sample" interval, the associated comparator cell is offset cancelled, by shorting the input inverter's output and input together. During subsequent bit times, the input inverter is used as a linear amplifier, where the output reflects the value of the input voltage with respect to its offset point. The output of this inverter is then regenerated and buffered, by the flip-flop which follows it. The large geometry devices of the input inverter allow a relatively high transconductance to be obtained. Using the simple MOS models presented earlier, the effective G_m of the input stage may be shown to be over 2m Mhos with a 10V power supply. Similarly, the shorting switch possesses an "ON" resistance of between 500 and 600 Ω s. As discussed in Chapter 5, where comparators of this type were considered, this results in an effective settling time constant of,

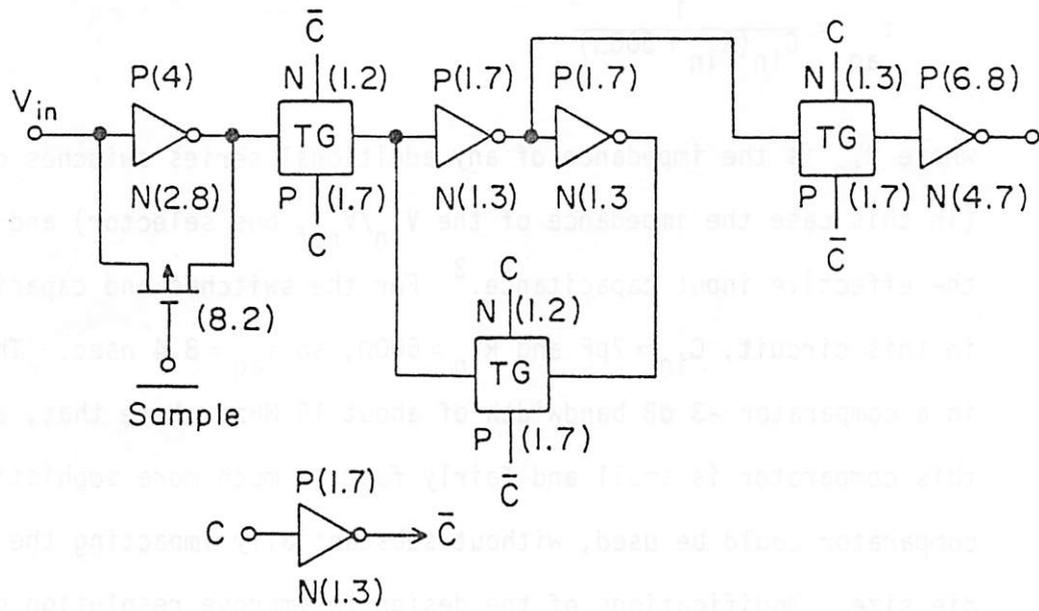


Fig. 6.20a Schematic of voltage comparator.

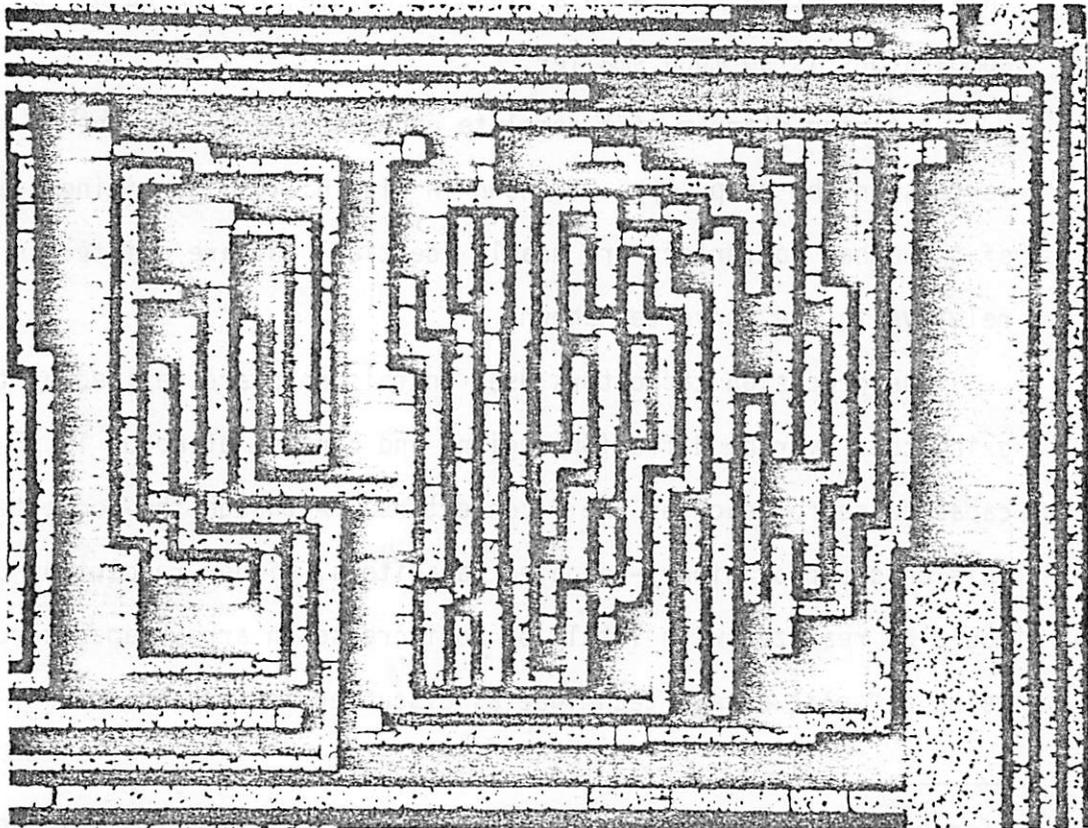


Fig. 6.20b Die photograph of comparator.

$$\tau_{aq} \approx \frac{1}{C_{in}(R_{in} + 500\Omega)} \quad (6.35)$$

where R_{in} is the impedance of any additional series switches or devices (in this case the impedance of the V_{in}/V_{ref} bus selector) and C_{in} is the effective input capacitance.² For the switches and capacitors used in this circuit, $C_{in} \approx 7\text{pF}$ and $R_{in} \approx 600\Omega$, so $\tau_{aq} \approx 8.4$ nsec. This results in a comparator -3 dB bandwidth of about 19 MHz. Note that, although this comparator is small and fairly fast, a much more sophisticated comparator could be used, without substantially impacting the overall die size. Modifications of the design to improve resolution or power-supply rejection, for instance, would be much easier to incorporate in this array method than with a flash technique, as the number of required comparators is so much smaller.

A timing diagram of a complete converter cycle for one of the converters on the chip (#1), is shown in Fig. 6.21. The timing diagrams of the other converters are nearly identical but are shifted in position relative to the converter shown.

The precision capacitors used in this chip are defined by thin oxide cuts over $n+$ diffusion regions and consist of about 10 pF of total capacitance for each of the 4 converters. Although the converters in this chip employ binary-weighted capacitors, higher resolution converters could be realized with little or no increase in array capacitance, by using a "split array" technique as discussed in Chapter 3.

²For bipolar input converters of the type used here, the value of C_{in} is less than the array capacitance, as the MSB capacitor is not charged to V_{in} .

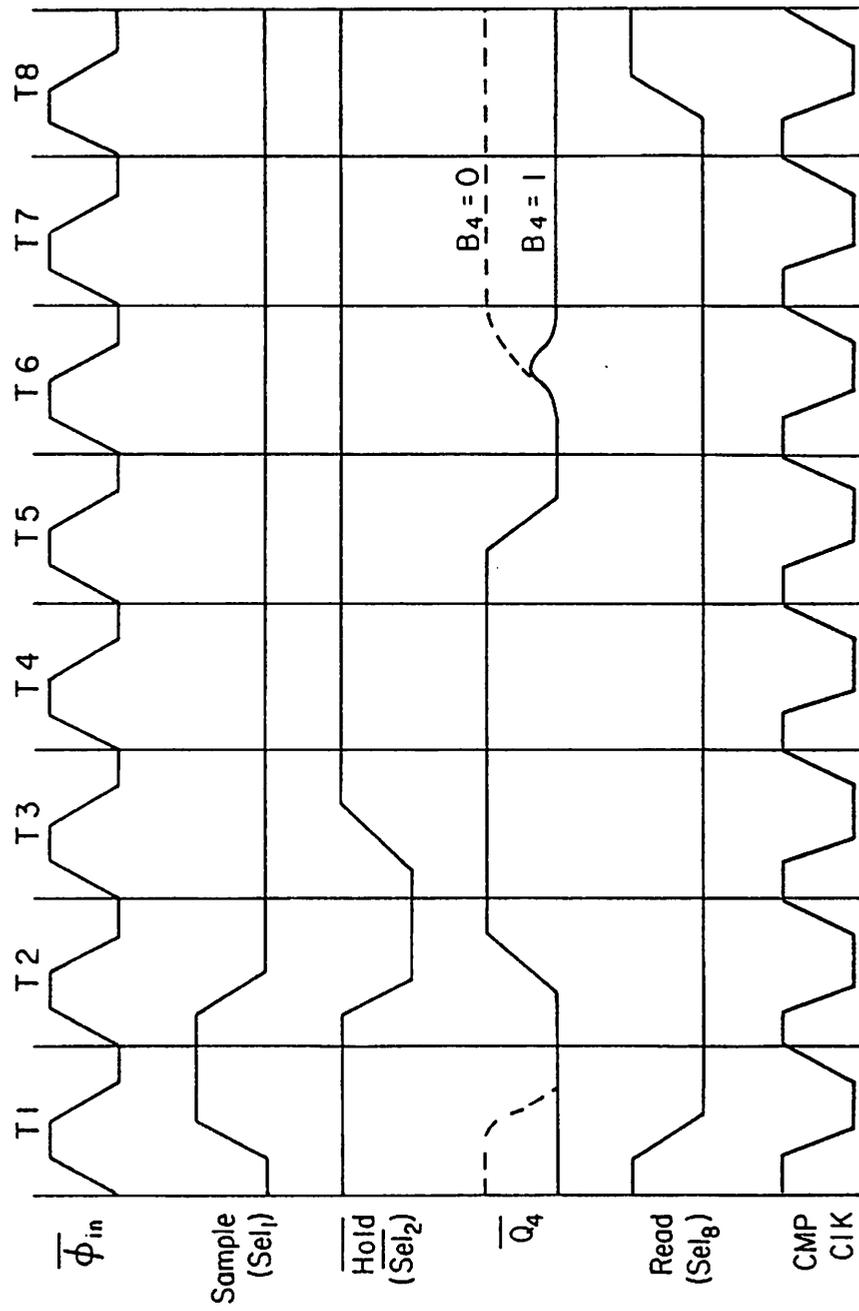


Fig. 6.21 Timing diagram for one conversion cycle of one A/D.

In order to allow easier converter testing, and to demonstrate a concept which would be useful in a commercial device, an input multiplexor was used between the 4 signal input pins and each of the 4 A/D converters. This multiplexor configuration is shown in Fig. 6.22. As is evident from this figure, each of the A/D converters may be tied to either of two signal input lines, based upon the state of an input select control pin (1 per converter). This allows the converters to be easily configured into various array configurations (common inputs), or operated independently. Circuitry for performing this multiplexing function is combined with the " V_{in}/V_{ref} " bus logic, as shown in Fig. 6.23. In this circuitry, ratioed CMOS logic is used to provide a rapid output of V_{in} during the sample interval, with a slow recovery to V_{ref} thereafter. This slow recovery is not detrimental to conversion time or accuracy, as no bit cells use the " V_{in}/V_{ref} " line until two time states after sampling. The two ratioed "AND" gates used in this circuit, are shown in Fig. 6.24.

The on-chip voltage reference uses a closed loop CMOS amplifier which has an implant-induced offset voltage of approximately .8V. A schematic of this amplifier is shown in Fig. 6.25, while measured amplifier characteristics are presented in Table 6.3. In order to provide improved transient settling response and an increase in reference voltage, as well as to allow a controllable amount of "dither" to be added to the quantized signals, this amplifier is used in the clocked circuit configuration shown in Fig. 6.26a. Operation of this circuit under DC loading is demonstrated by the oscilloscope photograph in Fig. 6.26b.

A die photograph of the complete converter array chip, is shown in Fig. 6.27.

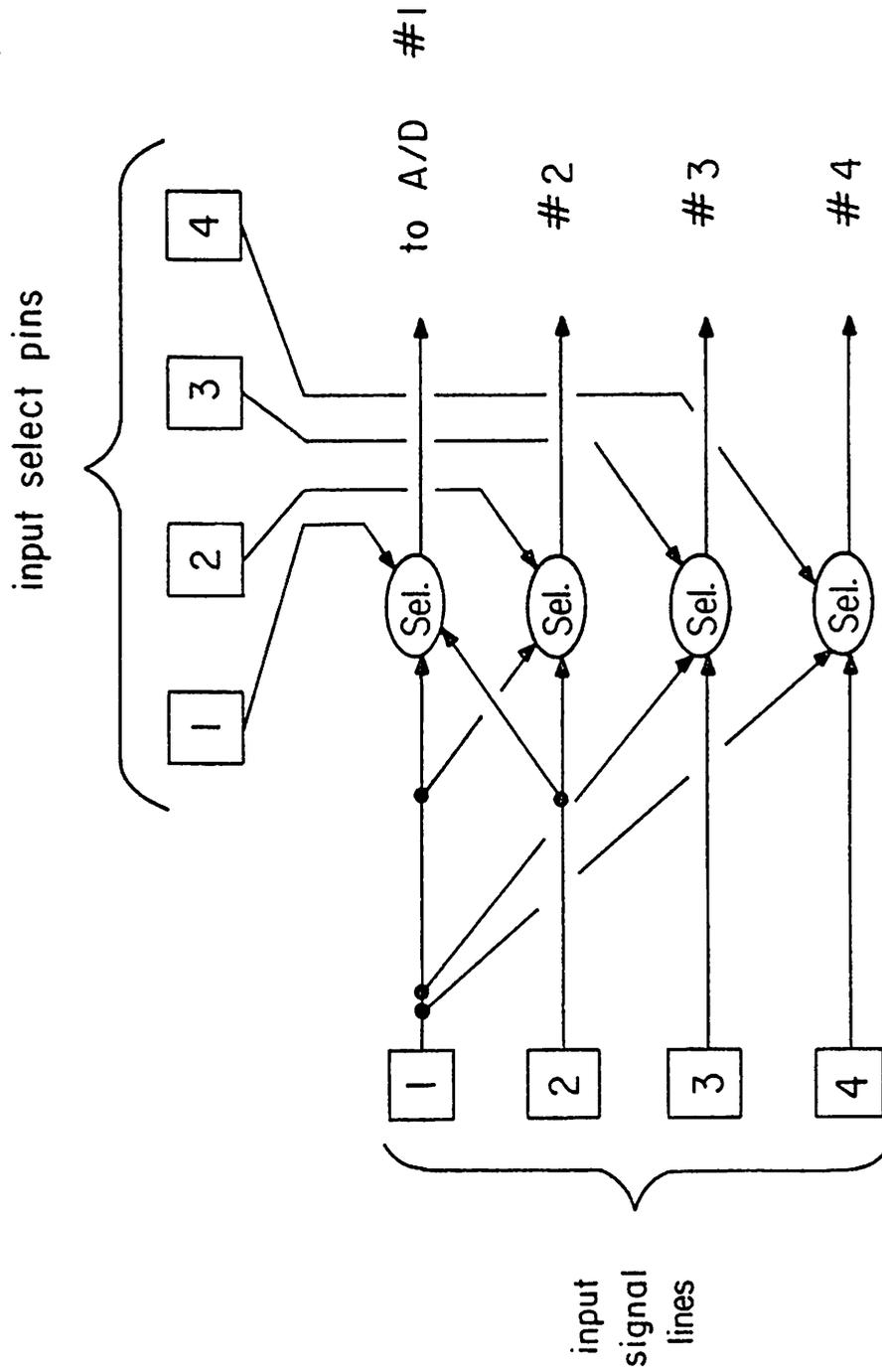


Fig. 6.22 Input multiplexor configuration. Each A/D converter may be programmed to select either of two input lines.

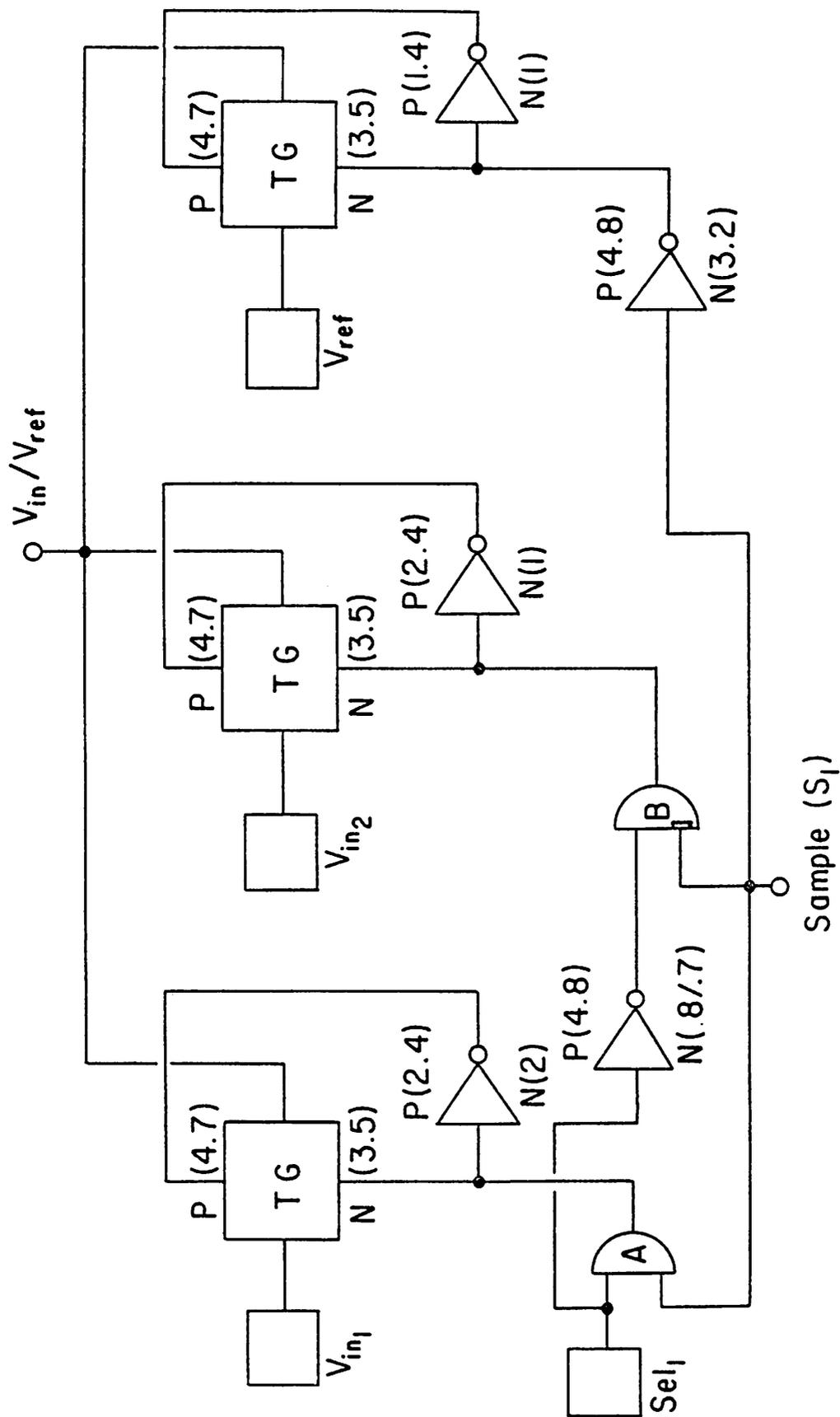


Fig. 6.23 V_{in}/V_{ref} bus logic for one converter. Circuit is combined with multiplexor logic and uses some ratioed logic for increased conversion rate (e.g. fast fall, slow rise where fall time is speed limiting).

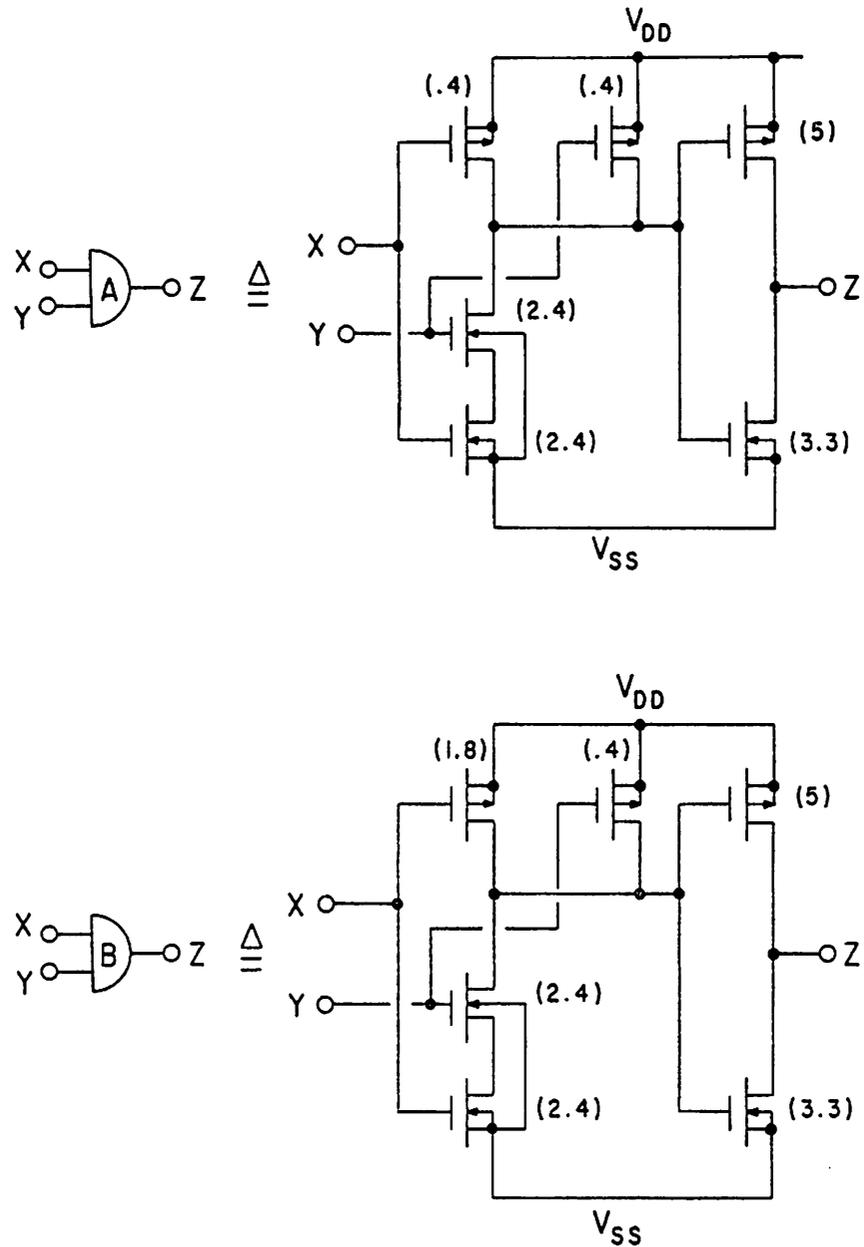


Fig. 6.24 Ratioed "AND" gates of previous figure. Skewing of W/L ratios from normal values allows faster transmission of V_{in} to bus upon receipt of "sample" signal.

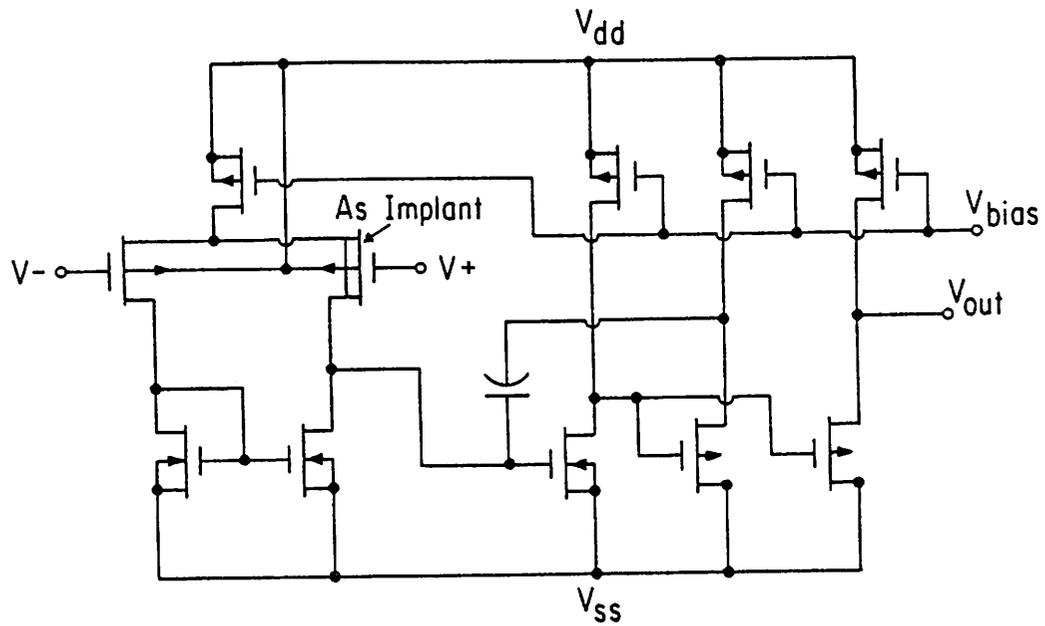


Fig. 6.25 CMOS op-amp with implant shifted offset voltage is used as a voltage reference.

TABLE 6.3

Gain	10,000 V/V
Bandwidth	1.5 MHz
Slew Rate	> 1 V/ μ sec
CMRR	60 dB
Power Dis.	1.5 mW
Die Size	300 mil ²

Measured op-amp characteristics.

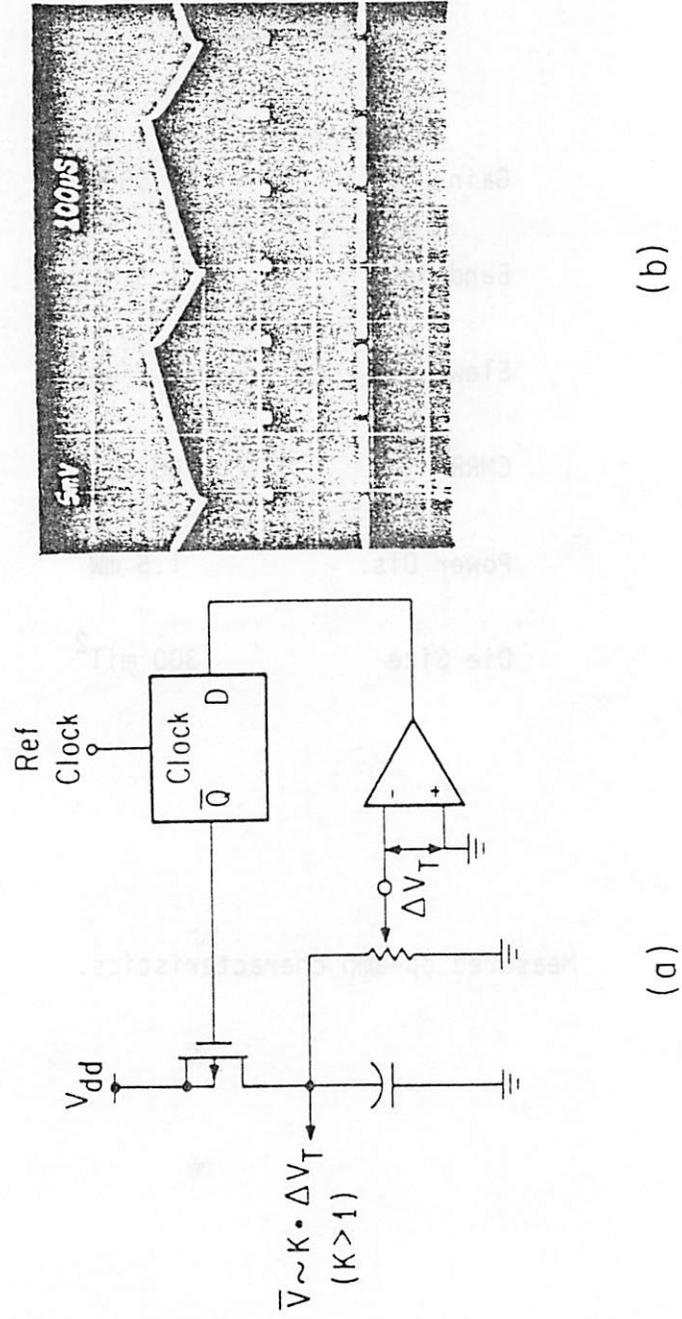


Fig. 6.26 Clocked reference configuration enables a controllable "dither" to be added to the quantized signal. (a) Circuit schematic. (b) Typical operation under a small DC load. Output waveform is at top, clock is at bottom.

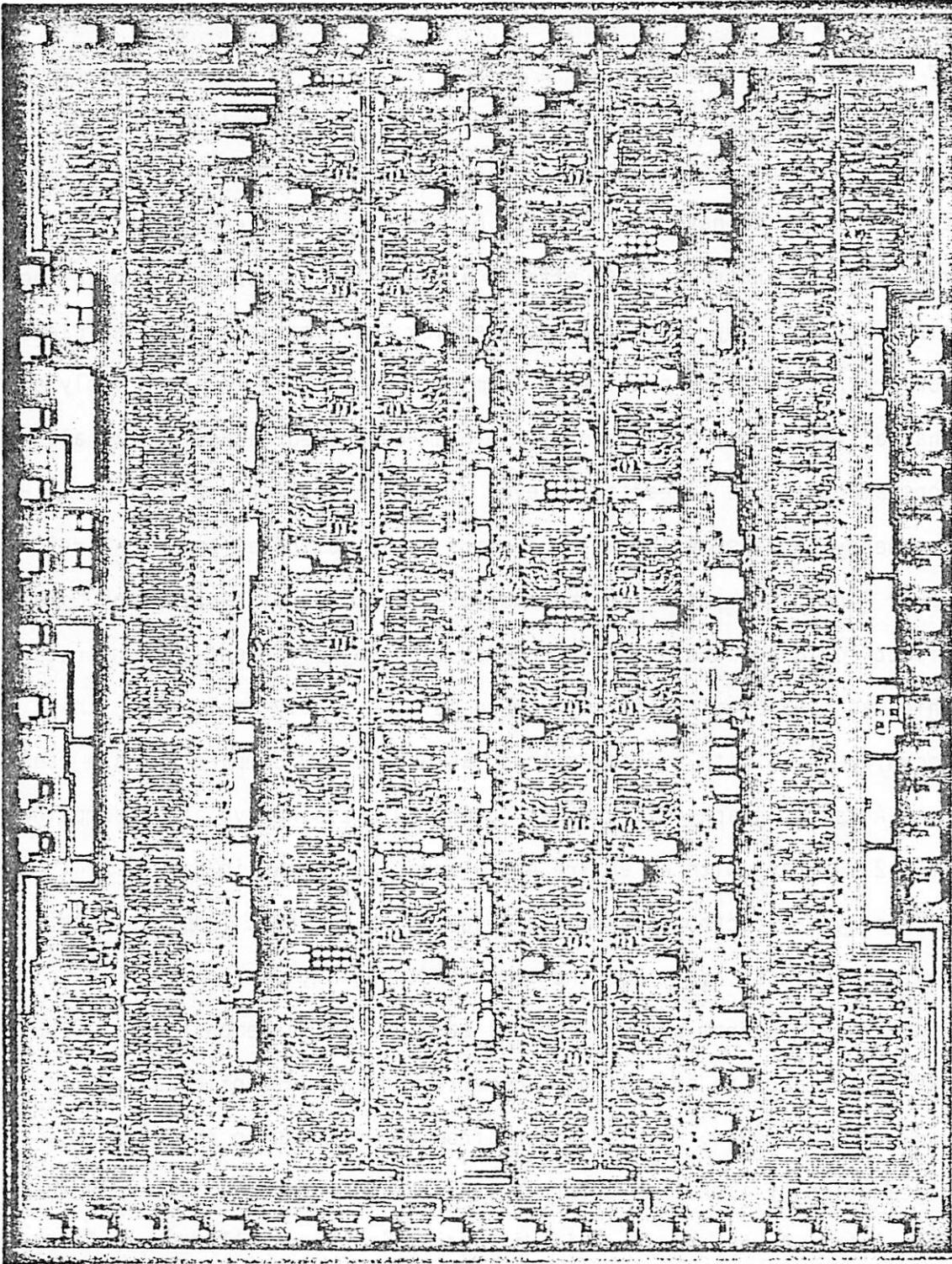


Fig. 6.27 Die photograph of monolithic converter array.

6.3.2 Experimental Results

Reconstructed ramp and sinewave input signals, which were digitized by this array chip, are shown in Fig.6.28. Low frequency linearity is maintained to approximately a 2.5 MHz conversion rate, with operation at reduced linearity (5-6 bits) continuing to 4 Hz. The low frequency linearity is dictated by binary-weighted capacitor matching, which in this case, appears to have been limited by consistent pattern-generator and mask run-out errors. These effects are usually pronounced in high speed circuits, which by necessity, must employ small capacitors. The capacitor errors are quite consistent from wafer to wafer, however, implying that a simple mask adjustment (or e-beam generated masks) should allow consistently good linearity in even higher resolution applications. High frequency conversion linearity appears to be dominated by signal-related power line noise affecting the single-ended comparator operation. This effect could be substantially reduced by using differential comparators, which as mentioned previously, could be more easily employed in this technique than with a flash method, as only a few comparators are used.

A spectrograph of a reconstructed 100 kHz sinewave after sampling at a 2 MHz rate with an entire 4-way array, is shown in Fig. 6.29a. Similarly, a spectrograph of the reconstructed output from a single converter within an array for the same input signal, is shown in Fig. 6.29. Power readings from several typical samples, which demonstrate 100 kHz SNR performance in both single converter and array configurations, are shown in Table 6.4. As is evident from the figures in these tables, a degradation of 1 or 2 dB occurs in the SNR in going from a single converter to a 4-way array. This difference is principally due to independent nonlinearities affecting the best fit converter gains and offsets,

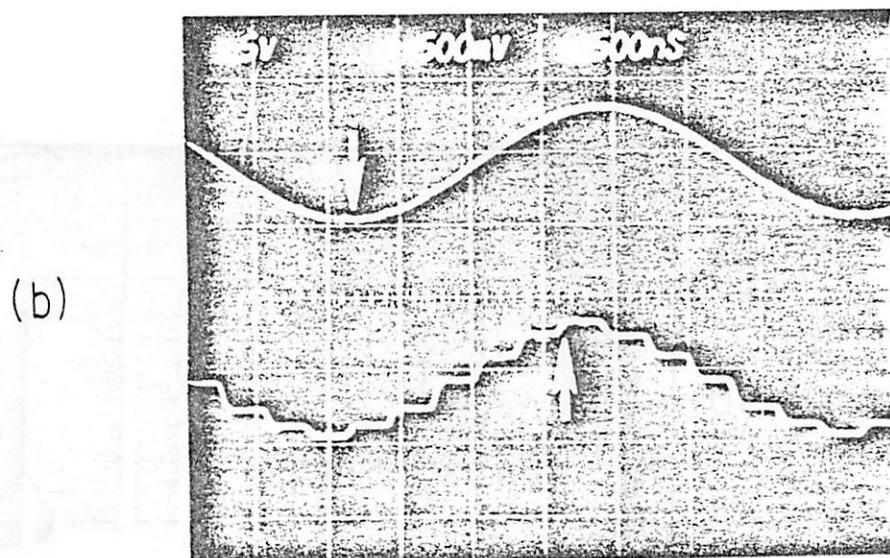
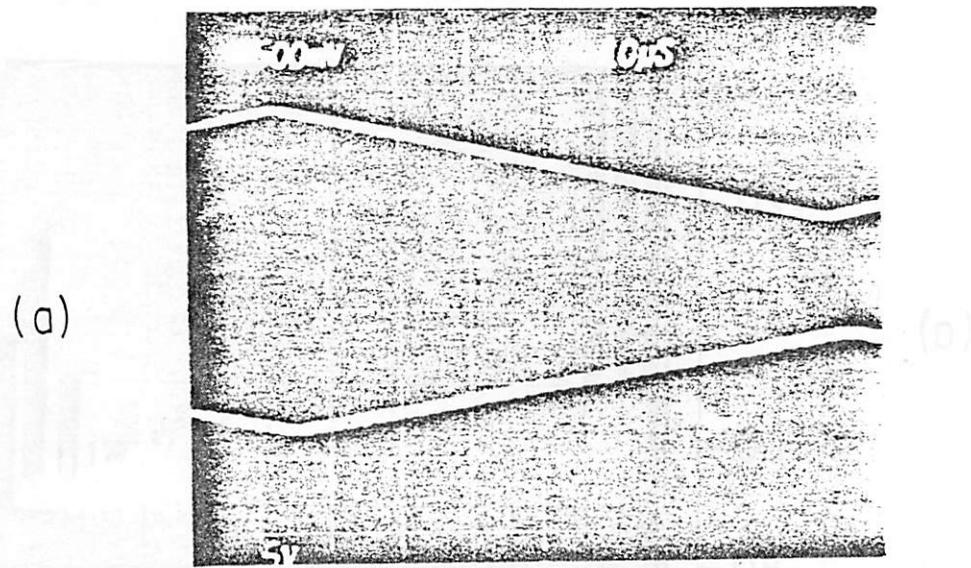


Fig. 6.28 Original and reconstructed waveforms after sampling with converter array chip (5V/div. on vertical scale, lower trace of each pair inverted). (a) Single converter within an array at a 625 kHz conversion rate. (b) Four-way array at 3.5 MHz conversion rate (points shown correspond).

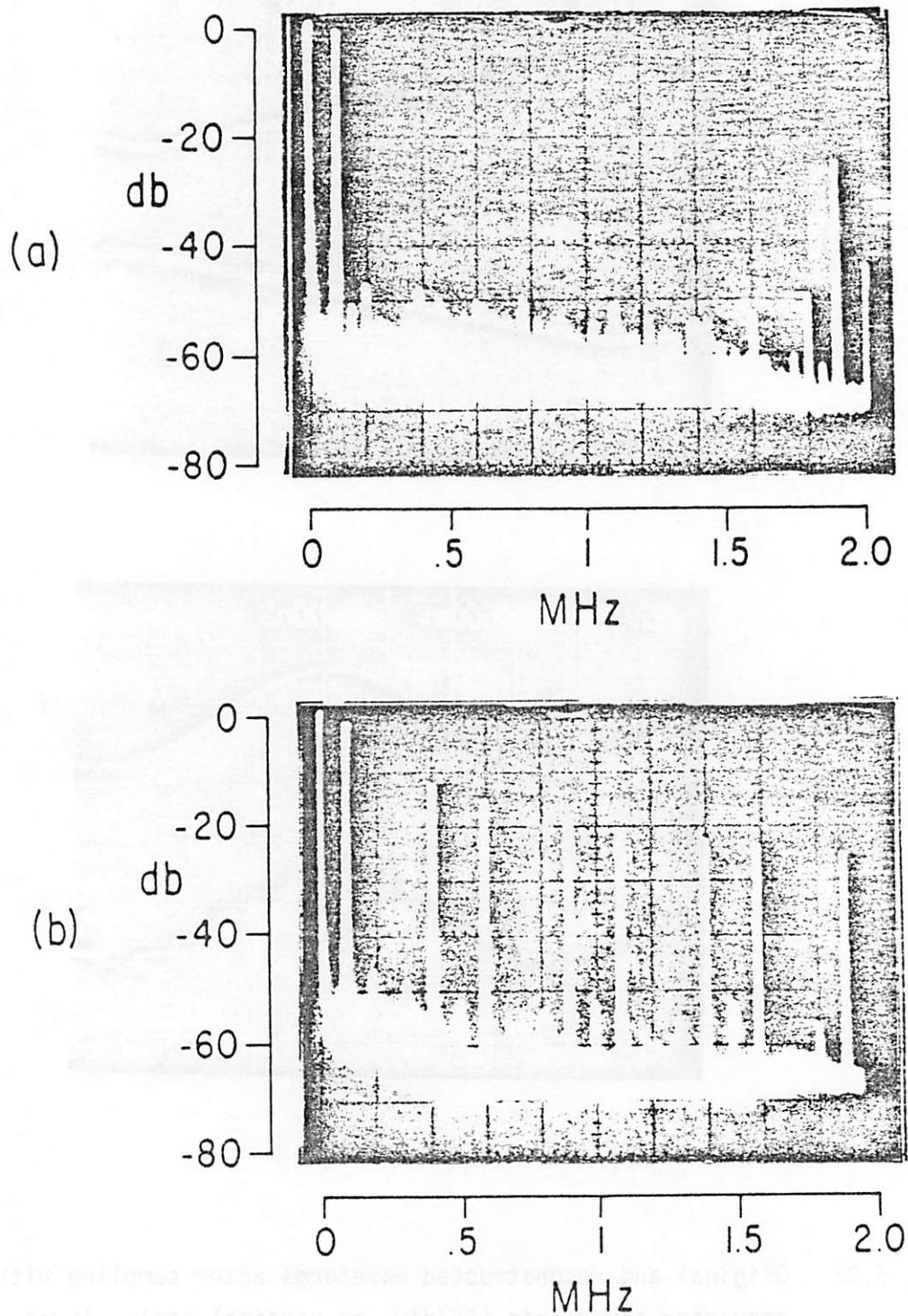


Fig. 6.29 Spectrum of reconstructed 100 kHz sine wave after sampling with array chip. (a) Four way array at 2 MHz sampling rate. (b) Single converter within an array at same clock rate as (a).

TABLE 6.4

1 Converter		2 Converters		4 Converters	
Converter	SNR	Converters	SNR	Converters	SNR
1	40.3				
2	41.9	1 and 3	39.4	1-4	38.7
3	39.15	2 and 4	40.3		
4	40.0				
Avg	40.3	Avg	39.85		38.7

1 Converter		2 Converters		4 Converters	
Converter	SNR	Converters	SNR	Converters	SNR
1	41.2				
2	42.2	1 and 3	39.4	1-4	39.4
3	41.5	2 and 4	40.9		
4	41.2				
Avg	41.5	Avg	40.15		39.4

Measured 100 kHz SNR performance of two converter array chips at a 4 MHz clock rate (2 MHz sampling rate for a 4-way array).

and a slight variance in feedthrough from the comparator reset switch. Signal-to-noise ratio performance as a function of input signal frequency is shown in Fig. 6.30. The degradation of performance at high frequencies is principally due to limited comparator bandwidth during the sample mode as discussed in Chapter 5, although some degradation due to sampling skew is evident as well. Channel isolation between converters is sufficiently high that no component of a full scale 100 kHz input signal to one converter is observable in the reconstructed quiet channel output of any other. A summary of the characteristics of this test converter array chip is given in Table 6.5.

The previous analysis regarding noise and distortion from nonideal arrays, was verified using this test chip. As mentioned previously, circuitry has been included on-chip for modifying the intrinsic offset of each A/D converter by fixed fractional LSB amounts. Similarly, the apparent gain of each converter may be modified by simply resistively attenuating the input signal at the input to each converter (off-chip). Spectral plots of reconstructed sinewaves after artificially inducing gain and offset variations in this manner are shown in Fig. 6.31. Note that noise spikes are at the frequencies indicated in Fig. 6.8, while the magnitude of the additional error power is approximately that indicated by Table 6.2. The effect of an induced offset mismatch on the quiet-channel noise output spectrum is shown in Fig. 6.32b, as compared to the nominal output noise spectrum in Fig. 6.32a. Note that the noise floor in each of these figures is that of the spectrum analyzer, and not broadband noise from the A/D or D/A converters.

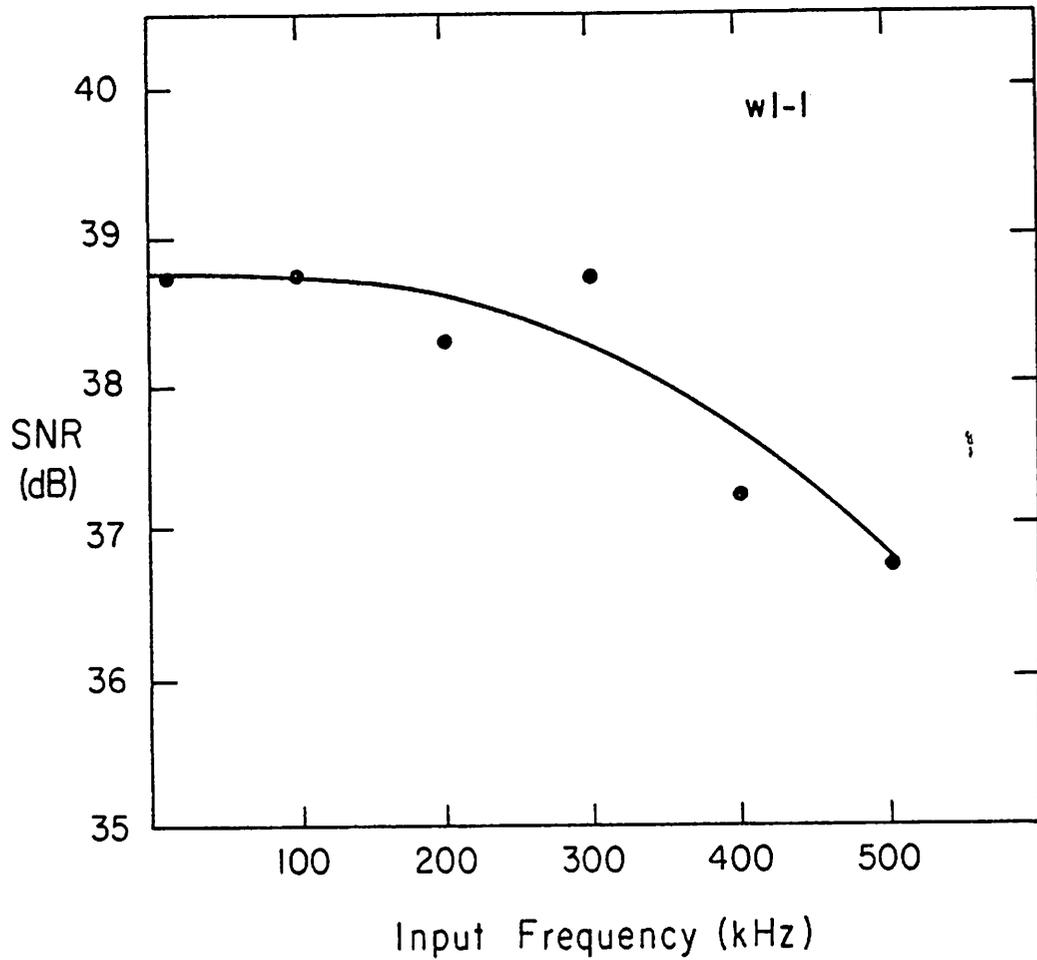


Fig. 6.30 Measured SNR performance as a function of input frequency for a 4-way array operating at a 2 MHz sampling rate.

TABLE 6.5

NUMBER OF MOS DEVICES	1250
TOTAL POWER CONSUMPTION	250 mW
MAXIMUM SAMPLING RATE PER CHIP	4 MHz
DIE SIZE	208x278 MILS
RESOLUTION (2.5 MHz SAMPLE RATE)	7 BITS
LINEARITY	1/2 LSB
S/N RATIO, FULL SCALE INPUT	39 db
PHASE SKEW AND JITTER	< 4 nsec
REFERENCE VOLTAGE DRIFT, 0-70° C	< 1 %

Summary of converter array chip characteristics.

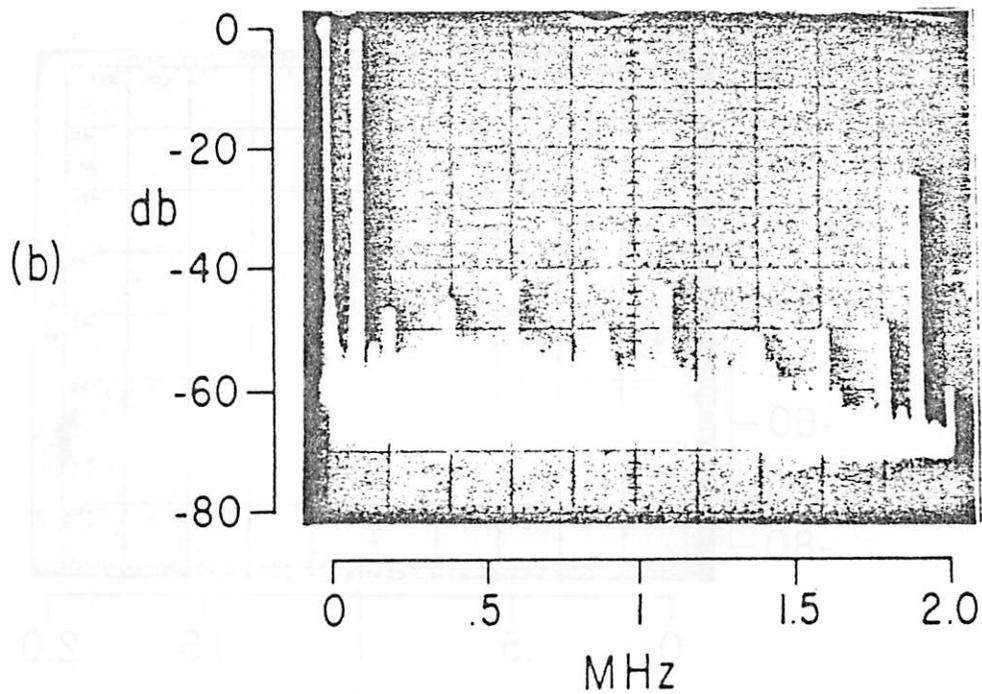
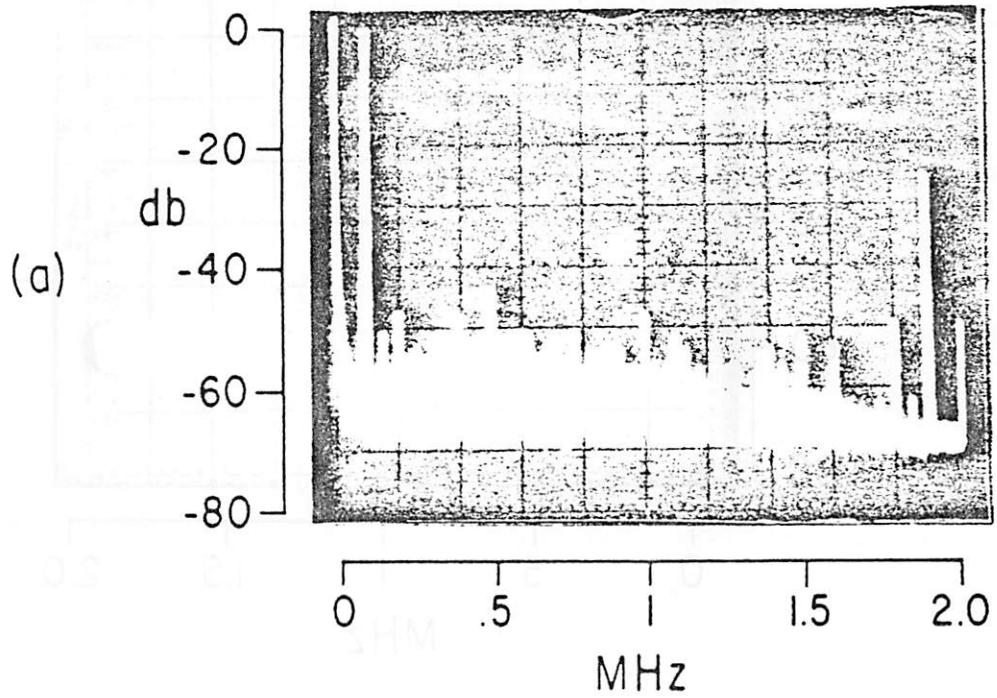


Fig. 6.31 Noise and distortion due to induced converter mismatch. Spikes shown are caused by mismatch indicated. (a) Gain variation (σ_a/\bar{a}) of approximately 1%. (b) Offset variation (σ_b) of approximately 0.5% at full-scale range.

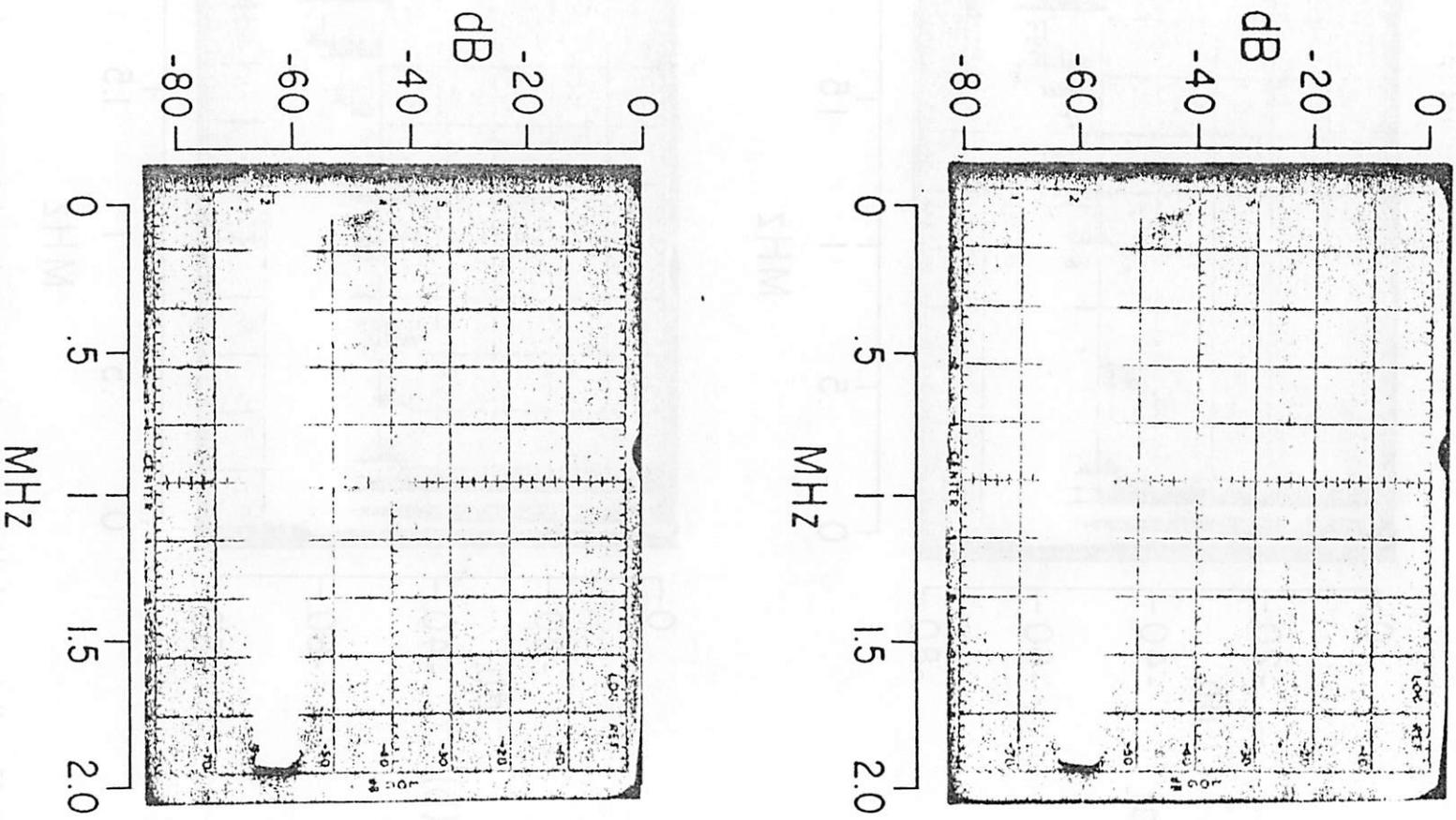


Fig. 6.32 Quiet channel output spectrum. (a) Nominal. (b) With induced offset mismatch.

6.3.3 Anticipated Characteristics of Implementations in An Advanced Process

If this basic design, which used a 10 micron, fully guard-banded CMOS process and program interconnected cells, were to be implemented in a modern 4-5 micron silicon-gate process using more area efficient techniques, a substantial improvement in die-size and performance would result. The die size would be reduced by a factor of four by simply scaling the major lithographic dimensions to half of their present size (shrinking 10 micron gates to 5 microns for instance). Further reductions in die size would occur if the field isolation regions were implanted or self-aligned, rather than consisting of diffused guard bands, and the computer generated cell-based layout was replaced with one drawn by hand. In fact, an overall reduction in die size by a factor of 6 appears to be conservative if the above improvements were incorporated into a new design.

As the maximum conversion rate is directly related to the capacitance which must be driven by internal nodes, the reductions in die size considered above, would all result in higher speed. Further improvements in conversion rate could be obtained by using optimized device sizes in all of the internal logic, rather than a standard cell approach, and by incorporating a lower resistivity interconnect than the P+ diffusions which were available here. These various improvements would allow an improvement in conversion rate by at least a factor of four. This would imply that an 8-bit video bandwidth (10-15 MHz) converter could be fabricated in a die size of less than 10,000 mil².

6.4 Application to a Digital Signal Processor

Monolithic digital signal processors which include on-chip A/D and D/A conversion circuits, would be useful in many applications if the circuits could be made inexpensively and with sufficiently high resolution. As high speed converters have tended to consume large die areas, and because very fast digital filters are difficult to implement in an MOS technology, however, single-chip processors have thus far been constrained to operate at near audio bandwidths [13]. Application of this array technique to a digital signal processor may allow higher frequency processors to become practical.

As this array technique will allow a fast converter to be realized in a smaller area than with other methods, more chip area is available for a digital processor within a maximum allowable die size. This would allow expanded capabilities in a single or few chip digital processor of the type illustrated in Fig. 6.33, although the digital processing section may still be difficult to implement in very fast systems. An alternative architecture, which takes advantage of the structure of a converter array, is shown in Fig. 6.34. In this case, the outputs from separate converters are processed independently and are only combined just before reconstruction. This architecture has the advantage of greatly reduced bandwidth in each digital processor and very simple expandability. This multiple-path technique has been employed with a variety of analog filter paths [14-16], for the realization of various bandpass and comb filtering functions, but has thus far not been applied to a digital processor. It would appear to be ideally suited to MOS implementations, as it trades several slower but potentially more area efficient circuits for a single high speed device (as was demonstrated above for A/D converters), and

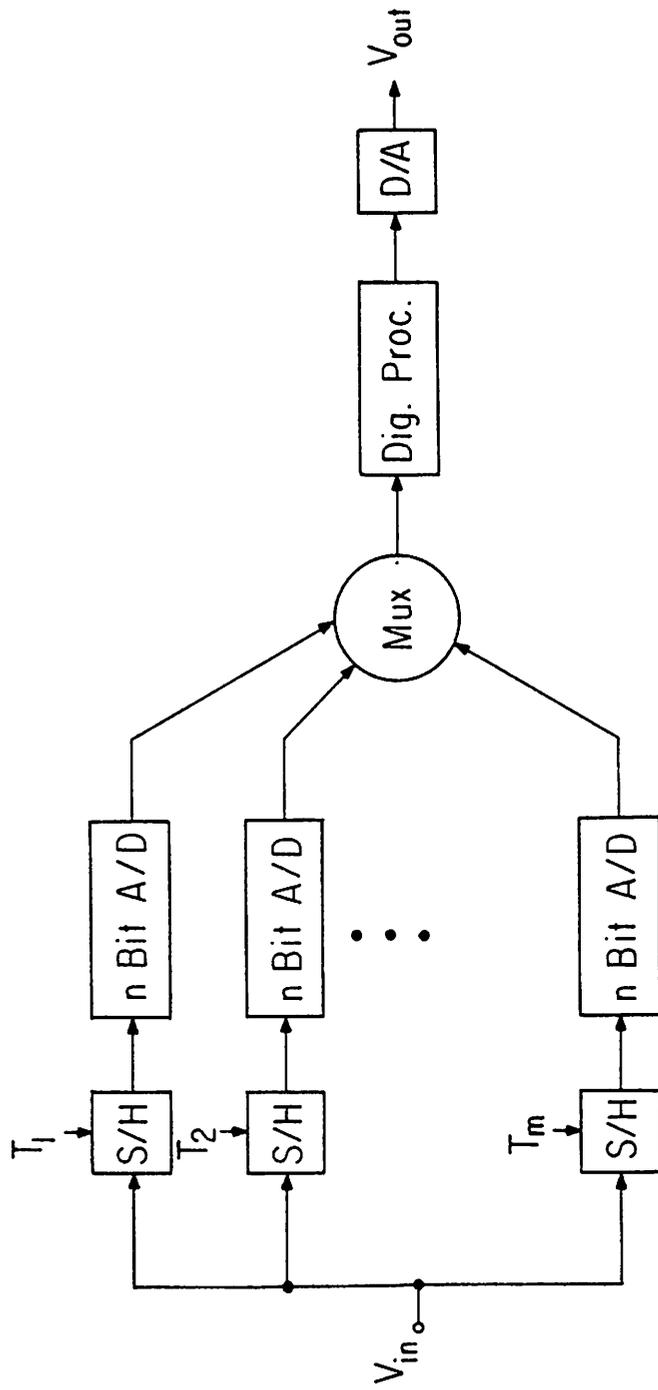


Fig. 6.33 Analog in/analog out digital signal processor employing a converter array front-end.

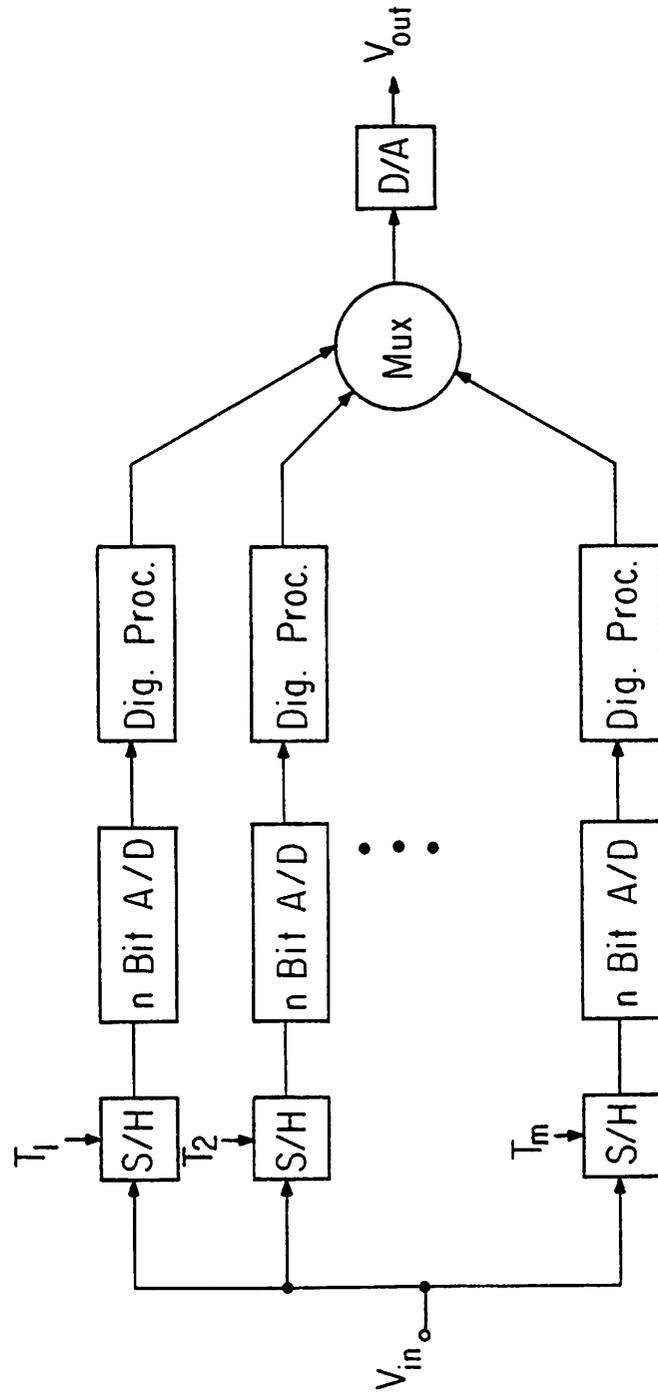


Fig. 6.34 Digital signal processor employing multiple path techniques throughout.

is easily expandable by simply interleaving additional processor paths or chips.

CHAPTER 7

CONCLUSIONS

Three basic high speed analog to digital conversion techniques have been analyzed with regard to implementation in a standard CMOS process. They are; the flash, successive approximation and pipeline converter methods. A new method for using A/D converters has also been analyzed and demonstrated; the time-interleaved converter array. This method, which employs a number of converters with synchronized sampling times to achieve a fast effective conversion rate, has been shown to be useful in achieving higher bandwidths, or smaller die sizes at a given bandwidth than are possible with any single converter. Although this method will, in some cases, result in increased noise or distortion, these effects are both predictable and consistent and may be minimized in the design of the array. This technique has been demonstrated in a 4-way successive approximation array chip, which has been used to verify the analysis of array characteristics and to show the feasibility of the method. Extensions of this work into implementations in an advanced process or as part of a digital processing system have also been discussed. This technique should be useful in reducing the cost of both high speed monolithic converter circuits and single- or few-chip digital signal processors which must operate on high frequency input signals.

REFERENCES

- [1] D.N. Gregory, J.L. Bliss, I.D.B. Millar and C.J. Allen, "Color picture synchronization by the natlock system," J. SMPTE, 78, 611-614, August 1969.
- [2] R.J. Bulter, "Remote color genlock," J. SMPTE, 78, 615-618, August 1969.
- [3] H. Schmid, "Synchronization of remote program sources for color TV broadcasting," J. SMPTE, 78, 619-620, August 1969.
- [4] D.E. Acker and R.H. McLean, "Digital time-base correction for video signal processing," J. SMPTE, 85, 146-160, March 1976.
- [5] K. Kano, T. Jmai, M. Inaba, A. Sugimoto, Y. Itoh, Y. Inoue, and J. Shimizu, "Television frame synchronizer," J. SMPTE, 84, 129-134, March 1975.
- [6] J.B. Mately, "A digital framestone synchronizer," J. SMPTE, 85, 385-388, June 1976.
- [7] A.A. Goldberg, "PCM NTSC television characteristics," J. SMPTE, 85, 141-145, March 1976.
- [8] R. Burkhardt and J. Wasser, "Digital television transmission with 34 μ bits/s," J. SMPTE, 89, 244-248, April 1980.
- [9] K. Yokoyama, S. Nakagawa, and H. Katayama, "An experimental digital videotape recorder," J. SMPTE, 89, 173-180, March 1980.
- [10] J.P. Rossi, "Sub-nyquist encoded PCM NTSC color television," J. SMPTE, 85, 1-6, January 1976.
- [11] V.G. Devereux, "Differential coding of PAL video signals using intra-field prediction," Proc. of IEEE, 124, 1139-1147, December 1977.

- [12] P. Rainger and P.A. Ratliff, "Low bit-rate system for digital coding of the television picture," J. SMPTE, 88, 550-554, August 1979.
- [13] J.P. Rossi, "Digital television image enhancement," J. SMPTE, 84, 545-551, July 1975.
- [14] A. Iga et al., "A consumer PCM audio unit connectable to home-use video tape recorder," presented at the Conference of the Acoustical Society of Japan, October 1977, paper 3-2-9.
- [15] Y. Tsuchiya et al., "Professional PCM audio unit using error correcting code," presented at the Conference of the Acoustical Society of Japan, October 1977, paper 3-2-10.
- [16] B.A. Blesser, "Digitization of audio," J. Audio Engineering Soc., Vol. 26, No. 10, pp. 750-771, October 1978.
- [17] T. Doi et al., "On several standards for converting PCM signals into video signals," J. Audio Engineering Soc., Vol. 26, No. 9, pp. 641-649, September 1978.
- [18] J.A. McCracken, "A high-performance digital audio recorder," J. Audio Engineering Soc., Vol. 26, No. 7/8, pp. 560-561, July/August 1978.
- [19] R.W. Brodersen, P.R. Gray, and D.A. Hodges, "MOS switched-capacitor filters," Proceedings of the IEEE, pp. 61-71, January 1979.
- [20] W.C. Black, Jr., D.A. Allstot, S.H. Patel, and J.B. Wieser, "CMOS PCM channel filters," ISSCC Dig. of Technical Papers, pp. 84-85, February 1980.
- [21] T. Doi, et al.
- [22] G.W. McNally, "Microprocessor mixing and processing of digital audio signals," J. Audio Engineering Soc., Vol. 27, No. 10, pp. 793-830, October 1979.

- [23] Y. Kawakami et al., "A Single-Chip Digital Signal Processor for Voiceband Applications," ISSCC Dig. of Technical Papers, pp. 40-41, February 1980.
- [24] National Semiconductor Data Acquisition Handbook, Chapter 15, 1978.
- [25] T. Redfern, National Semiconductor Corp., private communication.
- [26] J.R. Naylor, "Testing digital/analog and analog/digital converters," IEEE Trans. on Circuits and Systems, Vol. CAS-25, No. 7, pp. 526-538, July 1978.
- [27] J.G. Peterson, "A monolithic video A/D converter," IEEE J. Solid-State Circuits, Vol. SC-14, No. 6, pp. 932-937, December 1979.
- [28] M.O. Felix, "Differential phase and gain measurements in digitized video signals," SMPTE Journal, Vol. 85, pp. 76-79, February 1976.
- [29] W.A. Kester, "Characterizing and testing A/D and D/A converters for color video applications," IEEE Trans. on Circuits and Systems, Vol. CAS-25, No. 7, July 1978.
- [30] R.A. Nordstrom, "High-speed integrated A/D converter," ISSCC Dig. of Technical Papers, pp. 150-151, February 1976.
- [31] J.G. Peterson, op.cit. [27].
- [32] A.G.F. Dingwall, "Monolithic expandable 6 bit 20 MHz CMOS/SOS A/D converter," IEEE J. Solid-State Circuits, Vol. SC-14, pp. 926-932, December 1979.
- [33] J.G. Graeme, G.E. Tobey, and L.P. Huelsman, Operational Amplifiers, McGraw-Hill, New York, 1971.
- [34] Y. Nishikawa and J.E. Solomon, "A general purpose wideband operational amplifier," ISSCC Dig. of Technical Papers, pp. 144-145, February 1973.

- [35] K.E. Kuijk, "A fast integrated comparator," IEEE J. Solid-State Circuits, Vol. SC-8, pp. 458-462, 1973.
- [36] J.G. Peterson, op.cit. [27].
- [37] M.A. Polinsky, O.H. Schade Jr., and J.P. Keller, "CMOS/bipolar monolithic integrated circuit technology," Proc. Inter. Electron Dev. Meeting, p. 229, December 1973.
- [38] W.C. Black, Jr., R.H. McCharles and D.A. Hodges, "CMOS (and bipolar) processes for high performance analog LSI," Proc. Inter. Electron Dev. Meeting, pp. 331-334, December 1976.
- [39] J.G. Peterson, op.cit. [27].
- [40] A.G.F. Dingwall, op.cit. [32].
- [41] R. Kanachero and P.R. Gray, University of California, Berkeley, private communication.
- [42] J.G. Peterson, op.cit. [27].
- [43] R.M. McCharles, Hewlett-Packard Laboratories, private communication.
- [44] R.M. McCharles, V.A. Saletore, W.C. Black, Jr., and D.A. Hodges, "An algorithmic analog to digital converter," ISSCC Dig. of Technical Papers, pp. 96-97, February 1977.
- [45] D.R. Breuer and J.D. Hyde, "10-bit 5-megasample/second monolithic A/D converter," ISSCC Dig. of Technical Papers, pp. 152-153, February 1976.
- [46] P.H. Saul, P.J. Ward, and A.J. Fryers, "A 5 ns monolithic D/A subsystem," ISSCC Dig. of Technical Papers, pp. 18-19, February 1980.
- [47] J.L. McCreary and P.R. Gray, "All-MOS charge-redistribution A/D conversion techniques," IEEE J. Solid-State Circuits, Vol. SC-10, pp. 371-379, December 1975.

- [48] J.L. McCreary, "Successive approximation analog-to-digital conversion techniques in MOS integrated circuits," Ph.D. dissertation, University of California, Berkeley, 1975.
- [49] W.C. Black, Jr. and D.A. Hodges, "Time interleaved converter arrays technique," ISSCC Dig. of Technical Papers, pp. 14-15, February 1980.
- [50] B.M. Gordon, "Linear electronic analog/digital conversion architectures, their origins, parameters, limitations and applications," IEEE Trans. on Circuits and Systems, Vol. CAS-25, No. 7, pp. 391-418, July 1978.
- [51] F.D. Waldhauer, U.S. Patent #3,187,325, June 1, 1965.
- [52] R.M. Jepperson, U.S. Patent #3,119,105, January 21, 1964.
- [53] R.J. Van de Plassche and R.E.J. Van der Grift, "A high-speed 7-bit A/D converter," IEEE J. Solid-State Circuits, Vol. SC-14, No. 6, pp. 938-943, December 1979.
- [54] *ibid.*
- [55] N.F. Foster, "Piezoelectric and piezoeresistive properties of films," a chapter of 'Handbook of Thin Film Technology,' McGraw-Hill Book, Co., New York, 1968.
- [56] G. Kelson, H.H. Stellrecht and D.S. Perloff, "A monolithic 10-bit digital-to-analog converter using ion implantation," IEEE J. Solid-State Circuits, Vol. SC-6, pp. 396-403, December 1973.
- [57] J.A. Schoeff, "An inherently monotonic 12-bit DAC," IEEE J. Solid-State Circuits, Vol. SC-14, pp. 904-911, December 1979.
- [58] G. Erdi, "A precision trim technique for monolithic analog circuits," ISSCC Digest of Technical Papers, pp. 193-192; February 1975.
- [59] D.T. Comer, "A monolithic 12-bit D/A Converter," ISSCC Digest of Technical Papers, pp. 104-105, February 1977.

- [60] A.P. Brokaw, "A monolithic 10-bit A/D using I²L and LWT thin-film resistors," IEEE J. Solid-State Circuits, Vol. SC-13, pp. 736-745, December 1978.
- [61] F. van de Wiele, R. van Overstraeten, and M. de Man, "Graphical method for the determination of junction parameters and of multiplication parameters," Solid State Electron., Vol. 13, pp. 25-36, 1970.
- [62] E.H. Stevens, "Depletion region thickness in diffused junctions," IEEE Trans. on Elect. Dev., Vol. ED-20, No. 7, pp. 659-660, July 1973.
- [63] H. Lawrence and R.M. Warner, Jr., "Diffused junction depletion layer calculations," Bell Syst. Tech. J., Vol. 39, pp. 389-403, 1960.
- [64] J.C. Irvin, "Resistivity of bulk silicon and of diffused layers in silicon," Bell Syst. Tech. J., Vol. 41, pp. 387-401, 1962.
- [65] J.P. Lee and S.M. Sze, "Depletion layer capacitance of cylindrical and spherical P-N junctions," Solid State Electron., 10, p. 1105, 1967.
- [66] Y. Ameniya and K. Kato, "Compensation for voltage dependence of diffused feedback resistors in operational amplifiers," IEEE J. Solid-State Circuits, Vol. SC-14, pp. 1118-1120, December 1979.
- [67] W.W. Gartner, Transistors, Principles, Design and Applications, Van Nostrand Co., Inc., Princeton, 1960.
- [68] K. Tan and P.R. Gray, "Fully integrated analog filters using bipolar-JFET technology," IEEE J. Solid-State Circuits, Vol. SC-13, pp. 814-821, December 1978.

- [69] J.L. McCreary and P.R. Gray, "All-MOS charge redistribution analog-to-digital conversion techniques -- Part I," IEEE J. Solid-State Circuits, Vol. SC-10, pp. 371-379, December 1975.
- [70] W.C. Black, Jr. et al., op.cit. [20].
- [71] T. Redfern, National Semiconductor Corp., Santa Clara, Ca., re. U.S. Patent Application.
- [72] G.F. Landsburg and G. Smarandoiu, "A two-chip CMOS CODEC," ISSCC Digest of Technical Papers, February 1978.
- [73] W.C. Black, et al., op.cit. [49].
- [74] J.L. McCreary, A.M. Voshchenkov, (unpublished).
- [75] J.L. McCreary, private communication.
- [76] S.M. Sze, Physics of Semiconductor Devices, J. Wiley, New York, 1969.
- [77] A. Many, et al., Semiconductor Surfaces, J. Wiley, New York, 1965.
- [78] J.L. McCreary, A.M. Voshchenkov (unpublished).
- [79] ibid.
- [80] A.L. Grebene, Analog Integrated Circuit Design, Van Nostrand Reinhold, New York, 1972.
- [81] American Institute of Physics Handbook, 2nd Ed., McGraw-Hill, New York, 1963.
- [82] P. Andreatch and H.J. McSkimin, J. Applied Physics, Vol. 47, pp. 1299-1301, 1976.
- [83] J.L. McCreary, op.cit. [74].
- [84] H. Schichman and D.A. Hodges, "Modeling and simulation of insulated-gate field-effect transistor switching circuits," IEEE J. Solid-State Circuits, Vol. SC-3, pp. 285-289, September 1968.

- [85] D. Frohman-Bentchkowski and A.S. Grove, "Conductance of MOS transistors in saturation," IEEE Trans. Electron Devices, Vol. ED-16, pp. 108-113, January 1969.
- [86] Y.A. El-Mansy and A.R. Boothroyd, "A simple two-dimensional model for IGFET operation in the saturation region," IEEE Trans. Electron Devices, Vol. ED-24, pp. 254-261, March 1977.
- [87] R.R. Troutman, "Low-level avalanche multiplication in IGFETs," Proc. Inter. Electron Dev. Meeting, pp. 43-46, December 1978.
- [88] W.C. Black, Jr., "High voltage metal gate CMOS process for large scale analog circuit applications," M.S. Thesis, University of California, Berkeley, June 1977.
- [89] *ibid.*
- [90] W.C. Black, Jr. et al., *op cit* [38].
- [91] J.E. Soloman, "The monolithic op amp: A tutorial study," IEEE J. Solid State Circuits, Vol. SC-9, pp.314-332, December 1974.
- [92] B.J. Hosticka, "MOS sampled data recursive filters using state variable techniques," Ph.D. dissertation, University of California, Berkeley, Memorandum No. ERL-M77/65, August 1977.
- [93] D. Senderowicz, D.A. Hodges, and P.R. Gray, "High-performance NMOS operational amplifier," IEEE J. Solid State Circuits, Vol. SC-13, pp. 760-766, December 1978.
- [94] W.C. Black, Jr., D.J. Allstot, and R.A. Reed, "A high-performance low power CMOS channel filter," IEEE J. Solid State Circuits, Cember 1980.
- [95] *ibid.*
- [96] H. Ohara, W. Baxter, C. Rahim, J. McCreary and P.R. Gray, "A new second generation low power MOS switched capacitor PCM channel Filter," Proc. Inter. Conf. on Communications, June 1980.

- [97] K.C. Hseuh, University of California, private communication.
- [98] Y.P. Tsividis and P.R. Gray, "An integrated NMOS operational amplifier with internal compensation," IEEE J. Solid State Circuits, Vol. SC-11, pp. 748-753, December 1976.
- [99] D. Senderowicz, et al, op cit [93].
- [100] W.C. Black, Jr., et al, op cit [94].
- [101] J.S. Brugler and P.G.A. Jespers, "Charge pumping in MOS devices," IEEE Trans. on Elect. Dev., Vol. ED-16, pp. 297-302, March 1969.
- [102] Y.S. Yee, L.M. Terman and L.G. Heller, "A 1 mV MOS comparator," IEEE J. Solid State Circuits, Vol. SC-13, pp. 294-297, June 1978.
- [103] M.B. Das and J.M. Moore, "Measurement and interpretation of low-frequency noise in FET's," IEEE Trans. Electron Dev., Vol. ED-21, pp. 247-257, April 1974.
- [104] F.M. Klaassen, "Characterization of low 1/f noise in MOS transistors," IEEE Trans. Electron Dev., Vol. ED-18, pp. 887-891, October 1971.
- [105] J. Bertails, "Low frequency noise considerations for MOS amplifier design," IEEE J. Solid State Circuits, Vol. SC-14, pp. 773-776, August 1979.
- [106] *ibid.*
- [107] F. Breole, Sandia Laboratories, private communication.
- [108] A.G.F. Dingwall, op cit. [32].
- [109] D.G. Messerschmitt, "Effects of bit-robbing in tandem digital switches," IEEE Trans. on Communications, Vol. COM-27, pp. 1025-1028, July 1979.
- [110] W.W. Goodall, "Television by pulse code modulation," Bell System Technical Journal, Vol. 30, No. 1, pp. 33-49, January 1951.

- [111] W.C. Black, Jr. et al, op cit. [49].
- [112] W.C. Black, Jr. and D.A. Hodges, "Time interleaved converter arrays," IEEE J. Solid State Circuits, December 1980.
- [113] M. Townsend, M.E. Hoff, Jr. and R.E. Holm, "An NMOS microprocessor for analog signal processing," IEEE J. Solid State Circuits, Vol. SC-15, No. 1, pp. 33-38, February 1980.
- [114] L.E. Franks and J.W. Sandberg, "An alternative approach to the realization of network transfer functions: The N-path filter," Bell Syst. Technical Journal, Vol. 39, pp. 1321-1350, September 1960.
- [115] W.E. Heinlein, K.H. Moehrmann and W.H. Holmes, "Double-turned N-path bandpass filters using a single gyrator," IEEE Trans. on Circuit Theory, pp. 728-729, November 1971.
- [116] D.J. Allstot and K.S. Tan, "A switched capacitor N-path filter," Proc. Inter. Circuits and Systems Conference, March 1980.

APPENDIX A

SIGNAL TO NOISE RATIO SIMULATION PROGRAM

(Flash Simulator Shown)

Initial variable definitions

N = # of bits

L = # of converters

E = relative component error (σ_R/\bar{R} if using normal distribution routine)

V1 = Reference voltage

A9 = $V_{in(peak)}/V_{max}$

Intermediate and output variable definitions

R(I,J) = Jth resistor of converter i

A ϕ (I) = Best fit offset of Ith converter

A1(I) = Best fit gain of Ith converter

A ϕ MA = Average array offset

A1MA = Average array gain

$$EP1 = \frac{1}{L} \sum_{i=1}^L \langle \delta_i^2 \rangle = \langle \overline{\delta_i^2} \rangle$$

$$EP2 = \frac{1}{L} \sum_{i=1}^L \langle x \delta_i \rangle (a_i - \bar{a})$$

$$EP3 = \frac{1}{L} \sum_{i=1}^L \langle \delta_i \rangle (b_i - \bar{b})$$

$$EP4 = \frac{1}{L} \sum_{i=1}^L \langle x \rangle (a_i - \bar{a})(b_i - \bar{b})$$

$$EP5 = \frac{1}{L} \sum_{i=1}^L \langle x^2 \rangle (a_i - \bar{a})^2$$

$$EP6 = \frac{1}{L} \sum_{i=1}^L (b_i - \bar{b})^2$$

$$EPX = \sum_{i=1}^6 EP_i$$

```

C*****
C*
C*   SIGNAL TO NOISE RATIO SIMULATION PROGRAM
C*
C*****
      DIMENSION R(10,1024),T(10)
      DIMENSION SIG(10),A0(10),A1(10),EP(6,10)
      DOUBLE PRECISION P1,S,S1,S2,S3,S4,S5,S6
      REAL J1,J2
      N=7           ! N = # OF BITS
      K=N+1
      L=2           ! L = # OF CONVERTERS
      E=1./32.     ! E = REL. RES. ERROR
      V1=5.        ! V1 = REFERENCE VOLTAGE
      A9=1.0       ! A9 = VIN(PEAK)/VMAX
      I1=1
      OPEN(UNIT=20,FILE='SIG1.DAT',ACCESS='SEQOUT')
      WRITE(20,200) N,L,E
200   FORMAT(2I5,F10.8)
      R1=1.        ! DEFINE UNIT RESISTOR
      DO 50 JJ=1,50 ! LOOP THROUGH # OF ARRAYS
      TYPE 100
100   FORMAT(1H )
      DO 10 I1=1,L
      I9=I1
      CALL ERR(K,R,R1,N,L,E,I9) !ARRAYS WITH ERRORS
10   CALL TOT(K,T,R,N,L,I9)   !ARRAY TOTALS
      BOFF=0.
      DO 50 BOFF=0,0.0        !LOOP THROUGH OFFSETS
      ROFF=BOFF
      TYPE 101,ROFF,A9,E
101   FORMAT(1H ,'/ROFF = ',F10.6,'/ A9 = ',F10.6,
1' E = ',F10.6)
      Q2=1           !SWITCH /1/ FOR INDIV. CONVERTERS
      C             !SWITCH /0/ FOR WHOLE ARRAY
      CALL LIN(K,T,R,N,L,V,V1,D,Q2,A0,A1,N1,A9,A1M,A0M
1,AOFF) !FIND BEST FIT LINES/EACH CONVERTER
      DO 20 I1=1,L
      Q2=I1        !SWITCH I1 FOR CONVERTER I1
      C           !SWITCH /0/ FOR WHOLE ARRAY
      CALL CHEK(K,T,R,N,L,V,V1,D,Q2,A0,A1,A9,S1,S,J2,
1J1,P1,N1,S2,S3,S4,S5,S6,ROFF,A1M,A0M) !FIND ERR.
      EP(1,I1)=S1;EP(2,I1)=S2;EP(3,I1)=S3
      EP(4,I1)=S4;EP(5,I1)=S5;EP(6,I1)=S6
      S=10*ALOG10((.5*A9+A9*V1*V1/S))
      S1=10*ALOG10((.5*A9+A9*V1*V1/S1))
      SIG(I1)=10.*(S1/10.)
      TYPE 102,I1,A1(I1),A0(I1),S1,J2      !OUTPUT
102   FORMAT(1H ,'/CONVERTER # ',I5.4(F12.8))
      WRITE(20,201) A1(I1),A0(I1),S1,J2,J1 !OUTPUT
      @

```

```

201     FORMAT(2F10.6,3F10.6)
20     CONTINUE
      A1MA=0.;A0MA=0.
      DO 30 I=1,L
      A1MA=A1MA+A1(I)
30     A0MA=A0MA+A0(I)
      A1MA=A1MA/L;A0MA=A0MA/L !AVG. GAIN AND OFFSET
      TYPE 103,A1MA,A0MA
103    FORMAT(1H ,/AVERAGE GAIN AND OFFSET /,2(F12.8))
      EP1=0.;EP2=0.;EP3=0.;EP4=0.;EP5=0.;EP6=0.
      DO 40 I=1,L
      EP1=EP1+EP(1,I)
      EP2=EP2+EP(2,I)♦(A1(I)-A1MA)
      EP3=EP3+EP(3,I)♦(A0(I)-A0MA)
      EP4=EP4+EP(4,I)♦(A1(I)-A1MA)♦(A0(I)-A0MA)
      EP5=EP5+EP(5,I)♦(A1(I)-A1MA)♦(A1(I)-A1MA)
40     EP6=EP6+EP(6,I)♦(A0(I)-A0MA)♦(A0(I)-A0MA)
      EP1=EP1/L
      EP2=EP2♦2./L
      EP3=EP3♦2./L
      EP4=EP4♦2./L
      EP5=EP5/L
      EP6=EP6/L
      EPX=EP1+EP2+EP3+EP4+EP5+EP6
      TYPE 104,EP1,EP2,EP3
104    FORMAT(1H ,/EP1 = /,F12.8,/ EP2 = /,F12.8,
1/ EP3 = /,F12.8)
      TYPE 105,EP4,EP5,EP6
105    FORMAT(1H ,/EP4 = /,F12.8,/ EP5 = /,F12.8,
1/ EP6 = /,F12.8)
      EPY=10♦ALOG10((.5♦A9♦A9♦V1♦V1/EPX))
      WRITE(20,202) A1MA,A0MA
      WRITE(20,202) EP1,EP2,EP3,EP4
      WRITE(20,202) EP5,EP6,EPX,EPY !OUTPUT
202    FORMAT(1H ,4(F12.8,3X))
      TYPE 106,EPX,EPY
106    FORMAT(1H ,/TOTAL ERROR POWER /,F12.8,F12.8)
      GO TO 50 !REMOVE FOR FULL ANALYSIS
      Q2=0
      CALL LINK(K,T,R,N,L,V,V1,D,Q2,A0,A1,N1,A9,A1M,A0M,
1A0FF)
      TYPE 107,A1M,A0M
107    FORMAT(1H ,/BEST FIT GAIN AND OFFSET/,F12.8,F12.8)
      Q2=0
      CALL CHEK(K,T,R,N,L,V,V1,D,Q2,A0,A1,A9,S1,S,J2,
1J1,P1,N1,S2,S3,S4,S5,S6,A0FF,A1M,A0M)
      EPX=S1;EPY=10♦ALOG10((.5♦A9♦V1♦A9♦V1/S1))
      TYPE 106,EPX,EPY;TYPE 101
50     CONTINUE
      A9=9999.
2

```

```

204      WRITE(20,204) A9,A9      !WRITE 'FINISHED' CODE
      FORMAT(2F10.2)
      CLOSE(UNIT=20)
      STOP
      END
C*****
      SUBROUTINE ERR(K,R,R1,N,L,E,I1)
      DIMENSION R(10,1024)
      E1=E
      DO 1 I=1,IFIX(2.*N+.01)
      CALL STAT(E1,X)           !CALL STAT FOR GAUSSIAN
C      X=E1*(RAN(0)-.5)*2      !THIS LINE FOR BOX DIST.
      IF(X.LT.-1.) X=-1.      !TRUNCATE ERR. FOR R>0
      R(I1,I)=R1*(1.+X)
1      CONTINUE
      RETURN
      END
C*****
      SUBROUTINE TOT(K,T,R,N,L,I1)
      DIMENSION T(10),R(10,1024)
      T(I1)=0
      DO 1 I=1,IFIX(2.*N+.01)
      T(I1)=T(I1)+R(I1,I)
1      CONTINUE
      RETURN
      END
C*****:
      SUBROUTINE ATOD(K,T,R,N,L,I1,V,V1,D)
      DIMENSION T(10),R(10,1024)
      RT=0.
      DO 1 I=1,IFIX(2.*N+.01)-1
      RT=RT+R(I1,I)
      V5=RT/T(I1)*2.*V1-V1
      IF(V5.LT.V) GO TO 1
      D=I-1
      RETURN
1      CONTINUE
      D=I-1
      RETURN
      END
C*****
      SUBROUTINE DTOR(K,N,W,V1,D)
      W=D/(FLOAT(IFIX(2.*N+.01)))*2.*V1-V1
      RETURN
      END
C*****
      SUBROUTINE PROB(N1,V1,V5,P,A9,ROFF)
      DOUBLE PRECISION P
C      A9=% OF FULL SCALE
      IF(V1*A9*V1*A9.GT.(V5-ROFF)*(V5-ROFF)) GO TO 1
      @

```

```

      P=0.000
      GO TO 2
1     P=1/(3.1415926535*SQRT(A9*A9*V1*V1-(V5-AOFF)*(V5-AOFF)))
      P=P*A9*V1*2./FLOAT(N1) !MAKE 1./FLOAT(N1) FOR
                                !A RAMP INPUT
2     CONTINUE
      RETURN
      END
C *****
SUBROUTINE LINK(K,T,R,N,L,V,V1,D,Q2,A0,A1,N1,A9,A1M,A0M
1,AOFF)
      DIMENSION R(10,1024),T(10)
      DIMENSION A0(10),A1(10)
      DOUBLE PRECISION P,T1,T2,T3,T4,T5,T6
      N1=0;T1=0.;T2=0.;T3=0.;T4=0.;T5=0.;T6=0.
      TEMP1=V1*A9
      TEMP2=TEMP1/10000.
      N2=IFIX(2.*(TEMP1/TEMP2)+1.000001)
      NUMB=2001.*L
      DO 1 I1=1,L
      IF(Q2.EQ.0) GO TO 2
      N1=0
      T1=0.;T2=0.;T3=0.;T4=0.;T5=0.
      T6=0
2         DO 4 V6=-1.*TEMP1+AOFF,TEMP1+AOFF,TEMP2
              V=V6
              N1=N1+1
              I9=I1
              CALL ATOD(K,T,R,N,L,I9,V,V1,D)
              CALL DTOD(K,N,W,V1,D)
              X=V6
              Y=W
              V5=V6
              CALL PROB(N2,V1,V5,P,A9,AOFF)
              IF(Q2.EQ.0) P=P/L
                  T1=T1+X*P
                  T2=T2+Y*P
                  T3=T3+X*Y*P
                  T4=T4+X*X*P
                  T5=T5+Y*Y*P
4         T6=T6+P !MAY MONITOR T6 FOR CHEK OF INT(P)=1
              CONTINUE
      IF(Q2.EQ.0)GO TO 1
      A1(I1)=(T3-T2*T1)/(T4-T1*T1)
      A0(I1)=T2-A1(I1)*T1
1     CONTINUE
      IF(Q2.NE.0) GOTO 3
      A1M=(T3-T2*T1)/(T4-T1*T1)
      A0M=T2-A1M*T1
C     TYPE 10,A1M,A0M
3

```

```

3      RETURN
      END
C*****
SUBROUTINE CHEK(K,T,R,N,L,V,V1,D,Q2,A0,A1,A9,
1S1,S,J2,J1,P1,N1,S2,S3,S4,S5,S6,AOFF,A1M,A0M)
DIMENSION R(10,1024),T(10)
DIMENSION A1(10),A0(10)
DOUBLE PRECISION P1,J3,S1,J,S,J4,S2,S3,S4,S5,S6
REAL J1,J2
S1=0;S=0;J2=0;J1=0;S2=0.;S3=0.;S4=0.;S5=0.;S6=0.
TEMP1=V1*A9
DO 1 V5=-TEMP1+AOFF,TEMP1+AOFF,TEMP1/10000
      I2=1;I3=L
      IF(Q2.NE.0) I2=Q2
      IF(Q2.NE.0) I3=Q2
          DO 4 I1=I2,I3
              V=V5
              I9=I1
              CALL ATOD(K,T,R,N,L,I9,V,V1,D)
              CALL DTOA(K,N,W,V1,D)
              J4=W-A1(I1)*V5-A0(I1)
IF(Q2.EQ.0) J4=W-A1M*V5-A0M
              J3=ABS(J4)
              V6=V5
6          CALL PROB(N1,V1,V6,P1,A9,AOFF)
7          S1=S1+J3*J3*P1
              S2=S2+J4*V5*P1
              S3=S3+J4*P1
              S4=S4+V5*P1
              S5=S5+V5*V5*P1
              S6=S6+P1
C      SET UP FOR IDEAL SINUSOIDS
      S4=AOFF;S5=AOFF*AOFF+(A9*V1*A9*V1)/2
      S6=1.
          J=ABS(W-V5+V1/(2.*N))
          S=S+J*J*P1
          IF(J.LT.J1) GO TO 3
          J1=J
3          IF(J3.LT.J2) GO TO 4
          J2=J3
4          CONTINUE
1          CONTINUE
      RETURN
      END
C*****
SUBROUTINE STAT(E1,X)
A00=2.515517
A01=.802853
A02=.010328
A01=1.432788

```


APPENDIX B.1

PLANAR W, E AND V DETERMINATION

From Guass' law

$$\nabla \cdot \underline{E} = \frac{\rho}{\epsilon_s} \quad (\text{B.1.1})$$

as

$$\rho(x) = q[N_s e^{-x^2/4DT} - N_B] \quad (\text{B.1.2})$$

then

$$E_1(x) = \frac{qN_s}{\epsilon} \int_{x_L}^x \left[e^{-x^2/4DT} - \frac{N_B}{N_s} \right] dx \quad (\text{B.1.3})$$

and

$$E_2(x) = \frac{qN_s}{\epsilon} \int_{x_R}^x \left[e^{-x^2/4DT} - \frac{N_B}{N_s} \right] dx \quad (\text{B.1.4})$$

where E_1 and E_2 represent the electric fields present on the left (diffusion) side and right (substrate) side of the metallurgical junction while x_L and x_R represent the edges of the depletion region.

As

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (\text{B.1.5})$$

and via substitution

$$\int_0^x e^{-t^2/4DT} = \sqrt{\pi DT} \text{erf}(x/\sqrt{4DT}) \quad (\text{B.1.6})$$

then the E fields may be expressed as

$$E_1(x) = \frac{qN_s}{\epsilon} \left[\sqrt{\pi DT} \left[\operatorname{erf} \left(\frac{x}{\sqrt{4DT}} \right) - \operatorname{erf} \left(\frac{x_L}{\sqrt{4DT}} \right) \right] - \frac{N_B}{N_s} (x - x_L) \right] \quad (\text{B.1.7})$$

and

$$E_2(x) = \frac{qN_s}{\epsilon} \left[\sqrt{\pi DT} \left[\operatorname{erf} \left(\frac{x}{\sqrt{4DT}} \right) - \operatorname{erf} \left(\frac{x_R}{\sqrt{4DT}} \right) \right] - \frac{N_B}{N_s} (x - x_R) \right] \quad (\text{B.1.8})$$

For a given x_R or x_L , the other component (x_L or x_R) may be determined by solving for the case $E_1(x_j) = E_2(x_j)$, which is the peak field. Taking

$$E_1(x_j) - E_2(x_j) = \frac{qN_s}{\epsilon} \left[\sqrt{\pi DT} \left(\operatorname{erf} \left(\frac{x_R}{\sqrt{4DT}} \right) - \operatorname{erf} \left(\frac{x_L}{\sqrt{4DT}} \right) \right) + \frac{N_B}{N_s} (x_L - x_R) \right] \quad (\text{B.1.9})$$

It is evident that zeros occur where the function

$$F = \sqrt{\pi DT} \left(\operatorname{erf} \left(\frac{x_R}{\sqrt{4DT}} \right) - \operatorname{erf} \left(\frac{x_L}{\sqrt{4DT}} \right) \right) + \frac{N_B}{N_s} (x_L - x_R) \quad (\text{B.1.10})$$

is zero. Taking x_R as given, and finding that a closed form of F' exists

$$\frac{\partial F}{\partial x_L} = e^{-x_L^2/4DT} - \frac{N_B}{N_s} \quad (\text{B.1.11})$$

it is apparent that Newtons method may be applied to the above to indicate the values of x_L which create zeros of F (see Appendix B.3).

The electrostatic potential may be determined by

$$-\nabla V = \mathcal{E} \quad (\text{B.1.12})$$

or

$$V_1 = - \int_{x_L}^{x_j} E_1(x) dx \quad (\text{B.1.13})$$

and

$$V_2 = + \int_{x_R}^{x_j} E_2(x) dx \quad (\text{B.1.14})$$

or

$$V_1 = \frac{-qN_s}{\epsilon_s} \left[\frac{1}{\sqrt{\pi DT}} \left[\int_{x_L}^{x_j} \operatorname{erf} \left(\frac{x}{\sqrt{4DT}} \right) dx \right. \right. \\ \left. \left. - \operatorname{erf} \left(\frac{x_L}{\sqrt{4DT}} \right) \int_{x_L}^{x_j} dx \right] - \frac{N_B}{N_s} \int_{x_L}^{x_j} x dx + \frac{N_B}{N_s} x_L \int_{x_L}^{x_j} dx \right] \quad (\text{B.1.15})$$

which because [B.1]

$$\int \operatorname{erf}(x) dx = x \operatorname{erf}(x) + \frac{1}{\sqrt{\pi}} e^{-x^2} + c \quad (\text{B.1.16})$$

becomes

$$V_1 = \frac{-qN_s}{\epsilon_s} \left[\frac{1}{\sqrt{\pi DT}} x_j \left(\operatorname{erf} \left(\frac{x_j}{\sqrt{4DT}} \right) - \operatorname{erf} \left(\frac{x_L}{\sqrt{4DT}} \right) \right) \right. \\ \left. + 2DT \left(e^{-x_j^2/4DT} - e^{-x_L^2/4DT} \right) - \frac{N_B}{2N_s} (x_j - x_L)^2 \right] \quad (\text{B.1.17})$$

and similarly

$$V_2 = \frac{+qN_s}{\epsilon} \left[\sqrt{\pi D T} x_j \left(\operatorname{erf} \left(x_j / \sqrt{4 D T} \right) - \operatorname{erf} \left(x_L / \sqrt{4 D T} \right) \right) \right. \\ \left. + 2 D T \left(e^{-x_j^2 / 4 D T} - e^{-x_L^2 / 4 D T} \right) - \frac{N_B}{2 N_s} (x_j - x_L)^2 \right] \quad (\text{B.1.18})$$

where it is the desired condition, that

$$F = V_B - V_1 - V_2 - \phi = 0 \quad (\text{B.1.19})$$

where

$$\phi = \frac{k T}{q} \ln \left[\frac{N(x_L) N(x_R)}{n_i^2} \right] \quad (\text{B.1.20})$$

As F is not differentiable by x_R (due to x_L being a function of x_R), Newton's method is not applicable without a numerical differentiation. It was found that the Secant method resulted in lower computational cost to convergence, allowing a fairly rapid determination of x_R for a specified V_B (see Appendix B.3).

APPENDIX B.2

CYLINDRICAL W, E AND V

From Gauss' law

$$\nabla \cdot \mathbf{E} = \rho/\epsilon \quad (\text{B.2.1})$$

it may be shown, that for a cylindrically symmetrical impurity distribution,

$$\frac{1}{r} \frac{\partial}{\partial r} [rE(r)] = \frac{\rho(r)}{\epsilon} \quad (\text{B.2.2})$$

or,

$$E_1(r) = \frac{1}{\epsilon r} \int_{r_L}^r r \rho(r) dr + c \quad (\text{B.2.3})$$

and

$$E_2(r) = \frac{1}{\epsilon r} \int_{r_R}^r r \rho(r) dr + c \quad (\text{B.2.4})$$

where E_1 and E_2 correspond to the E fields present on the resistor and substrate side of the metallurgical junction, and r_L and r_R correspond to the resistor and substrate depletion region edges. For the gaussian impurity distribution that has been assumed, the above reduces to:

$$E_1(r) = - \frac{2DTqN_s}{r\epsilon_s} \left[e^{-r^2/4DT} - e^{-r_L^2/4DT} \right] - \frac{qN_B}{2r\epsilon_s} (r^2 - r_L^2) \quad (\text{B.2.5})$$

$$E_2 = -\frac{2DTqN_s}{r\epsilon_s} \left[e^{-r^2/4DT} - e^{-r_r^2/4DT} \right] - \frac{qN}{2r\epsilon_s} (r^2 - r_r^2) \quad (\text{B.2.6})$$

For a given r_r or r_L , the other component (r_L or r_r) may be determined by solving for the case $E_1(r_j) = E_2(r_j)$, which is the peak field. Taking $E_1(r_j) - E_2(r_j)$, it is evident that zeros occur where the function

$$F = 2DTN_s \left[e^{-r_L^2/4DT} - e^{-r_r^2/4DT} \right] + \frac{N_B}{2} (r_L^2 - r_r^2) \quad (\text{B.2.7})$$

is zero. Taking r_r as given, and finding that a closed form of F' exists

$$\frac{\partial F}{\partial r_L} = -N_s e^{-r_L^2/4DT} r_L + N_B r_L \quad (\text{B.2.8})$$

it is apparent that Newton's method may be applied to the above to indicate the values of r_L which create zeros of F (see Appendix B.3).

The electrostatic potential may be determined by

$$-\nabla V = \underline{E} \quad (\text{B.2.9})$$

we then have

$$V_1(r) = \int_{r_L}^r E_1(r) dr \quad (\text{B.2.10})$$

$$= \frac{-2DTqN_s}{\epsilon} \left[\int_{r_L}^r \frac{e^{-r^2/4DT}}{r} dr - e^{-R_L^2/4DT} \cdot \ln \left(\frac{r}{r_L} \right) \right] - \frac{qN}{4\epsilon} (r^2 - r_L^2) + \frac{qNr_L^2}{2\epsilon} \ln (r/r_L) \quad (\text{B.2.11})$$

Further:

$$\int \frac{e^{-a^2 r^2} dr}{r} = \ln(r) + \sum_{i=1}^N \frac{(ar)^{2i} (-1)^i}{i!(2i)} + \epsilon \quad (\text{B.2.12})$$

where

$$|\epsilon| \leq \frac{(ar)^{2N}}{N!(2N)} \quad (\text{B.2.13})$$

Thus

$$V_1(r) = \frac{-q}{\epsilon} \left[-\frac{N_B}{4} (r^2 - r_L^2) + \frac{N_B}{2} r_L^2 \ln(r/r_L) + 2DTN_s e^{-r_L^2/4DT} \ln(r/r_L) - 2DTN_s [f(r) - f(r_L)] \right] \quad (\text{B.2.14})$$

and

$$V_2(r) = \frac{q}{\epsilon} \left[-\frac{N_B}{4} (r^2 - r_r^2) + \frac{N_B}{2} r_r^2 \ln(r/r_r) + 2DTN_s e^{-r_r^2/4DT} \ln(r/r_r) - 2DTN_s [f(r) - f(r_r)] \right] \quad (\text{B.2.15})$$

where

$$f(r) = \ln(r) + \sum_{i=1}^N \frac{(r^2/4DT)^i (-1)^i}{i!(2i)} \quad (\text{B.2.16})$$

where it is the desired condition, that

$$F = V_B - V_1 - V_2 - \phi = 0 \quad (\text{B.2.17})$$

where

$$\phi = \frac{kT}{q} \ln \left[\frac{N(r_L)N(r_r)}{n_i^2} \right] \quad (\text{B.2.18})$$

As in the planar case, the Secant method is used to determine the appropriate value of r_r given a specific V_B (see Appendix B.3).

APPENDIX B.3

PROGRAM FOR EVALUATION OF RESISTIVITY AND VOLTAGE
COEFFICIENT OF DIFFUSED RESISTORS, INCLUDING SIDEWALLS

Initial variable definitions

$$D1 = DT$$

$$Q1 = Q_T \text{ (total implant dose in ions/cm}^2\text{)}$$

M = mobility coef.

k = mobility coef. ($\mu = kN^m(x)$)

N = N_B (background doping in ions/cm³)

L = resistor length

W = resistor width (on mask)

V = V_B (applied bias)

Output variable definitions

$$A2 = \left. \frac{\partial x_L}{\partial V} \right|_{V=V_B} \quad \text{planar case}$$

$$A4 = \left. \frac{\partial r_L}{\partial V} \right|_{V=V_B} \quad \text{cylindrical case}$$

$$R1 = \rho_s \quad \text{planar}$$

$$F1 = \left. \frac{\partial \rho_s}{\partial V} \right|_{V=V_B} \quad \text{planar}$$

$$G = G_{\text{side}} \quad \text{cylindrical}$$

$$F2 = \left. \frac{\partial G}{\partial V} \right|_{V=V_B} \quad \text{cylindrical}$$

$R = R_T$ combined planar and cylindrical resistance

$R_2 = R / (\frac{L}{W} R_1)$ normalized resistance with respect to mask dimensions

$F_3 = \frac{\partial R_T}{\partial V} \Big|_{V=V_B}$ total resistor bias voltage coefficient

$R_3 = \frac{1}{R_T} \frac{\partial R_T}{\partial V} \Big|_{V=V_B}$ normalized bias voltage coefficient