HIGH FREQUENCY CMOS SWITCHED-CAPACITOR FILTERS

by

Tat Cheung Choi

Memorandum No. UCB/ERL M83/31

18 May 1983

HIGH FREQUENCY CMOS SWITCHED CAPACITOR FILTERS

by

Tat Cheung Choi

Memorandum No. UCB/ERL M83/31

18 May 1983

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

High Frequency CMOS Switched-Capacitor Filters

By

Tat Cheung Choi

B.E.E. (University of Minnesota) 1975

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Engineering

in the

GRADUATE DIVISION

OF THE

UNIVERSITY OF CALIFORNIA, BERKELEY

Approved: Robert W. Brodersen 5/18/83
Chairman Date
Paul R. Gray May 17, 1983
Alan H. Potts May 19, 1983

HIGH FREQUENCY CMOS SWITCHED-CAPACITOR FILTERS

Ph. D. Tat Cheung Choi Dept. of EECS

Robert W. Brodersen
Chairman of Committee

Abstract

Extension of switched-capacitor filtering techniques into the MHz range would allow a higher level of integration in high frequency communication systems. However, monolithic realization of high frequency switched-capacitor filters encounters numerous problems, namely, limitation of operational amplifier settling time, sensitivity to component variations, increase in noise due to higher Q, and presence of parasitic power supply coupling paths. New techniques are introduced to solve these problems. First, a high speed differential cascode CMOS amplifier is used to achieve fast settling time. Second, an identical resonator filter structure is used to solve the problem of sensitivity. Third, a T-network scheme is used to reduce component ratio spread. Fourth, Thevenin equivalent circuits are used to solve the D.C. stability problem in elliptic bandpass ladders. Fifth, frequency translation is used to relax the filter requirements as well as reduce the filter in-band noise. Last but not least, fully differential signal paths are used to reduce clock feedthroughs and increase power supply rejection. These concepts are demonstrated in an experimental filter fabricated in CMOS technology. This prototype filter is a sixth order 260 KHz elliptic bandpass filter with a Q of 40 and clocked at 4 MHz. The results meet all the filter specifications and show good correlation with theoretical values.

Acknowledgement

I would like to thank Prof. Bob Brodersen for his support and encouragement throughout the course of this work. He has been both mentor and friend, and is always there to lift your spirits when the experiments are not doing well. I would also like to thank Ron Kaneshiro for undertaking this joint project with me, Prof. Paul Gray for his numerous ideas and support, and Prof. Dave Hodges for his guidance as academic adviser.

My fellow students in the EECS Dept. have been extremely helpful and have contributed to this work in different ways. It is not possible to list all their names here. However, special thanks should go to John Hui, who taught me a lot about process technology, Ping Ko, for his insights in device physics, and David Soo, who shared with me his many ideas in circuit and filter design. In addition, I want to thank Simon Wong, Po Tong, Y. K. Ip, Ignatius Lee, and Sou C. Au, who have given me lots of encouragement and have widened my perspectives in many different ways. A word of thanks should also go to Willy Yau, who have helped me with the figures. Last but not least, I would like to thank my parents for my education and their patience with me through the years.

TABLE OF CONTENT

Chapter 1

Introduction

Recent advances in analog MOS integrated circuits have drastically brought down the cost of communication systems [12]. With improvements in circuit design and process technology, these analog IC's can incorporate more circuit functions than was previously possible, resulting in a higher level of integration with more precise and sophisticated circuitry. One of these recent developments has been in the area of frequency selective filters, which form a vital part of any communication system. Historically, filters have not been integrable because their frequency response depend on *RC* time constants, and precision resistors and capacitors are not available in integrated circuits. The smallest filters then were in hybrid form [15], which used laser-trimmed thin film resistors and capacitors to achieve accurate frequency response. With the advent of switched-capacitor techniques [1,8], it is now possible to realize monolithic filters in integrated circuits.

Switched-capacitor filters are discrete time filters which transform an *RC* time constant into the product of a clock frequency times a capacitor ratio. The clock frequency can be arbitrarily accurate since it is derived from a crystal. Whereas capacitor ratios only depend on ratios of geometries, and with proper layout, accuracies on the order of 0.1% are easily achievable[16]. The introduction of switched-capacitor filters have a significant impact on voice band communication systems, and they found wide acceptance in commercial communication circuits such as codec filters for PCM telephony [3-6], channel bank filters for speech recognition systems [7], and various kinds of miscellaneous applications like modems [13] and DTMF receivers [14]. So far, all these applications have been limited to the audio range. However, communication systems span

the entire spectrum from voice band to microwave frequencies. There are many filter applications in higher frequencies, especially in the MHz range. Some examples are AM and FM intermediate frequency filtering in radio receivers, clock recovery in data communication systems, video processing in television receivers and channel filters for FDM telephony. Presently, these filtering functions are realized by external components such as LC filters and crystal filters. There is a need to find monolithic realization of these filters to allow a higher level of integration in these high frequency communication systems.

This dissertation is aimed at extending switched-capacitor techniques into the MHz range. In the past, switched-capacitor filters are limited to the voice band because of limitations of the operational amplifier settling time. Further, since high frequency applications often require highly selective filters, the problems of sensitivity to component values, dynamic range and stray power supply coupling paths must be overcome. This project uses a CMOS switched-capacitor filtering technique to allow operations in the MHz range [2,11]. Here, a high speed differential cascode CMOS amplifier is used to achieve fast settling time, and an identical resonator scheme is used to solve the problem of sensitivity. Also, fully differential signal paths are used for better power supply rejection and to minimize parasitic couplings between signal paths in high Q filters. A complete description of the amplifier design can be found in [26].

The organization of this dissertation is as follows. In Chapter 2, the design considerations for high frequency switched-capacitor filters are discussed. Here, the properties of the LDI transformation and the effects of non-ideal ladder terminations are first described. Next, the problem of D.C. stability in elliptic bandpass ladder filters is considered. This is followed by a section on the issue of sensitivity in high Q bandpass filters. Lastly, a decimation scheme is introduced to relax the filter requirements for high frequency high Q circuits.

Chapter 3 is concerned with the practical implementation of high frequency switched-capacitor filters. A fully differential filter implementation is described to alleviate the problems of settling, clock noise and PSRR. Next, a T-network scheme is introduced to reduce the component ratio spread in these high Q filters. To help reduce the clocking rate required, a double-sampling LDI integrator is discussed which will effectively double the sampling rates of conventional integrators. This is followed by detail design examples for an AM IF filter and a FM IF filter.

The interaction of CMOS technology with high speed analog application is the subject of discussion in Chapter 4. This chapter compares the NMOS and the CMOS technologies for this kind of application. Also, analog CMOS design considerations like latch-up and short channel transistors are addressed in detail.

Chapter 5 gives some experimental results on a test vehicle used to demonstrate our design concepts. This is a 260 KHz high Q bandpass filter intended for AM IF applications. Notice that the conventional IF frequency for AM radio is 455 KHz. Here, the 260 KHz IF filter is for car radios. This filter has very tight requirements, and the specifications call for Q's on the order of 40 and stopband rejection greater than 55 dB. The experimental chip easily meets these requirements.

Chapter 6 discusses the conclusions that we drawn from our studies. This is followed by two appendices, one is a description of the process flow of the CMOS process which was developed during this work, and the other is a list of design rules suitable for this CMOS process.

Chapter 2

Design Considerations for High Frequency Switched-Capacitor Filters

2.1 Introduction

Switched-capacitor ladder filters are sampled-data realizations of the "leap-frog" active filter structure which is a simulation of a doubly-terminated LC ladder. Since these filters are essentially discrete-time circuits, their resemblance to their continuous-time counterparts depends primarily on the transformation from the s-domain to the z-domain. In switched-capacitor filters, this transformation can be directly implemented by choosing a particular structure for a discrete-time integrator. A widely used structure is the Lossless Discrete Integrator (LDI) based on a digital filter structure introduced by Bruton [17]. Filters based on this integrator have been demonstrated to be very accurate realizations of LC prototype filters if the ratio of sample rate of the integrator to band edge frequencies is large [18]. In high frequency applications, it is often desirable to reduce this ratio as much as possible since this relaxes the settling requirements of the filter amplifiers. This, however, complicates the design procedure because of the non-linear characteristics of the LDI transformation. In addition, unavoidable errors associated with the ladder termination resistances become important as this ratio is reduced [9].

Another difficulty stems from the fact that most high frequency applications require very narrow band filters. In high frequency communication systems, frequency multiplexing is often used so that the same channel can be shared by multiple users. To separate the different frequency bands, very selective (high Q) filters are needed. This leads to sensitivity problems because of the inherently increased sensitivity of high Q filters both to the ratios of the capacitors in the filters as well as to the gain and settling behavior of the operational

amplifiers.

To alleviate this problem of sensitivity, the use of N-path or pseudo N-path filtering techniques have been suggested [19-22]. In N-path filters, the signal is translated from its initial center frequency to a range centered about zero frequency, low pass filtered, and then retranslated back up to the original center frequency. In order that all spurious quadrature components introduced by the modulation cancel out, multiple paths are required which are spaced equally in the phase of the modulating signals. The great advantage of this technique is that the center frequency is determined precisely by an externally supplied clock, with no dependence on capacitor ratios, and that the actual filter itself is a lowpass operating at a frequency much lower than the center frequency. The disadvantages are that the modulating signal is at the filter center frequency where carrier feedthrough in the modulators will directly degrade the dynamic range of the filter, that the parallel paths must match to a high degree of accuracy, and that parasitic passbands exist starting at a frequency twice the passband frequency. To solve the problem of carrier feedthrough, pseudo N-path techniques have been suggested. In these filters, the modulating signal will be exact multiples of the center frequency so that carrier feedthrough will not be inside the filter passband. To achieve the same delay as in N-path filters, multiple clock phases are used in the parallel paths. In order for pseudo N-path filters to be useful, it is important that these clock phases must be equally spaced. A slight offset will introduce an alias in the middle of the passband. For high frequency applications, it is difficult to produce clock signals that are exactly equally spaced, and in addition, the finite rise and fall times will be a problem. Also, unavoidable phase jitters will inadvertently cause feedthroughs in the passband frequencies. It appears likely that these N-path or pseudo N-path filters will be useful only in extremely high Q applications where extreme accuracy of the center frequency is paramount. However, in cases where the

required Q is more moderate and dynamic range is important, conventional filtering techniques are more likely to be used.

Another problem arises from the use of elliptic bandpass filters, which are more efficient in realizing sharp roll-off characteristics. The realization of elliptic bandpass filters in active ladder configuration can give rise to unstable D.C. conditions, thus driving the op. amps. into saturation. The successful implementation of high frequency, high Q filters will require an effective approach to solve this problem.

This chapter addresses the above problems. The source of these problems will be described in full detail together with design techniques that are used to solve them. In section 2.2, the LDI transformation is reviewed and it is shown that designs using this transformation for filters which have a low ratio of passband edge to sample rate involves a frequency warping very similar to the more familiar bilinear transformation [9]. In section 2.3, a first order solution to the termination problem is given which yields significant improvements in both passband and stopband responses when a low ratio of passband to sample rate is required [9]. Section 2.4 introduces a solution to the D.C. stability problem in elliptic bandpass ladders. This is followed by a discussion in section 2.5 on the issue of sensitivity in high Q filters, together with an identical-resonator scheme for these applications. Lastly, in section 2.6, a decimation method to relax the filter requirements in high frequency applications is described.

2.2 The LDI Transformation

2.2.1 PROPERTIES OF THE LDI TRANSFORMATION

The LDI transformation was introduced by Bruton [17] as a building block in digital ladder design. It was discussed by Jacobs et al. when used to implement switched-capacitor integrators [8]. The transformation between the s- and z- domains is defined by

$$s \iff \frac{1}{T_s}\frac{1-z^{-1}}{z^{-1/2}} \tag{2.1}$$

which should be compared to the bilinear transformation

$$s \iff \frac{2}{T_s}\frac{1-z^{-1}}{1+z^{-1}} \tag{2.2}$$

where T_s is the sampling period.

In transforming a continuous-time system to a discrete-time system, we generally require that the essential properties of the continuous-time system be preserved in its discrete-time counterpart. A good transformation from the s-domain to the z-domain has two basic properties, namely, the imaginary axis of the s-plane is mapped onto the unit circle of the z-plane, and a stable continuous-time filter is transformed into a stable discrete-time filter [23]. The bilinear transform as well as the LDI transform have both of these properties.

To see this for the case of the LDI transform, let Ω be the frequency of the digital filter. The unit circle in the z-plane is described by

$$z = e^{j\Omega T_s}. \tag{2.3}$$

Substituting this into (2.1), we have

$$s = j\frac{2}{T_s}\sin\frac{\Omega T_s}{2}. \tag{2.4}$$

Now defining $s = \sigma + j\omega$, where ω is the frequency in the continuous-time domain,

we obtain by equating real and imaginary parts,

$$\sigma = 0 \tag{2.5}$$

$$\omega = \frac{2}{T_s}\sin\frac{\Omega T_s}{2}. \tag{2.6}$$

Thus for values of z on the unit circle, we have $\sigma = 0$. Hence, the imaginary axis of the s-plane is mapped onto the unit circle of the z-plane. Also note that as the continuous-time filter frequency ranges from

$$-2f_s < \omega < 2f_s \tag{2.7}$$

that the discrete-time filter frequency ranges over its entire principle range of

$$-\frac{f_s}{2} < \frac{\Omega}{2\pi} < \frac{f_s}{2}. \tag{2.8}$$

Because there is a unique mapping between the $j\omega$ axis and the unit circle, there is no "aliasing" of the frequency response. The effects of the transformation can therefore be understood as a frequency axis warping given by (2.6).

We next consider the stability of the discrete-time filter so transformed. Defining $z = Re^{j\vartheta}$ where R and ϑ are the polar coordinates of z, we have from (2.1)

$$s = \frac{1}{T_s}(\sqrt{R}e^{j\frac{\vartheta}{2}} - \frac{1}{\sqrt{R}}e^{-j\frac{\vartheta}{2}})$$

or

$$s = \frac{1}{T_s}(\sqrt{R} - \frac{1}{\sqrt{R}})\cos\frac{\vartheta}{2} + j\frac{1}{T_s}(\sqrt{R} + \frac{1}{\sqrt{R}})\sin\frac{\vartheta}{2}. \tag{2.9}$$

Thus for

$$\mathrm{Re}\{s\} < 0$$

we have

$$\frac{1}{T_s}\left(\sqrt{R}-\frac{1}{\sqrt{R}}\right)\cos\frac{\vartheta}{2}<0.$$

But for $-\pi<\vartheta<\pi$,

$$\cos\frac{\vartheta}{2}>0$$

therefore

$$\sqrt{R}-\frac{1}{\sqrt{R}}<0$$

$$\sqrt{R}<\frac{1}{\sqrt{R}}$$

$$R<1.$$

But $z = Re^{j\vartheta}$, $R < 1$ implies $|z| < 1$, we therefore have

$$\mathrm{Re}\{s\}<0 \;\Rightarrow\; |z|<1. \tag{2.10}$$

Thus a stable continuous-time filter is transformed into a stable discrete-time filter.

The two properties summarized in (2.6) and (2.10) are analogous to those of the bilinear transformation. The difference between the two lies in the different ways by which they warp the frequency axis. The LDI transformation maps only the portion for $|\omega| < 2f_s$ and has a sinusoidal type of characteristics. The portion for $|\omega| > 2f_s$ is simply not mapped. On the other hand, the bilinear transformation maps the entire $j\omega$ axis onto the unit circle, and Ω and ω are related by

$$\omega=\frac{2}{T_s}\tan\frac{\Omega T_s}{2}. \tag{2.11}$$

In fact, the distortion introduced by a tangent type of characteristics is greater than that introduced by a sinusoid, especially at higher frequencies. To compare these two transformations, their corresponding mappings of the continuous-time filter frequencies are plotted in Fig. 2.1.

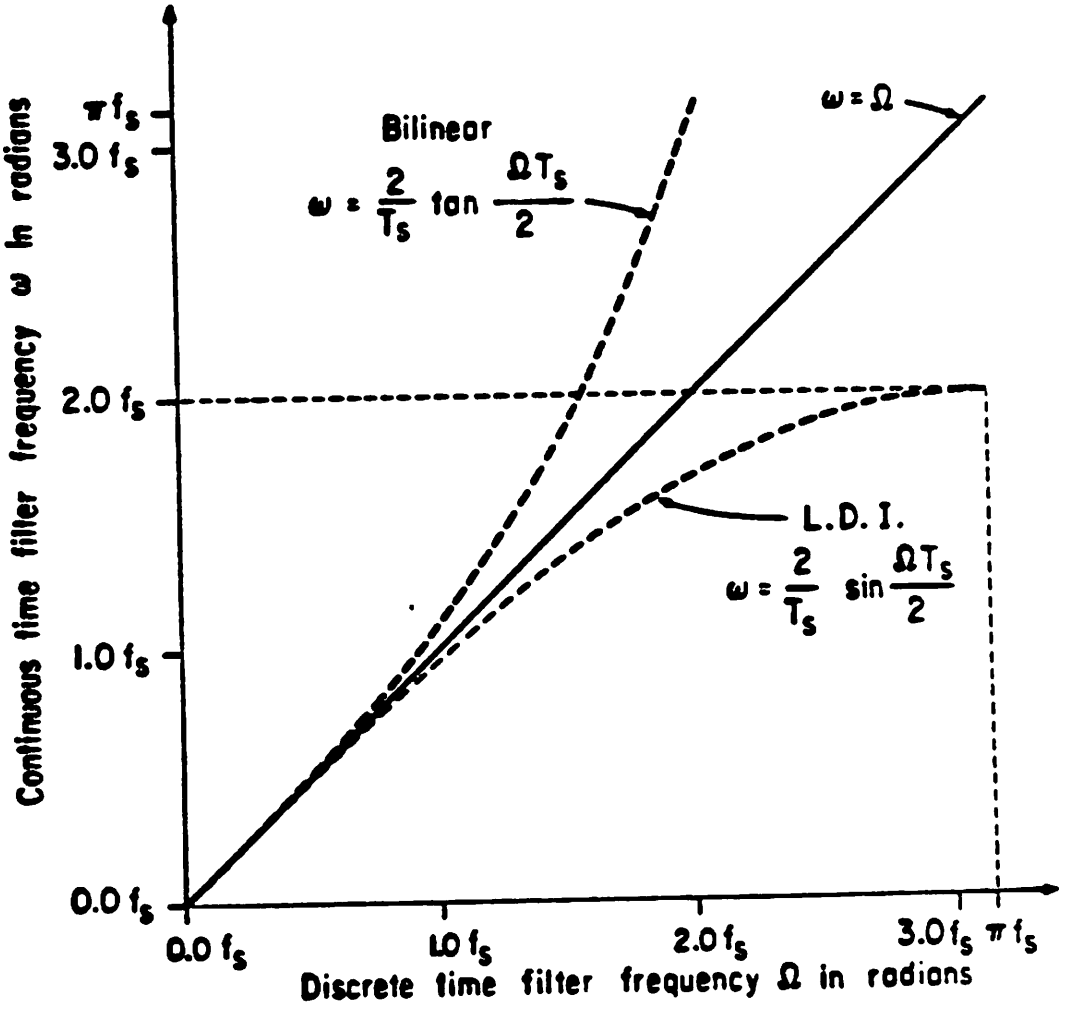


Fig. 2.1 Warping of the analog frequency axis by the bilinear and LDI transformations.

2.2.2 LDI LADDER CONFIGURATIONS

Design methodology for switched-capacitor ladder filters was developed by Jacobs et al. using signal flow graph techniques [8,25]. These filters simulates a doubly-terminated LC ladder so that the low sensitivity properties of the LC network are carried over to its active counterpart [24]. The basic building block of a switched-capacitor ladder filter is an LDI integrator, which can be realized by a single-ended op. amp. as well as a fully differential op. amp. Until recently, single-ended op. amps. are predominantly used because of their simple structure and ease of design. However, fully differential structures can offer better clock noise rejection, better PSRR, wider bandwidth, as well as use of chopper-stabilized techniques for flicker noise suppression [10]. Their disadvantage is that they require twice as many switches and capacitors, and therefore a larger die area is needed for their implementation. Here, we will consider the implementation of LDI ladders using both types of op. amps.

Earlier versions of LDI ladders use an integrator shown in Fig. 2.2, where V_1 is the non-inverting input and V_2, the inverting input. It takes half a cycle delay for the signal to propagate from the input to the output, and the transfer function can be described by

$$V_o(z) = \frac{C_1}{C_2} \frac{z^{-1/2}}{1-z^{-1}} (V_1(z) - V_2(z)) . \qquad (2.12)$$

To achieve the LDI transformation in a ladder structure, proper phasing of the switches is required. The most basic structure in a leapfrog filter is a two-integrator loop, shown in Fig. 2.3. To realize the LDI transform, alternate clock phasings are used in the two integrators. Notice that as the output of op. amp. 1 changes to a new value, it is immediately sampled by the other integrator. In so doing, we realize one cycle of delay per loop, as required by the LDI transform. The z-domain block diagram for this integrator loop is shown in Fig. 2.4.

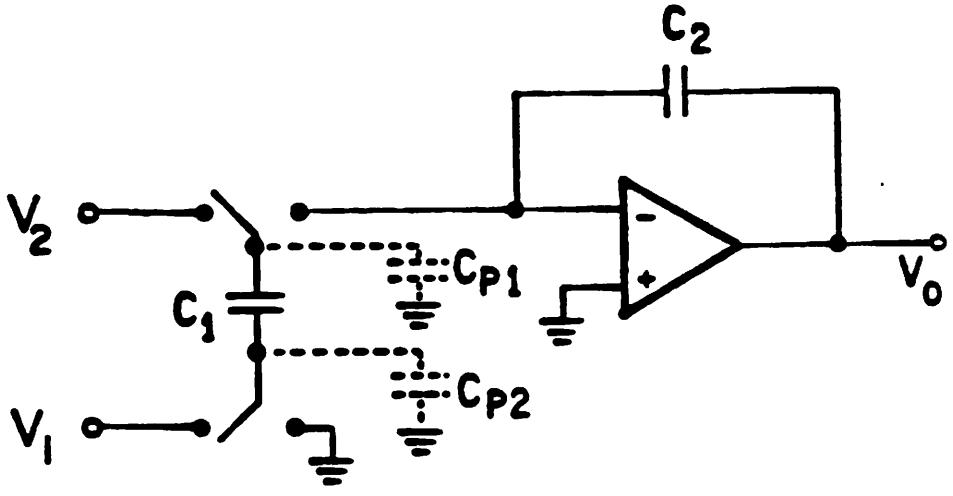


Fig. 2.2 An LDI integrator which is sensitive to parasitics.

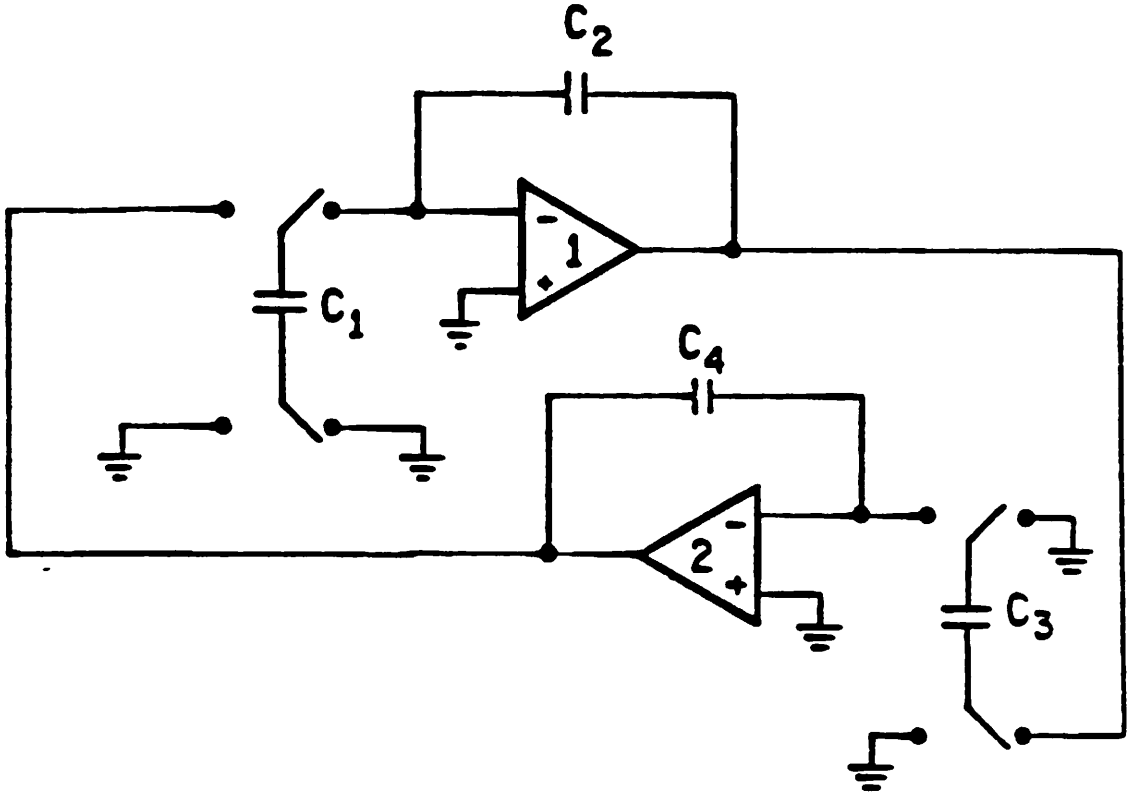


Fig. 2.3 A two-integrator loop with proper switch phasing to realize the LDI transformation.

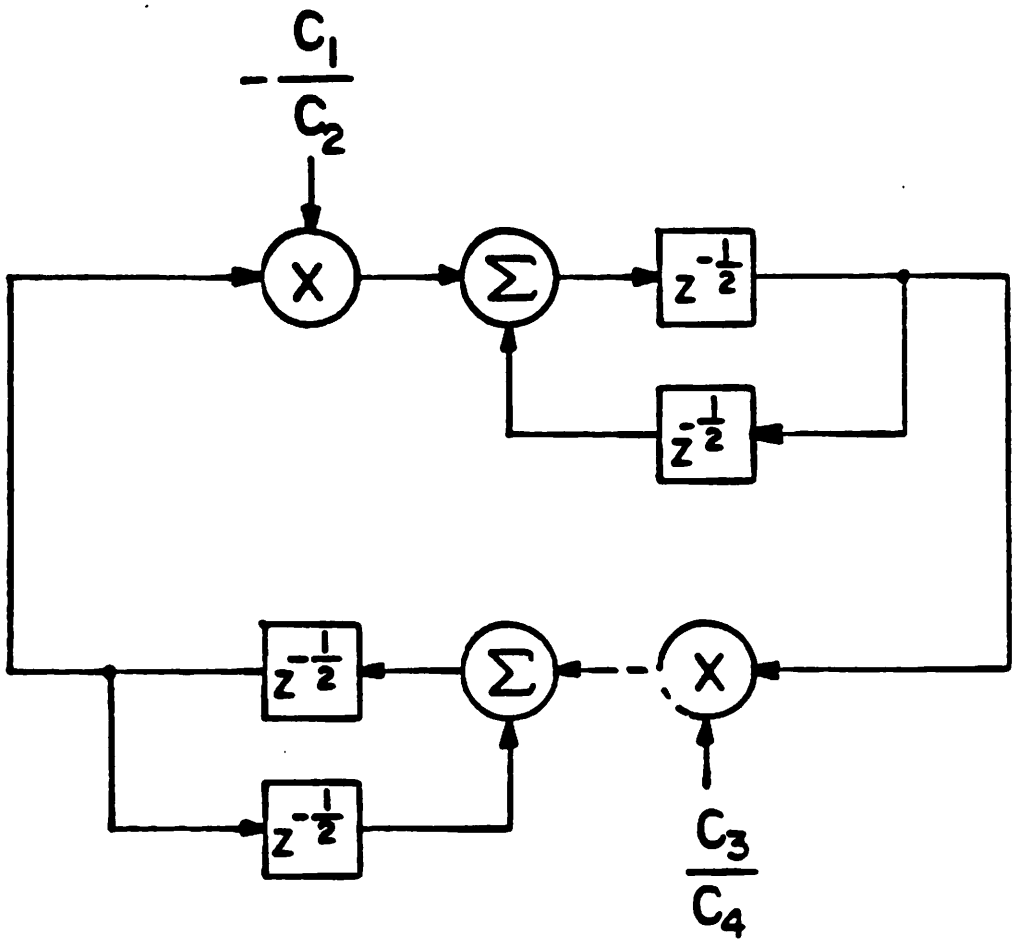


Fig. 2.4 s-domain block diagram for an integrator loop which realizes the LDI transformation.

The problem with this kind of integrator is that it is sensitive to parasitics associated with the top plate of the sampling capacitor (C_{p1} in Fig. 2.2). The nature of parasitic capacitances in switched-capacitor networks have been thoroughly discussed [1,3], and parasitic-free integrators have been developed [14]. Their details will not be elaborated here. Instead, we want to demonstrate that connecting these integrators in a ladder structure will still realize the LDI transformation.

Fig. 2.5 shows a parasitic-free LDI integrator, where V_1 is the non-inverting input and V_2, the inverting input. Its transfer function can be described by

$$V_o(z) = \frac{C_1}{C_2} \frac{1}{1-z^{-1}} (z^{-1/2} V_1(z) - V_2(z)) . \tag{2.13}$$

Notice that the non-inverting input has a half cycle delay to the output whereas the inverting input is delay free. For leapfrog filters, it is almost always true that in any loop (except one including the terminations), there is an equal number of inverting and non-inverting integrators. To realize the LDI transform, proper phasing of the switches is needed. For a two integrator loop shown in Fig. 2.6, it takes half a cycle delay to propagate the signal from the non-inverting input of integrator 2 to the op. amp.output. There is no delay involved to travel from the inverting input of integrator 1 to the op. amp. output, but the signal from integrator 2 is not sampled until half a cycle later. As before, it takes one full cycle to go around the loop, realizing the LDI transform. The z-domain block diagram for this circuit is shown in Fig. 2.7, which is identical to the one shown in Fig. 2.4.

An opposite phasing scheme shown in Fig. 2.8 can also achieve one delay per loop. Although both op. amp. outputs change at the same time, the signal from output of op. amp. 1 is not sampled until half a cycle later. It takes an additional half delay to go from the non-inverting input of the integrator 2 to the op.

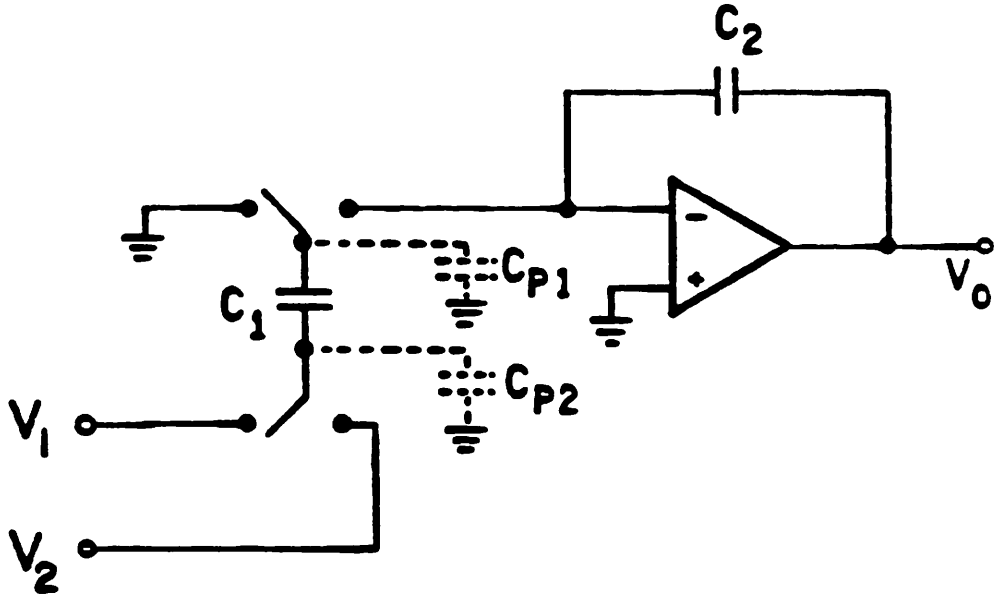


Fig. 2.5 A parasitic-free LDI integrator.

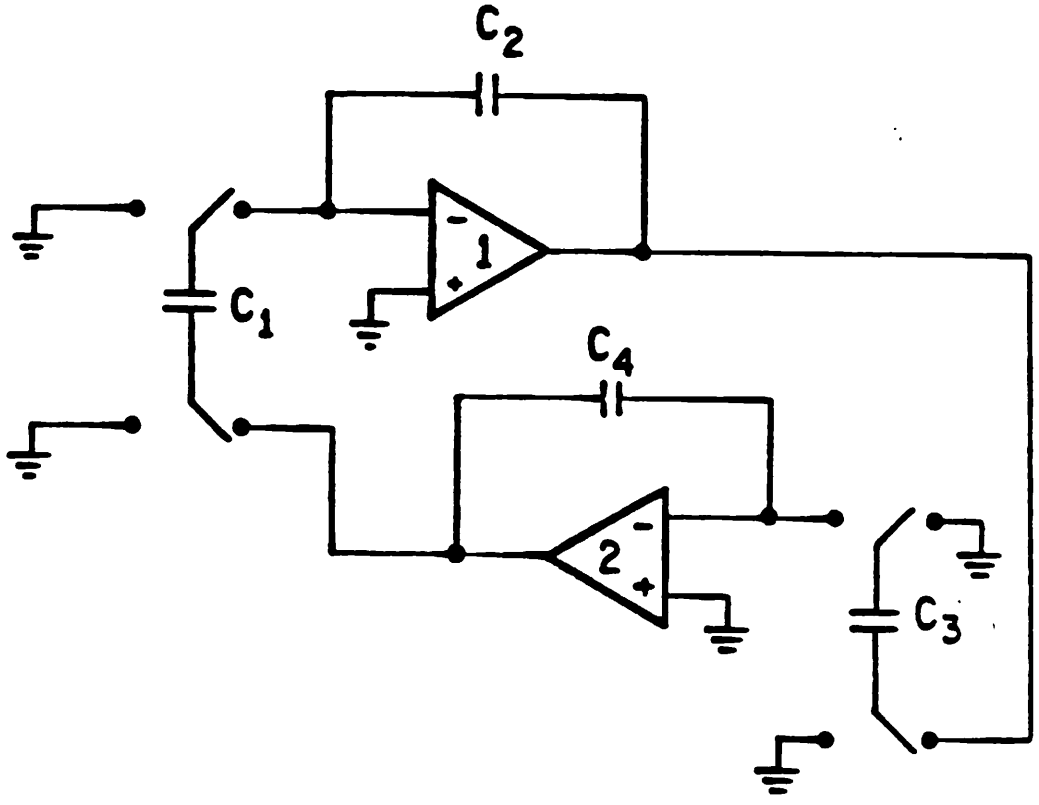


Fig. 2.6 A two-integrator loop with parasitic-free integrators.

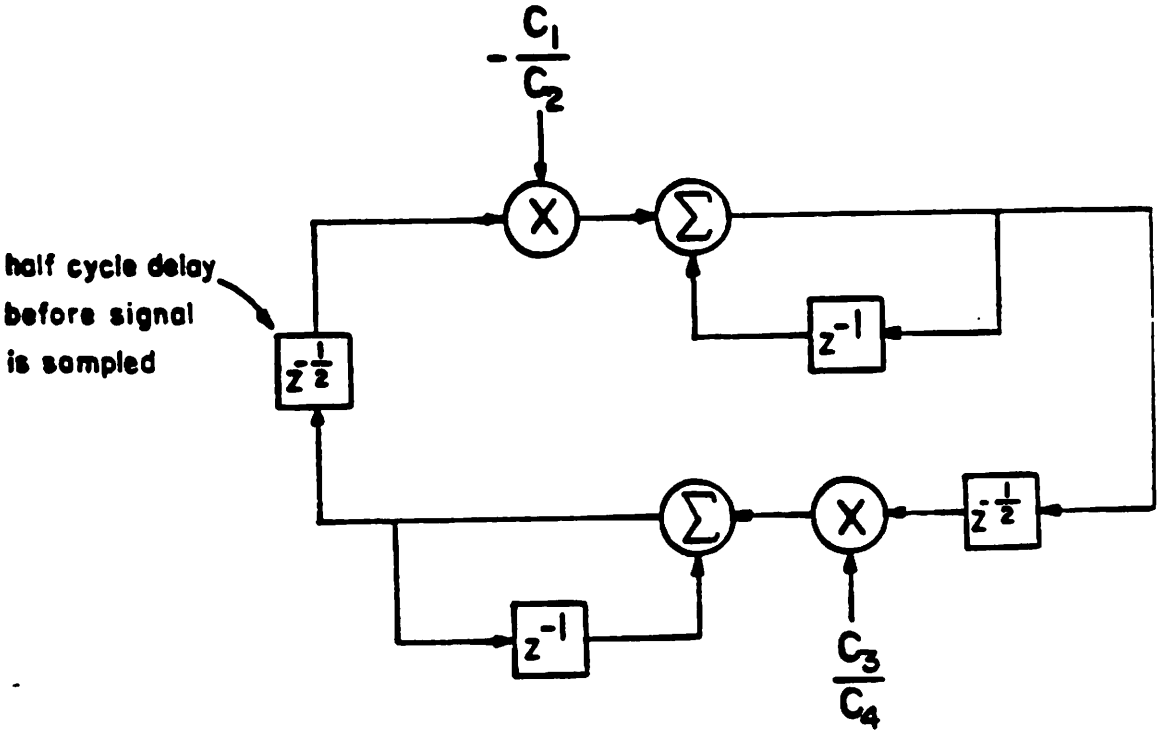


Fig. 2.7 z-domain block diagram of the integrator loop shown in Fig. 2.6.

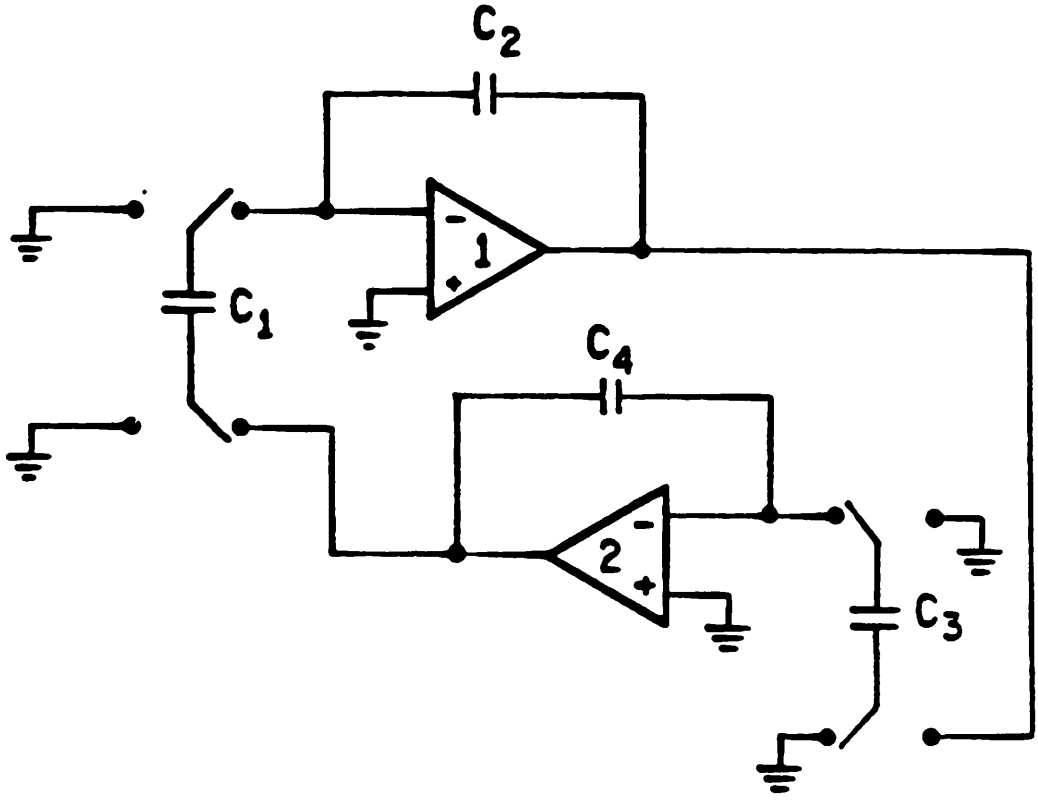


Fig. 2.8 A two integrator loop with opposite phasing to that of Fig. 2.6.

amp. output. Hence, the total delay per loop is still 1 cycle. For all-pole filters, there is no difference between this clocking scheme and that shown in Fig. 2.6. But for elliptic filters, care must be taken with this kind of clocking scheme to make sure that the LDI transformation is realized. In addition, since the outputs of both op. amps. change at the same time, its settling behavior is further complicated. Thus this clocking scheme is not preferred and the one shown in Fig. 2.6 should be used.

The use of fully differential switched-capacitor technique has been described earlier for voice band filters [10]. Although twice as many capacitor and switches are needed, they are more versatile in realizing the LDI transform. Like the single-ended case, it can take half a cycle delay or no delay to go from the sampling inputs to the op. amp. output. However, since the signals are fully differential, we can not distinguish between an inverting or a non-inverting input. A change of sign can be easily achieved by inverting the differential signal. A fully differential integrator is shown in Fig. 2.9. The input V_1 takes a half cycle delay to the output whereas the input V_2 is delay free, and the transfer function is described by

$$V_o(z) = \frac{C_1}{C_2} \frac{1}{1-z^{-1}} (z^{-1/2} V_1(z) + V_2(z)) . \tag{2.14}$$

This realization is totally insensitive to parasitics. To achieve the LDI transform, switch phasings for a two integrator loop is shown in Fig. 2.10. As mentioned before, signal inversion can be done by simply flipping the differential inputs, hence there is no need of the delay free integrator input. The clocking scheme of Fig. 2.10 is identical to that of Fig. 2.3 and the z-domain block diagram is given by Fig. 2.4. Alternately, a phasing scheme similar to that of Fig. 2.6 can also be used by the differential integrators, but delay free paths tend to complicate the settling behavior of the amplifiers. However, these delay

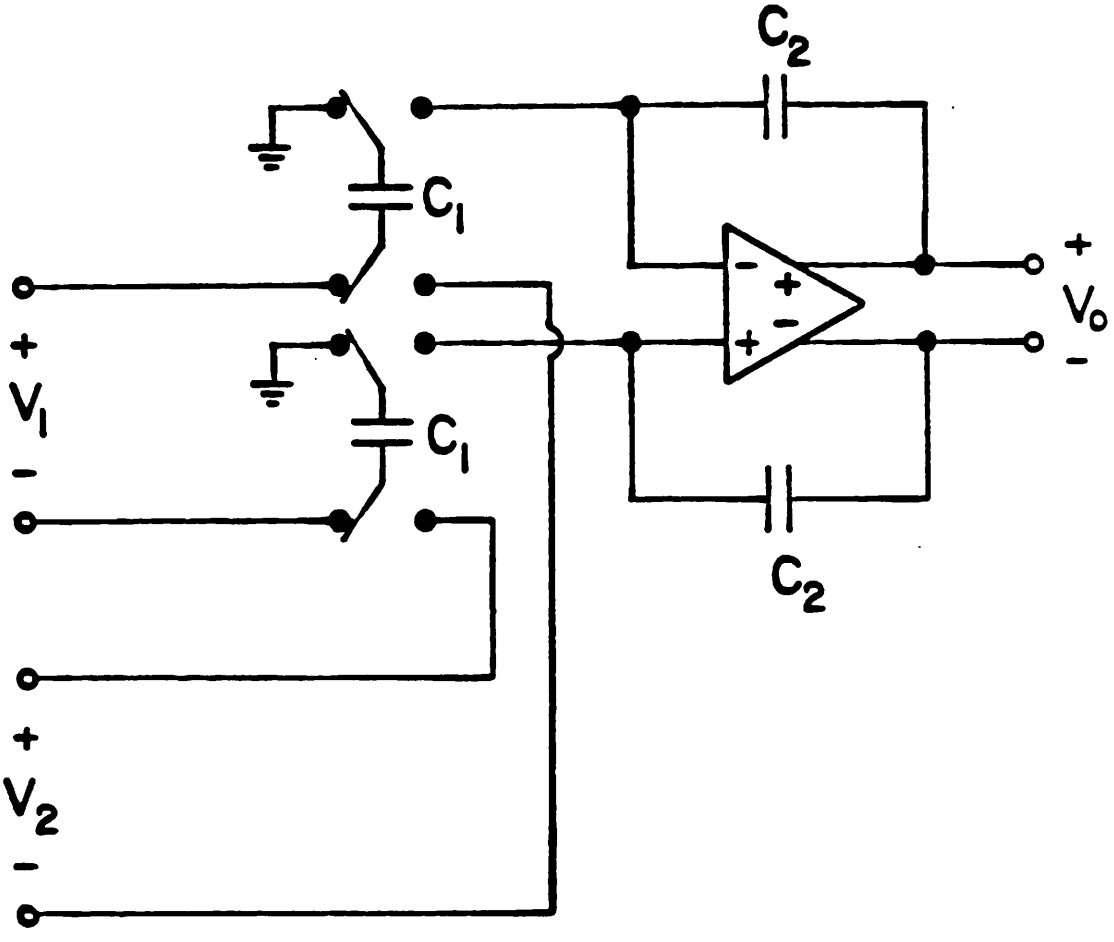


Fig. 2.9 A fully differential LDI integrator.

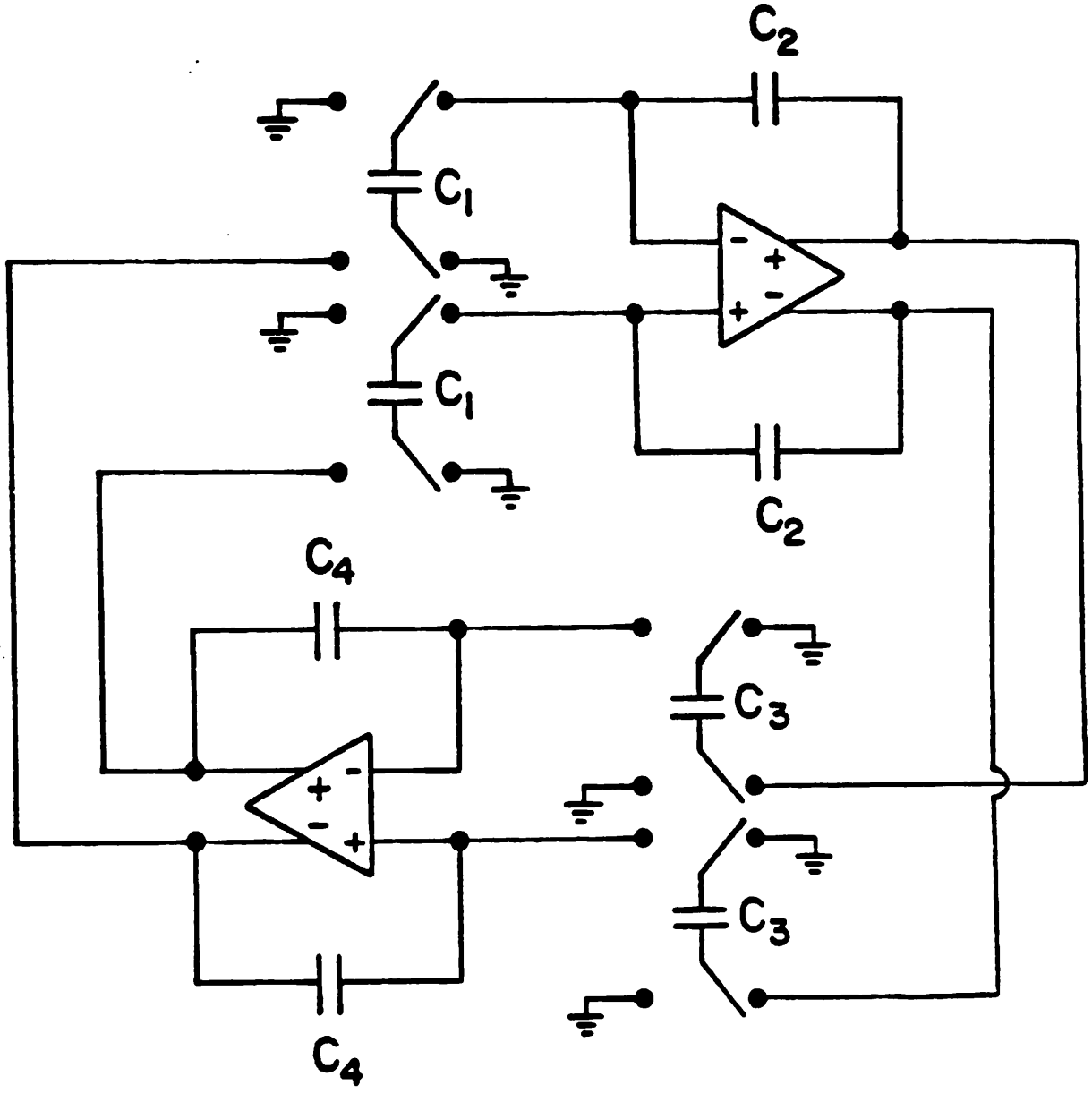


Fig. 2.10 A two-integrator loop with fully differential integrators.

free inputs do provide summing signal paths without extra switches and capacitors. Hence, they can be used if there is a special need to save die area.

2.3 LDI Ladder Terminations

2.3.1 THE LDI TERMINATION WITH AN EXTRA HALF DELAY

The integrators described in the previous section can be used to implement the LDI transformation exactly throughout the ladder. Unfortunately, this is not true for simulation of the resistive terminations.

The load termination of a doubly terminated LC ladder is shown in Fig. 2.11a. Its signal flow graph representation [8,18] is shown in Fig. 2.11b. This can be redrawn as Fig. 2.11c after resistive scaling. Using the LDI transformation as defined by (2.1), the z-domain flow diagram is shown in Fig. 2.11d which then can be represented in block diagram form as in Fig. 2.11e. An ideal LDI termination would then resemble this block diagram which dictates half a cycle delay in the termination feedback loop. Yet, since terminations in the LDI ladder are realized by the type of circuit shown in Fig. 2.12a, it requires one full delay to go around the termination, thus effectively adding half a delay to that loop. The circuit in Fig. 2.12a can be described by the difference equation

$$C_2 V_{out}(nT_s) = C_2 V_{out}[(n-1)T_s] + C_1 V_{in}[(n-\tfrac{1}{2})T_s] - C_3 V_{out}[(n-1)T_s] \quad (2.15)$$

or

$$\frac{V_{out}(z)}{V_{in}(z)} = \frac{C_1}{C_2} \frac{z^{-1/2}}{1 - z^{-1} + \frac{C_3}{C_2} z^{-1}}. \quad (2.16)$$

The z-domain block diagram for (2.16) is shown in Fig. 2.12b, showing an undesirable extra half delay in the termination loop.

The effect of an extra half delay in the termination depends heavily on the sample rate and can best be demonstrated by a specific example. The responses of a fifth order Chebyshev low pass filter are plotted in Fig. 2.13a-c. The nominal filter is designed to have a ripple of 0.5 dB and a ripple bandwidth of

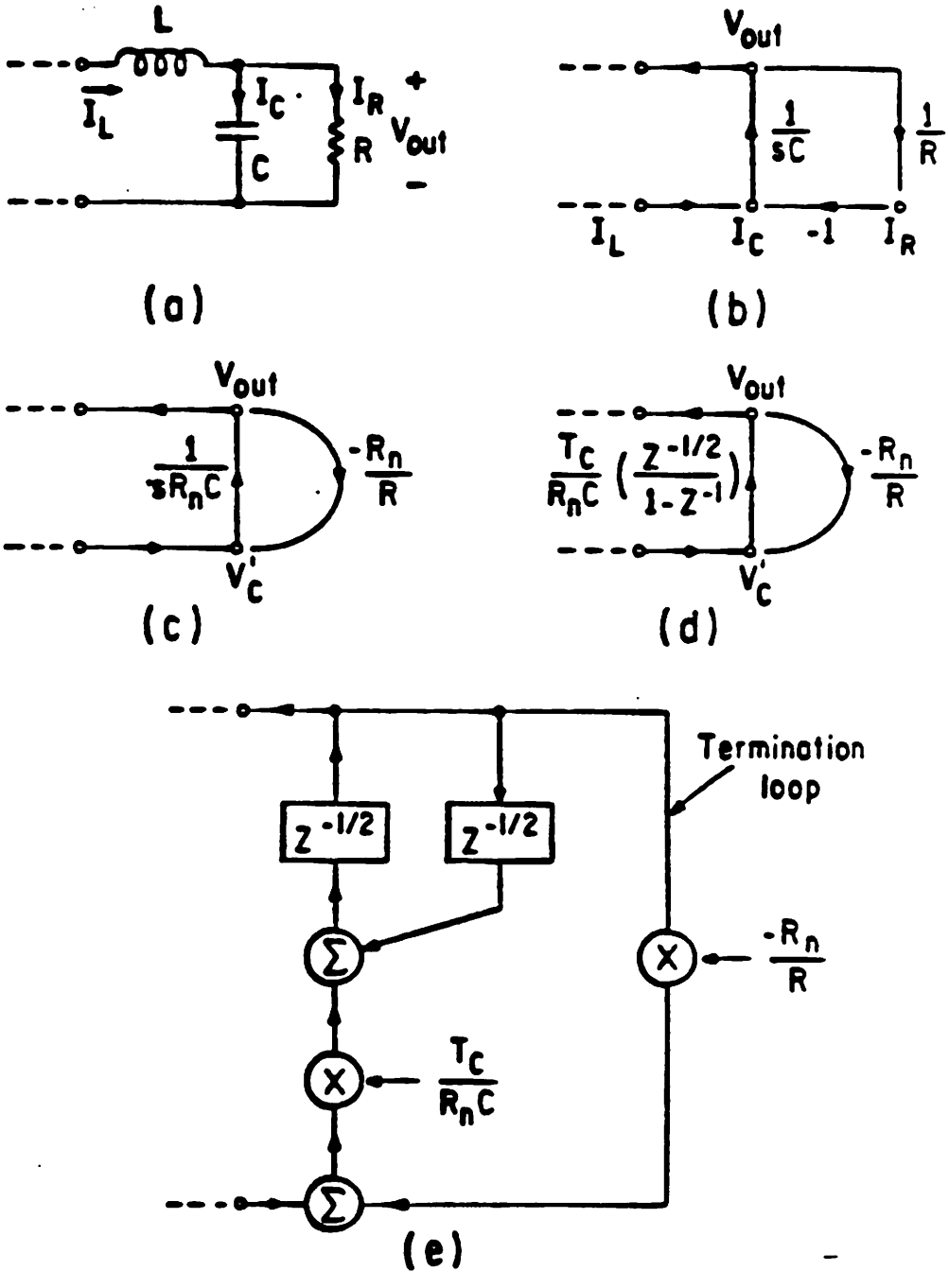


Fig. 2.11 Load termination of a doubly terminated LC ladder.
(a) RLC circuit.
(b) Signal-flow-graph representation.
(c) Resistively scaled flow graph.
(d) z-domain flow graph after LDI transformation.
(e) z-domain block diagram.

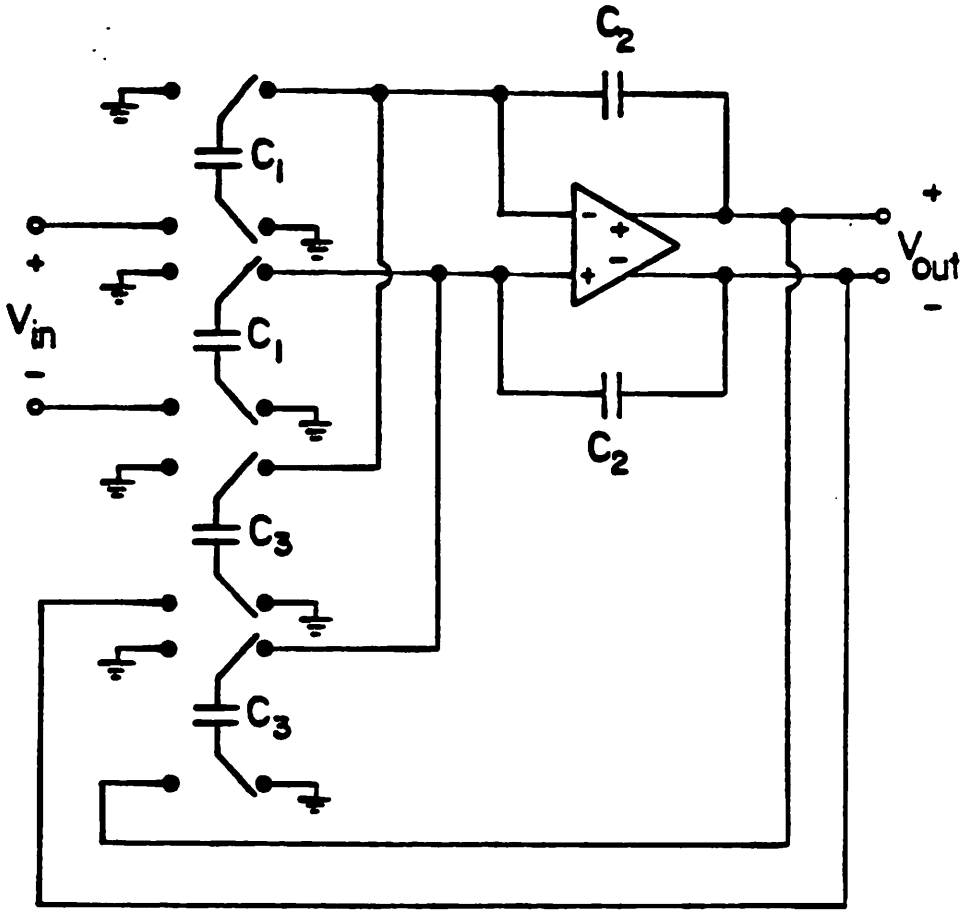


Fig. 2.12 (a) Fully differential LDI termination with an extra half delay.

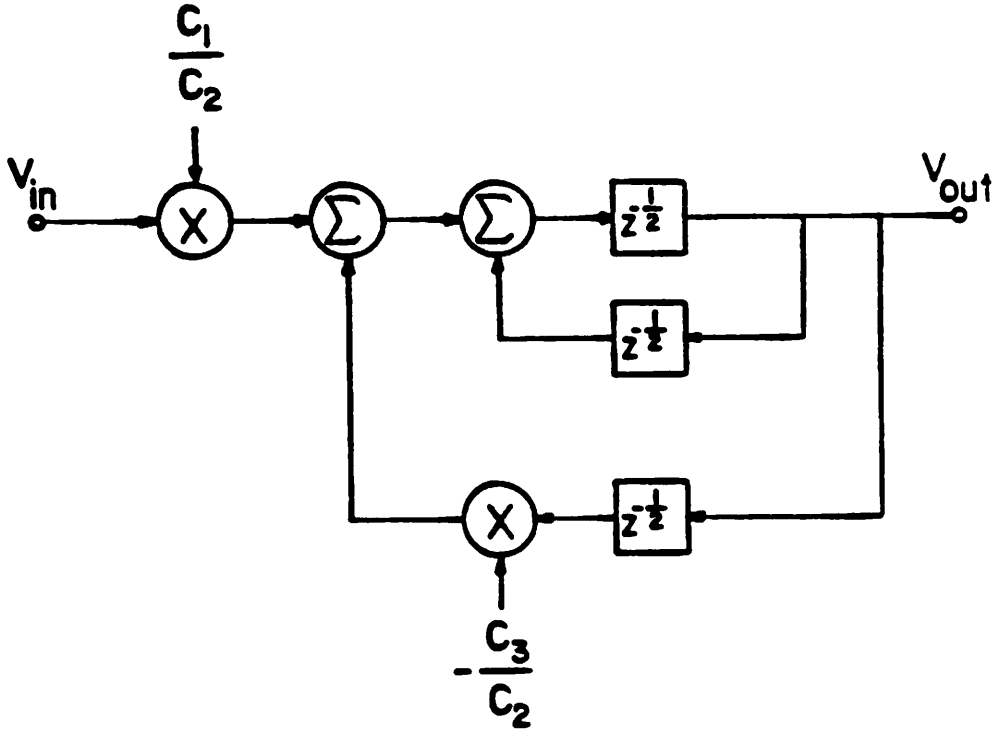


Fig. 2.12 (b) z-domain block diagram representation.

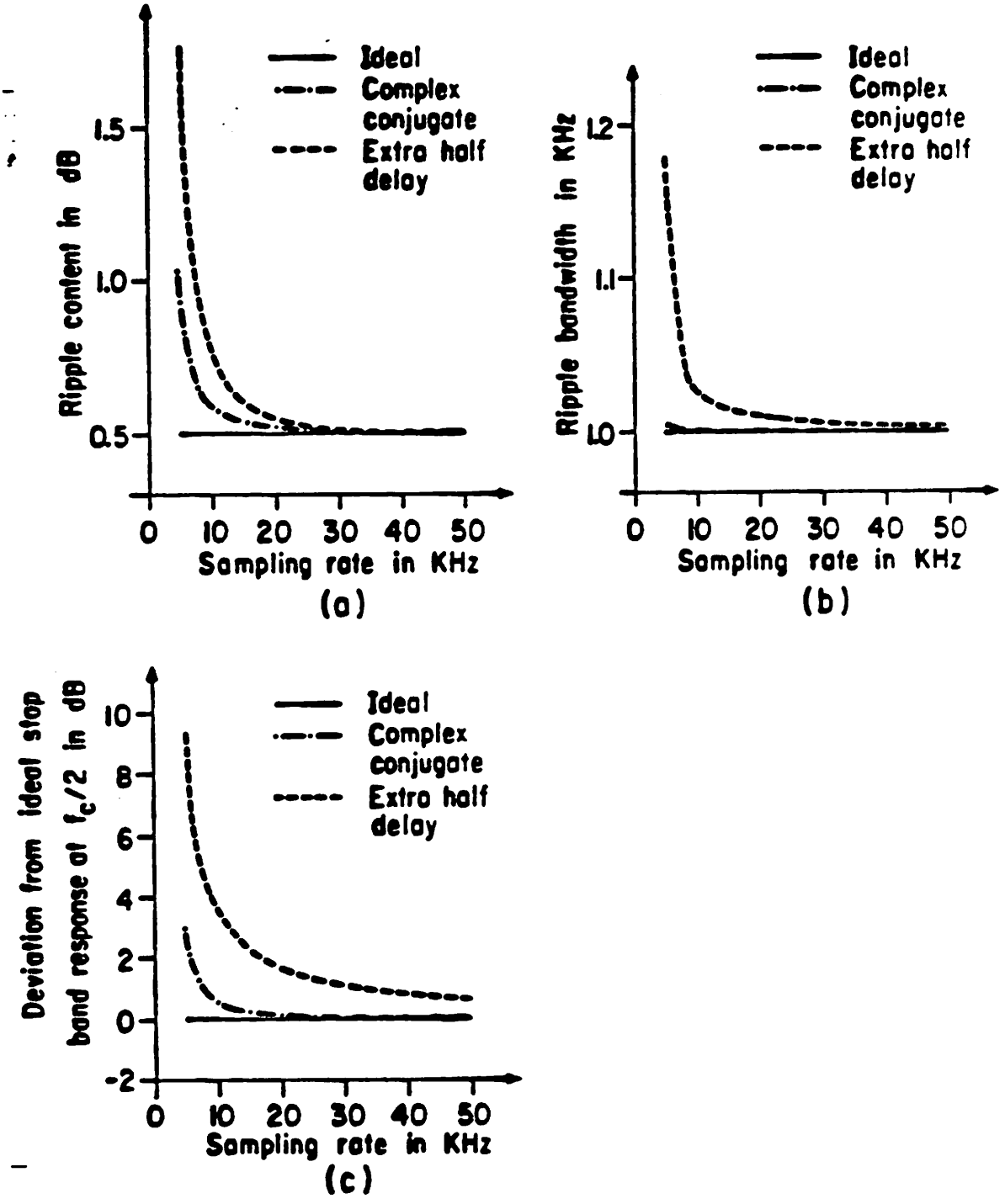


Fig. 2.13 Termination errors as a function of the clock rate for a fifth-order Chebyshev low pass.

(a) Error in ripple content.

(b) Error in ripple bandwidth.

(c) Error in stopband response.

1 KHz. For the case with ideal terminations, the nominal characteristics are achieved independent of the sampling rate. Whereas for the case with an extra half delay in the terminations, we deviate from the nominal curves. With a sampling rate only five times the passband edge, we have a ripple content of 1.76 dB, a bandwidth of 1.17 KHz and a stopband rejection with 9.5 dB less than that for the nominal case.

For low frequency filters, a sample rate which is only five times the cutoff frequency is rarely used because of anti-alias considerations. Typical circuits use a clock rate which is about 40 times that of the frequencies of interest [2]. In these cases, termination error can simply be ignored, as we can see from Fig. 2.13a-c. However, at high operating frequencies, it is not possible to use such a high ratio and for these circumstances, the termination error has introduced a significant deviation of the responses.

2.3.2 THE COMPLEX-CONJUGATE TERMINATION

The termination problem arises from the fact that feedback loops in a circuit are either delay free or it takes a full clock cycle to go around a loop. There simply is no "half-a-clock-cycle" delay loop. However, delay free loops give us an alternative to a full clock cycle delay in termination realizations. If we use a delay free loop, we have a minus one-half delay instead of an extra half delay in the termination. The circuit which realizes this is shown in Fig. 2.14a. This circuit is described by

$$C_2 V_{out}(nT_s) = C_2 V_{out}[(n-1)T_s] + C_1 V_{in}[(n-\tfrac{1}{2})T_s] - C_3 V_{out}(nT_s) \qquad (2.17)$$

or

$$\frac{V_{out}(z)}{V_{in}(z)} = \frac{C_1}{C_2} \frac{z^{-1/2}}{1 - z^{-1} + \frac{C_3}{C_2}} \qquad (2.18)$$

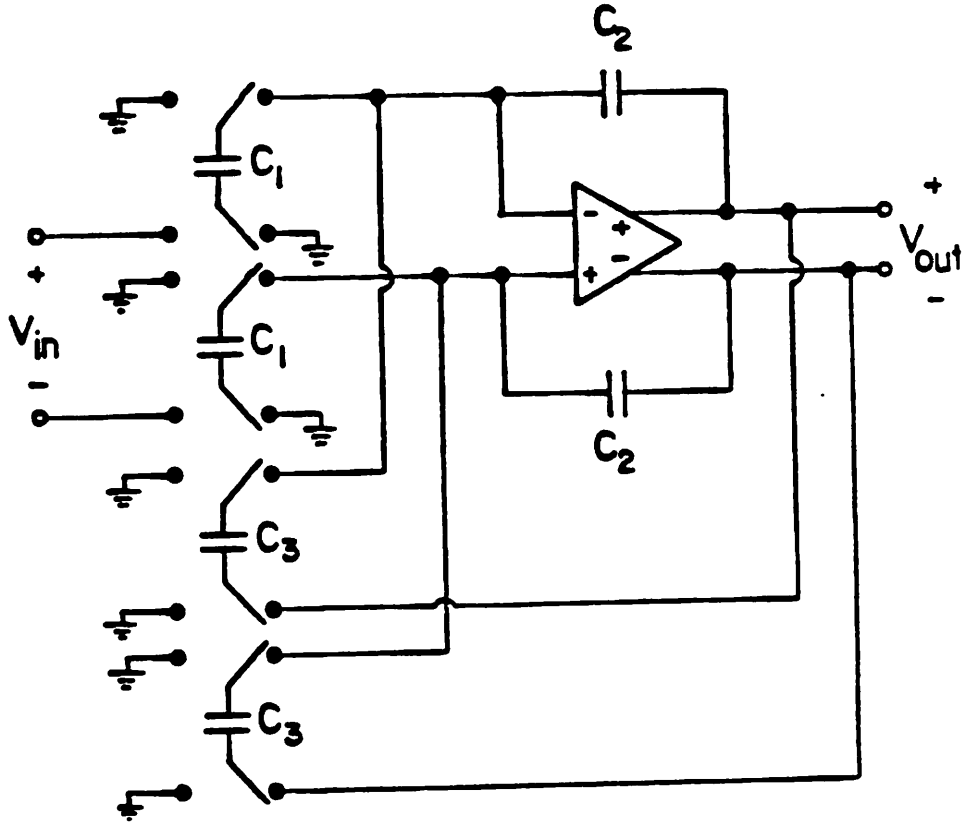


Fig. 2.14 (a) Fully differential LDI termination with a minus one-half delay.

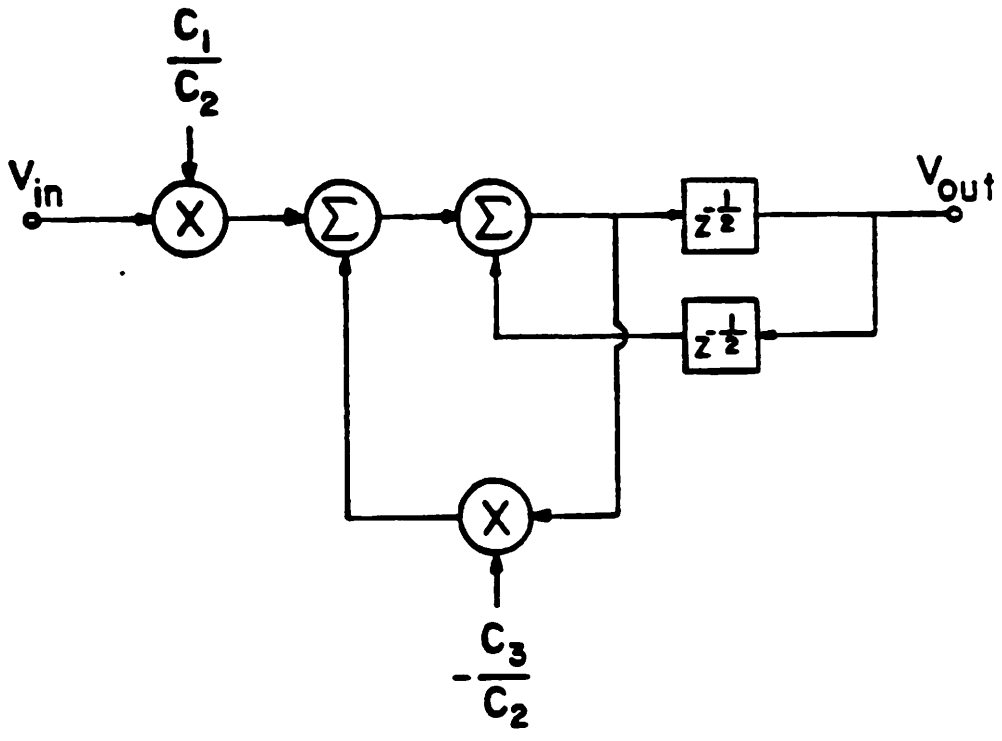


Fig. 2.14 (b) z-domain block diagram representation.

The z-domain block diagram for (2.18) is shown in Fig. 2.14b, showing a minus one-half delay in the termination loop.

The difference between the three different terminations can be clearly seen from their respective z-domain block diagrams (Figs. 2.11e, 2.12b, and 2.14b). For the ideal case, we have a multiplication constant $K = \frac{R_n}{R}$ in the termination loop. (Here, R_n is the resistive scaling factor, R is the termination resistance.) For the case with an extra half delay, we have $Kz^{-1/2}$ whereas for the case with a minus one-half delay, we have $Kz^{+1/2}$. For $z = e^{j\omega T_s}$, $Kz^{-1/2}$ and $Kz^{+1/2}$ are complex-conjugates of each other. From classical theory of doubly terminated networks, we would expect the error for a doubly terminated ladder with complex-conjugate terminations to be less than that for the case when both source and load terminations are not complex-conjugates. To understand this more clearly, let us consider the s-domain description of these non-ideal LDI termination errors.

An extra half delay in the termination path means that we have $Kz^{-1/2}$ in the termination feedback loop instead of K. Hence, we would like to know the s-domain description for $z^{-1/2}$ in terms of the LDI termination. This can be found by inverting (2.1) to obtain

$$s^{-1/2} = \left[1 + \frac{s^2 T_s^2}{4}\right]^{1/2} - \frac{sT_s}{2}. \tag{2.19a}$$

Thus depending on whether $K = \frac{R_n}{R}$ or $\frac{R}{R_n}$ (where R_n = resistive scaling factor, R = termination resistance), the s-domain counterpart for the termination with an extra half delay is a conductance of

$$\frac{1}{R\left[\left[1 + \frac{s^2 T_s^2}{4}\right]^{1/2} - \frac{sT_s}{2}\right]}$$

or a resistance of

$$R\left[\left[1 + \frac{s^2T_s^2}{4}\right]^{1/2} - \frac{sT_s}{2}\right],$$

respectively.

Similarly, for the case of a minus one half delay, we have $K = z^{+1/2}$ and using (2.1) as before

$$z^{+1/2} = \left[1 + \frac{s^2T_s^2}{4}\right]^{1/2} + \frac{sT_s}{2}. \tag{2.19b}$$

Again, depending on whether $K = \frac{R_n}{R}$ or $\frac{R}{R_n}$, the simulated termination in terms of the LDI transformation is

$$\frac{1}{R\left[\left[1 + \frac{s^2T_s^2}{4}\right]^{1/2} + \frac{sT_s}{2}\right]}$$

or

$$R\left[\left[1 + \frac{s^2T_s^2}{4}\right]^{1/2} + \frac{sT_s}{2}\right].$$

Comparing (2.19a) and (2.19b) we notice that for $s = j\omega$, the two terminations, namely the extra half delay and the minus one half delay, are complex-conjugates of each other. Consider the circuit shown in Fig. 2.15a. A matched condition exists when the load impedance Z_L is the complex-conjugate of the source impedance Z_S. Under this condition, we have maximum power transferred to the load. If we have a lossless two-port between the two impedances (Fig. 2.15b), the matched condition no longer holds. The impedance seen by the source termination, Z_{in}, is not necessarily a complex-conjugate of Z_S. Nevertheless, if we use a termination with an extra half delay in the source and one with minus one half delay in the load, we would expect our termination error would be smaller compared with the case when both terminations have

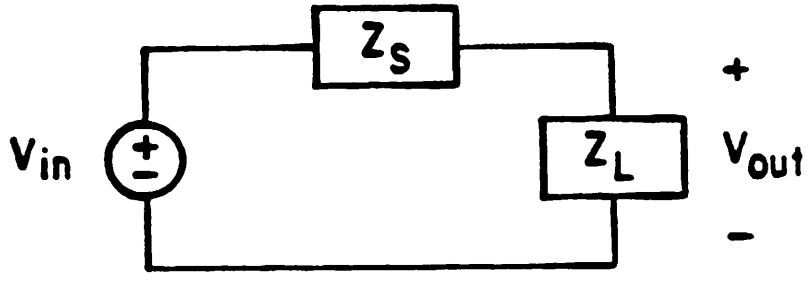


Fig. 2.15 (a) A voltage source (with source impedance) driving a load.

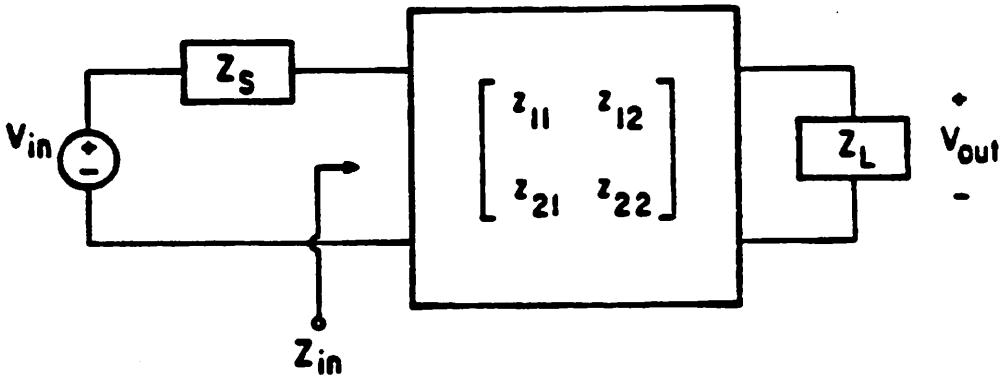


Fig. 2.15 (b) A doubly-terminated two-port network.

extra half delays. To see this more clearly, consider the transfer function for such a circuit will be given by

$$T(s) = \frac{V_{out}}{V_{in}} = \frac{Z_L z_{12}}{Z_S Z_L + z_{11} Z_L + z_{22} Z_S + |Z|} \tag{2.19c}$$

where the z_{ij}'s are the open circuit impedance parameters and $|Z| = z_{11}z_{22} = z_{12}z_{21}$. To simplify the case, choose as an example a symmetric circuit such that $z_{11} = z_{22}$ and let $Z_S = Z_L = R$. Then, for the ideal termination case,

$$T(s) = \frac{R z_{12}}{R^2 + 2z_{11}R + |Z|} \tag{2.19d}$$

$$\left|T(j\omega)\right|^2 = \frac{R^2 |z_{12}|^2}{(R^2 + |Z|)^2 + (2|z_{11}|R)^2}. \tag{2.19e}$$

(For an LC two-port, $z_{ij}(j\omega)$ is purely imaginary, and $Z(j\omega)$ is purely real.) For the case with extra half delays in the terminations, assuming $K = \frac{R}{R_n}$ such that

$$Z_L = Z_S = R\left[\left(1 + \frac{s^2T_s^2}{4}\right)^{1/2} - \frac{sT_s}{2}\right],$$

then

$$T'(s) = \frac{R\left[\left(1 + \frac{s^2T_s^2}{4}\right)^{1/2} - \frac{sT_s}{2}\right] z_{12}}{R^2\left[\left(1 + \frac{s^2T_s^2}{4}\right)^{1/2} - \frac{sT_s}{2}\right]^2 + 2z_{11}R\left[\left(1 + \frac{s^2T_s^2}{4}\right)^{1/2} - \frac{sT_s}{2}\right] + |Z|} \tag{2.19f}$$

where $T'(s)$ is the transfer function with extra half delay terminations. Assuming ωT_s small, we can derive

$$\left|T'(j\omega)\right|^2 \approx \left|T(j\omega)\right|^2 \left[1 + \frac{\omega^2 T_s^2 |Z| R^2 \pm 2\omega T_s |z^{11}| R(|Z| - R^2)}{R^2 |z_{12}|^2} \left|T(j\omega)\right|^2\right] \tag{2.19g}$$

For the case with complex-conjugate terminations, we have

$$T''(s) = \frac{R\left[\left(1 + \frac{s^2T_s^2}{4}\right)^{1/2} + \frac{sT_s}{2}\right]z_{12}}{R^2 + 2z_{11}R\left[1 + \frac{s^2T_s^2}{4}\right]^{1/2} + |Z|} \qquad (2.19h)$$

where $T''(s)$ is the transfer function with complex-conjugate terminations. Again assuming ωT_s small, we have

$$\left|T''(j\omega)\right|^2 \approx \left|T(j\omega)\right|^2\left[1 + \omega^2T_s^2\left|\frac{z_{11}}{z_{12}}\right|^2\left|T(j\omega)\right|^2\right]. \qquad (2.19i)$$

From (2.19g) and (2.19i), we see that if $\omega T_s \to 0$, then both $\left|T'(j\omega)\right|^2$ and $\left|T''(j\omega)\right| \to \left|T(j\omega)\right|$, i.e., the error decreases with increasing sample rate, as expected. We also notice that for the case with extra half delay terminations, we have errors on the order of ωT_s whereas for the case of complex-conjugate terminations, we have errors on the order of $(\omega T_s)^2$. Thus we would expect complex-conjugate terminations to yield less error.

For our specific example, this is seen to be true. The responses of our filter example with complex-conjugate terminations are also plotted in Fig. 2.13a-c. We indeed notice a significant improvement in all respects. With a clock rate only ten times the cutoff frequency, there is now a 0.59 dB ripple and the ripple bandwidth and stopband rejection are the same as those for the ideal case. A detail plot of the passband responses for three different terminations with a sample rate of 10 KHz is shown in Fig. 2.15c for better comparison.

Figs. 2.12a and 2.14a show terminations with fully differential integrators. The terminations for the single-ended case is not as straight forward since all termination loops require negative feedback and the inverting input for single-ended integrators are always delay free, thus realizing only minus one-half delay terminations. To realize an extra half delay in the termination loop, a more complicated clocking scheme shown in Fig. 2.16a is required. This circuit can be described by the difference equation

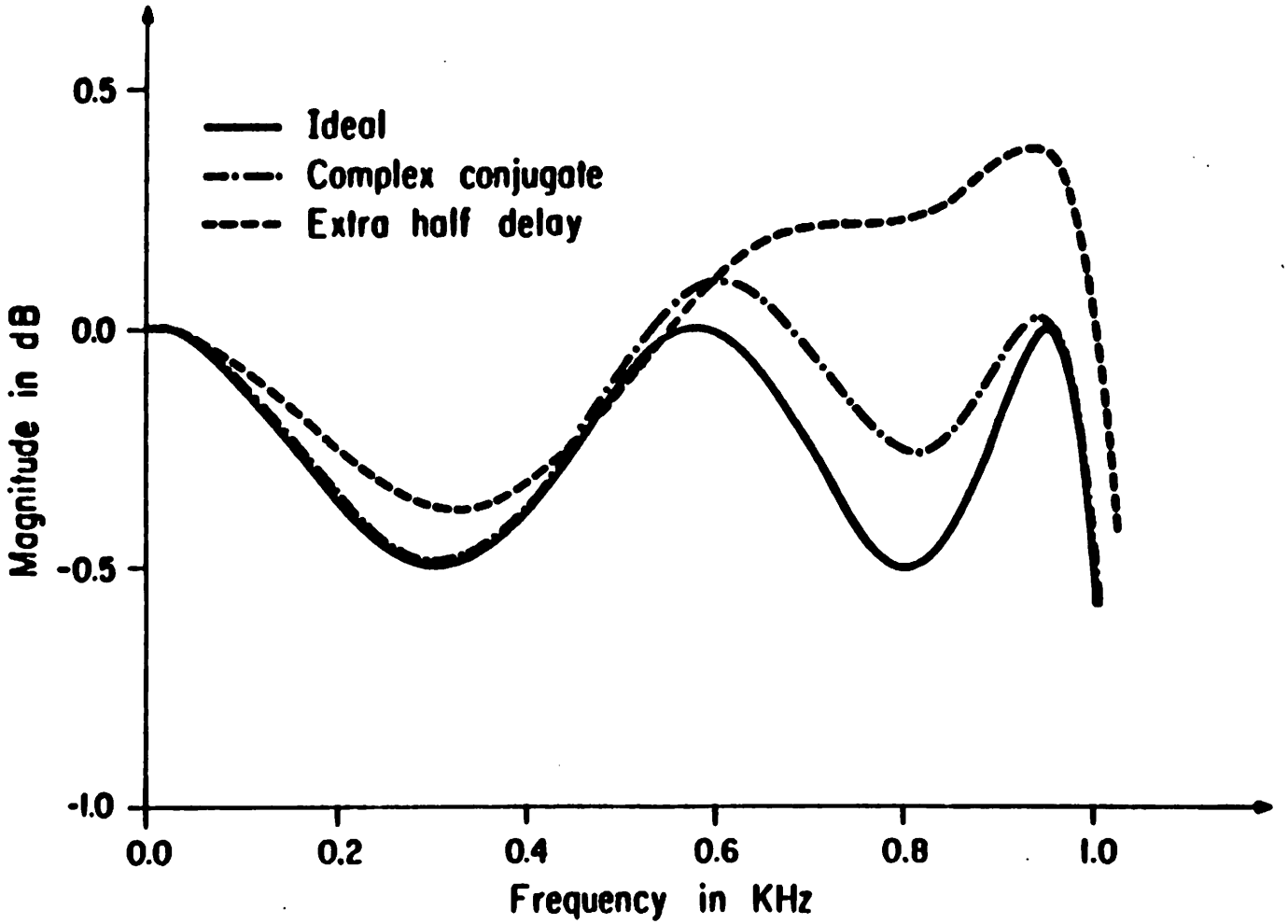


Fig. 2.15 (c) Detail passband response of the fifth order Chebyshev example with a 10 KHz sample rate.

$$C_2 V_{out}(nT_s) = C_2 V_{out}[(n-\tfrac{1}{2})T_s] + C_1 V_{in}[(n-\tfrac{1}{2})T_s] \tag{2.20}$$

but

$$C_2 V_{out}[(n-\tfrac{1}{2})T_s] = \frac{C_2}{C_2+C_3} V_{out}[(n-1)T_s] \tag{2.21}$$

thus

$$C_2 V_{out}(nT_s) = \frac{C_2^2}{C_2+C_3} V_{out}[(n-1)T_s] + C_1 V_{in}[(n-\tfrac{1}{2})T_s] \tag{2.22a}$$

which in z-transform notation is

$$\frac{V_{out}(z)}{V_{in}(z)} = \frac{C_1}{C_2} \frac{z^{-1/2}}{1-z^{-1}+\frac{C_3}{C_2+C_3}z^{-1}}. \tag{2.22b}$$

The z-domain block diagram for (2.22b) is shown in Fig. 2.16b, showing an extra half delay in the termination loop.

The case for minus one-half delays with single-ended integrators is similar to that for fully differential integrators. This circuit is shown in Fig. 2.17a and is also described by equations (2.17) and (2.18). The corresponding z-domain block diagram is shown in Fig. 2.17b, which is identical to Fig. 2.14b.

It should be noted that the circuits shown in Figs. 2.12a, 2.14a, 2.16a, and 2.17a are all insensitive to parasitic capacitances. The parasitic on the input side of C_3 is switched between grounds and that on the output side is switched between voltage sources. The "parasitic-free" properties of the LDI integrators are therefore carried over to the terminations, and thus the circuits can be realized with minimum sized capacitors resulting in maximum possible sampling rates.

For fully differential ladders, it does not matter how we place the complex-conjugate terminations, as long as we put either one (whether extra half delay or minus one-half delay) in the source and the other in the load, they will always

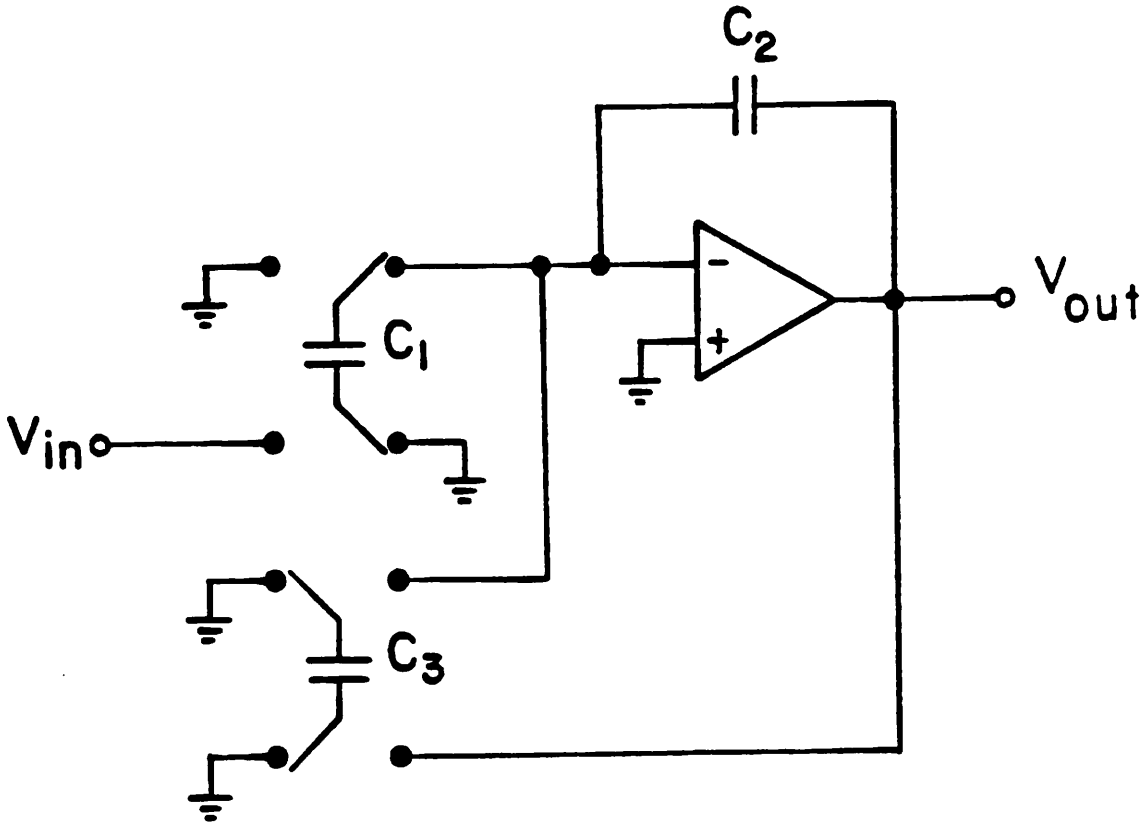


Fig. 2.16 (a) Single-ended LDI termination with an extra half delay.

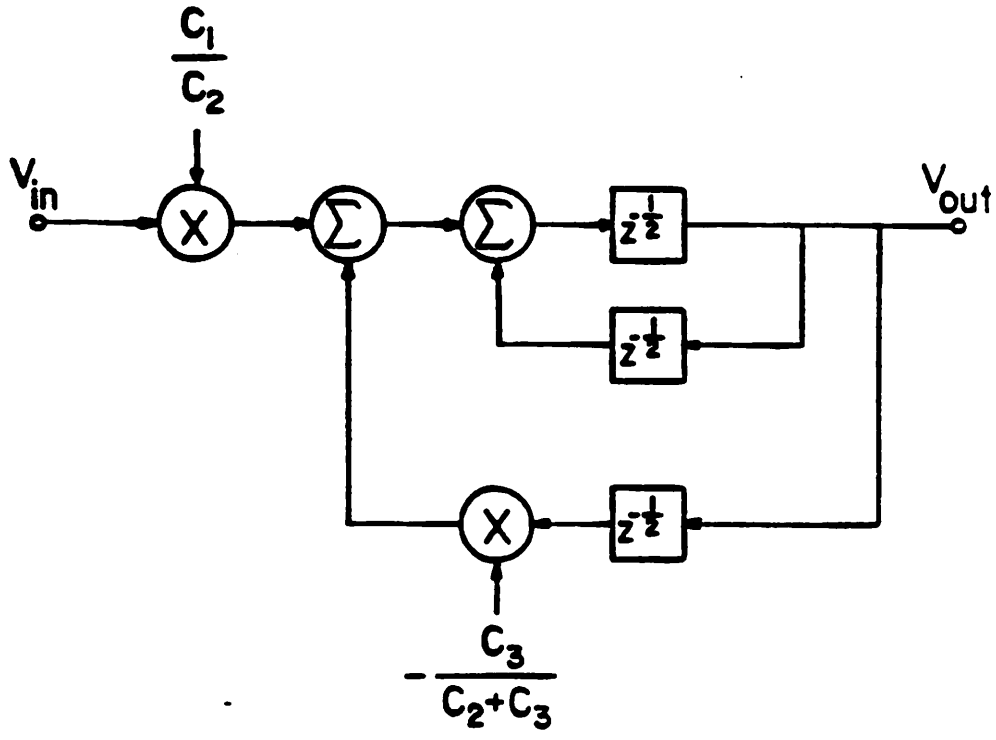


Fig. 2.16 (b) z-domain block diagram representation.

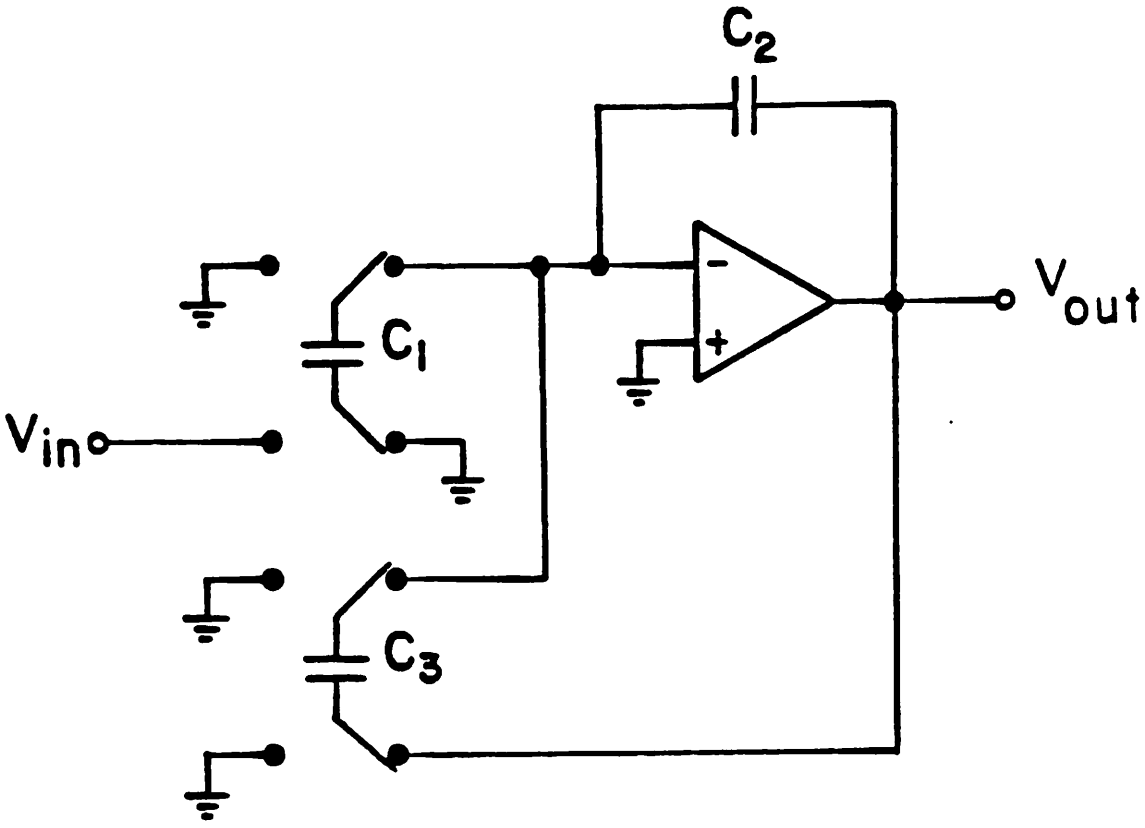


Fig. 2.17 (a) Single-ended LDI termination with a minus one-half delay.

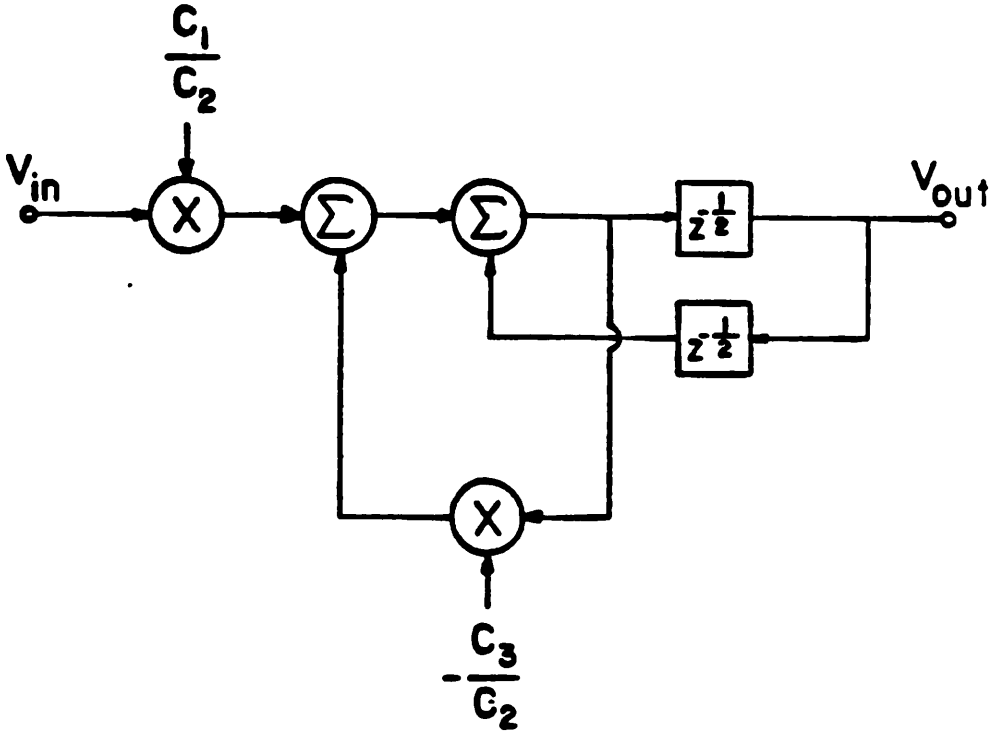


Fig. 2.17 (b) z-domain block diagram representation.

be complex-conjugates of each other. However, for single-ended integrators, we notice that the clocking scheme for Fig. 2.16a is such that we have charge sharing between C_2 and C_3 at the half cycle when the output of the op. amp. is not sampled. Therefore, we should use the termination with an extra half delay (Fig. 2.16a) at the source so that the filter output will only change once every clock cycle. As an illustration for a "parasitic-free" LDI ladder (with single-ended op. amps.), the complete circuit for the fifth order low pass filter example using parasitic free circuits and complex-conjugate terminations is shown in Fig. 2.18.

It should be noted that the output of Fig. 2.16a changes on every half cycle, but as long as it is not sampled during that period, as in the case for the all-pole low pass example (Fig. 2.18), this will not affect filter response. However, when there are continuous feed-forward paths such as in elliptic filters, care must be taken to achieve LDI transformation. On the other hand, for bandpass LDI ladders, both minus one-half delay and extra half delay terminations can be realized without the complex clocking scheme of Fig. 2.16a. This is because bandpass leapfrog structures have localized feedback loops which are not shared by the other parts of the circuit. Thus, the extra half delay termination can be accomplished by inverting the signal through the local loop and fed back via the non-inverting input, as illustrated in Fig. 2.19.

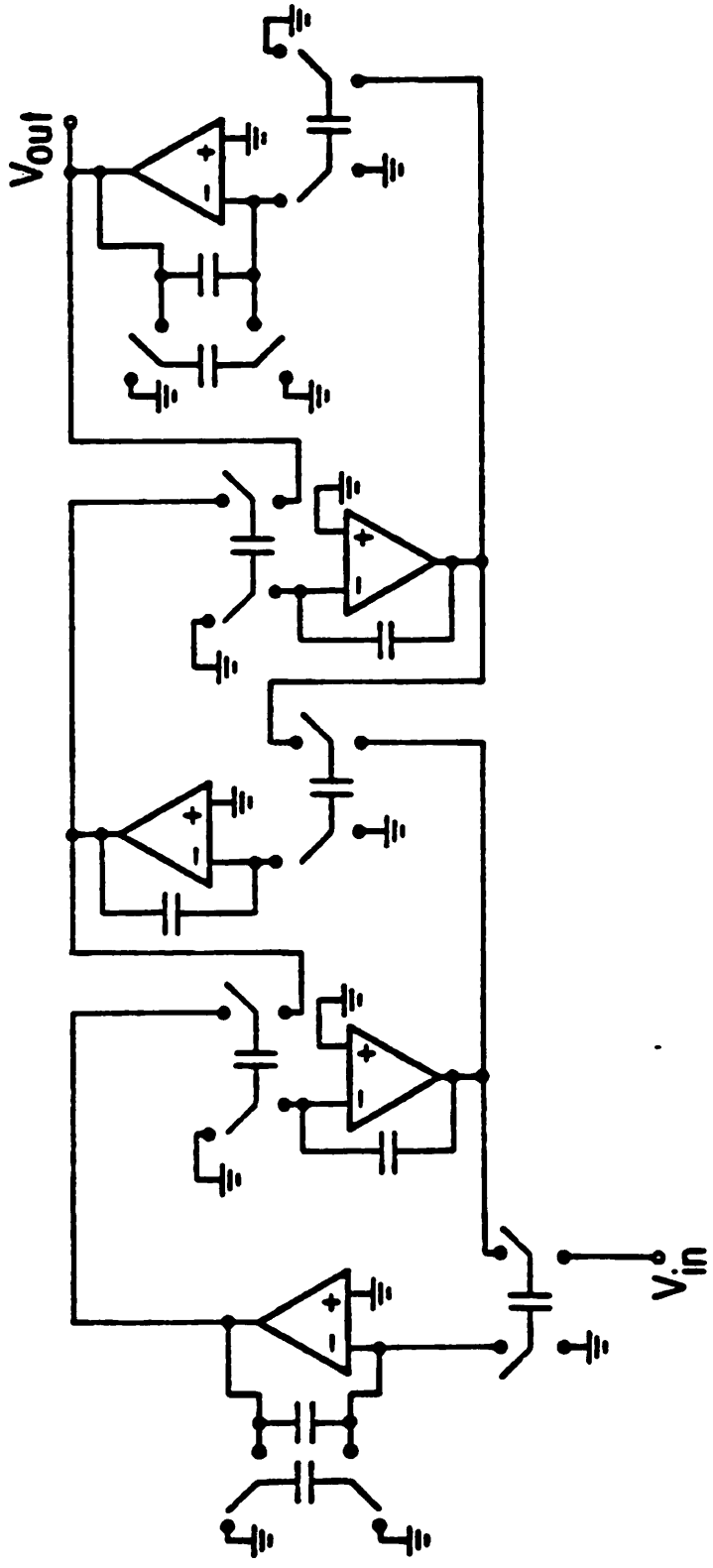


Fig. 2.18 A fifth order all-pole low pass ladder with complex-conjugate terminations (single-ended op. amp. case).

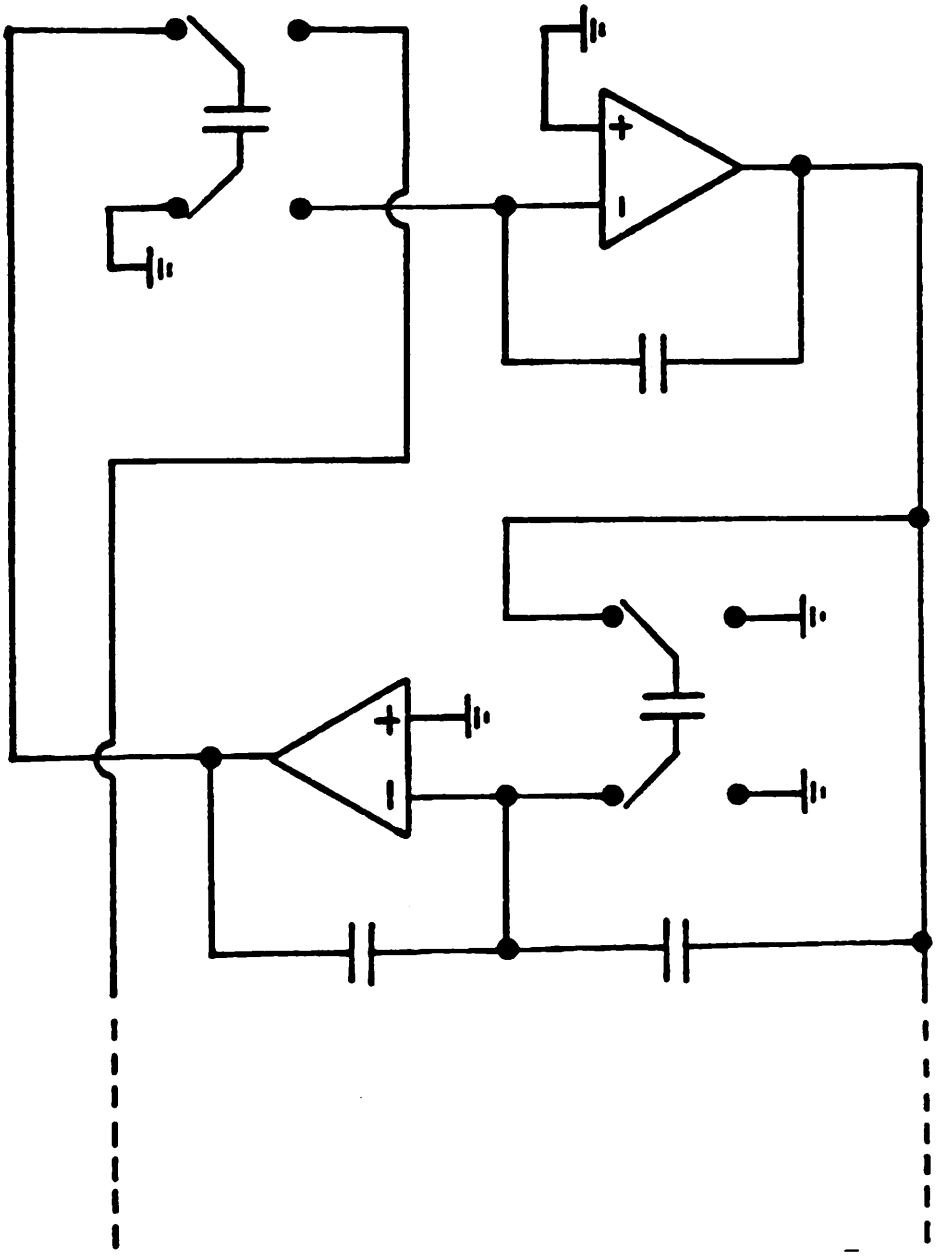


Fig. 2.19 Termination with an extra half delay in bandpass LDI ladders.

2.4 D.C. Stability in Elliptic Bandpass Ladder Filters

Elliptic bandpass filters are very useful in realizing narrow band filters because of the higher selectivity at the band edges obtained from the transmission zeros. However, implementation of elliptic bandpass filters in the leapfrog configuration can result in unstable D.C. conditions in the circuit. This problem arises from the presence of inductor loops in the LC ladder (capacitor loops may have similar problems), and is not unique to switched-capacitor filters. To see this, consider the circuits in Fig. 2.20 where an RC integrator is used to simulate the $i-v$ characteristics of an inductor. Here, active RC integrators are used for demonstration purposes, the same reasoning applies to switched capacitor integrators. Assuming the op. amp. is ideal, the transfer function of the integrator is given by

$$\frac{V_{out}(s)}{V_{in}(s)} = -\frac{1}{sRC}. \tag{2.23}$$

The $i-v$ characteristics of an inductor is given by

$$\frac{I(s)}{V(s)} = \frac{1}{sL}. \tag{2.24}$$

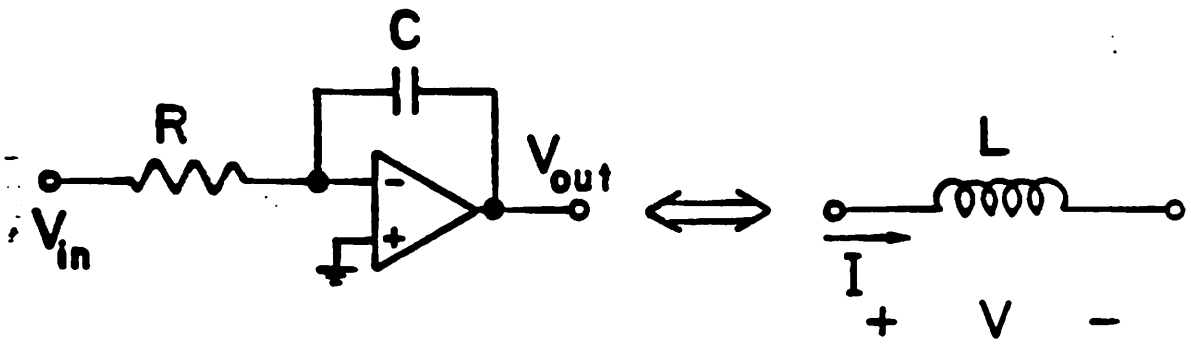
If $I(s)$ is resistively scaled to $V(s)$, then

$$\frac{V'(s)}{V(s)} = \frac{I(s)R'}{V(s)} = \frac{R'}{sL} \tag{2.25}$$

where R' is the scaling resistance. Equations (2.23) and (2.25) show that with ideal op. amp., an inductor can be exactly simulated by an active RC integrator, where the RC time constant is related to L by

$$RC = \frac{L}{R'}. \tag{2.26}$$

The minus sign in (2.23) is not important since the inductor current can be defined with a different orientation.



$$V_{out} \Leftrightarrow R' I$$

$$\frac{1}{RC} \Leftrightarrow \frac{R'}{L}$$

$$V_{in} \Leftrightarrow V$$

where R' = scaling resistance

Fig. 2.20 Simulation of an inductor $i-v$ characteristics by an active RC integrator.

If there is an offset voltage V_{os} in the op. amp. (Fig. 2.21a), then

$$\frac{V_{in}(s) - V_{os}}{R} = sC(V_{os} - V_{out}(s)) \tag{2.27}$$

or

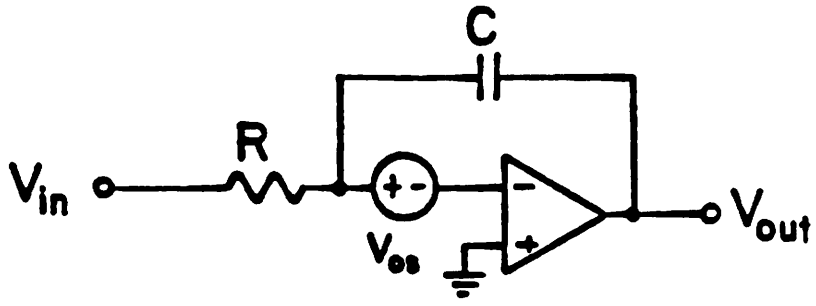
$$V_{out}(s) = -\frac{V_{in}(s) - V_{os}}{sRC} + V_{os} . \tag{2.28}$$

This relationship can be represented by the circuit in Fig. 2.21b. If the small offset voltage at the output is ignored, then

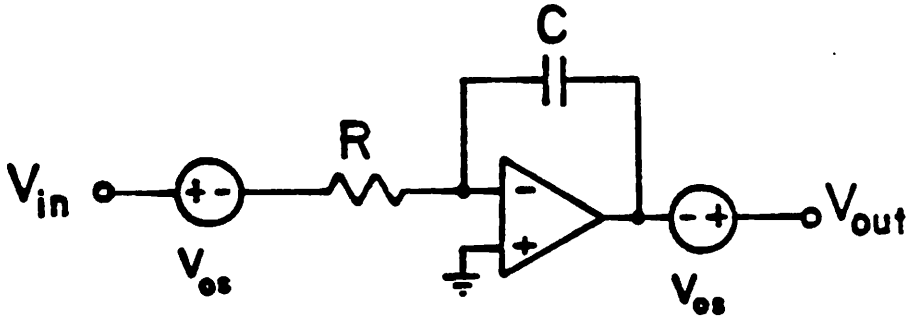
$$\frac{V_{out}(s)}{V_{in}(s) - V_{os}} \approx -\frac{1}{sRC} . \tag{2.29}$$

Comparing (2.29) with (2.25), this would mean that instead of $V(s)$, $(V(s) - V_{os})$ is being integrated. Hence, in addition to simulating an inductor, there is a small voltage source in series which is equal to the offset voltage of the op. amp. (Fig. 2.21c). If such integrators are used to implement an inductor loop as in Fig. 2.22a, it will result in the circuit in Fig. 2.22b where there are offset voltage sources in series with the inductors. It is very likely that these offsets do not add up to zero, then presence of a D.C. voltage will give infinite current in the inductors. In terms of the active equivalent circuit, there will be an op. amp. driven by a voltage equal to $A(\sum_i V_{osi})$ where A is the op. amp. gain. Hence, if the offset voltages do not sum to zero, then the op. amp. will be driven into saturation.

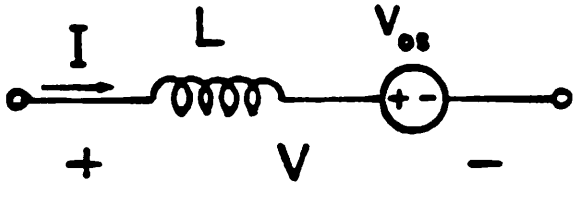
The same problem can be analyzed in the context of the active filter. Consider the inductor loop in Fig. 2.23a, with inductor voltages and currents defined as shown. This can be part of a doubly-terminated LC ladder. The signal flow graph representing their $i-v$ characteristics is shown in Fig. 2.23b. This is realized by active RC integrators as shown in Fig. 2.23c, where V_i's are the integrator inputs, and V_o's are the op. amp. outputs. Consider the D.C. condition of the circuit. In order to keep the output voltage V_{o1} in active region, its input voltage



(a)



(b)



(c)

Fig. 2.21 (a) An active RC integrator with offset voltage in the op. amp.
(b) Equivalent circuit of Fig. 2.21a.
(c) Effect of op. amp. offset voltage in the simulated inductor.

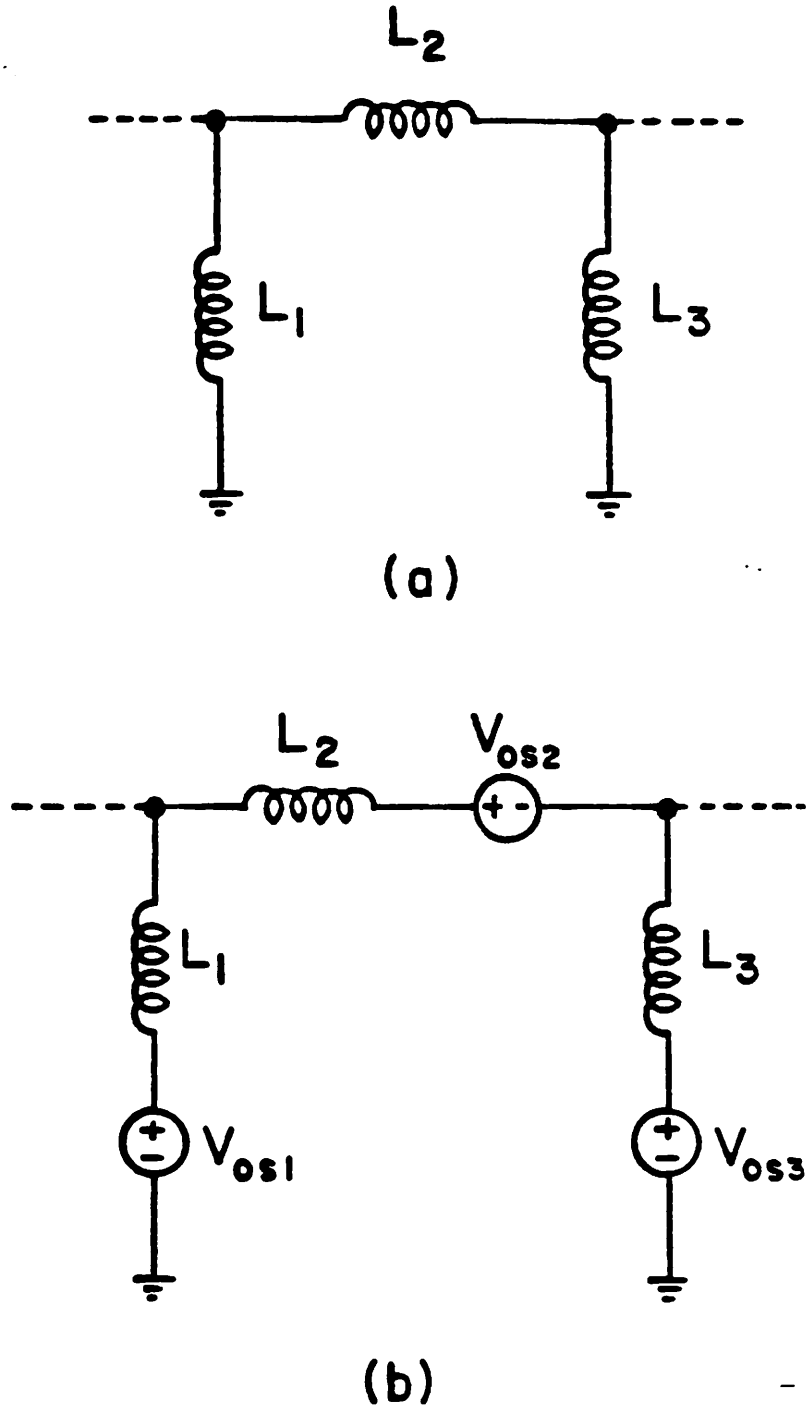
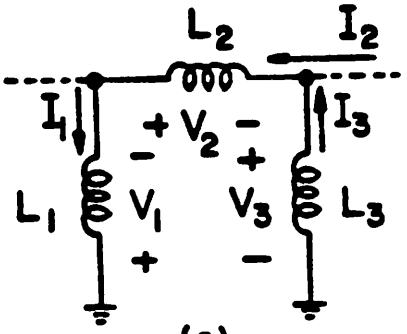
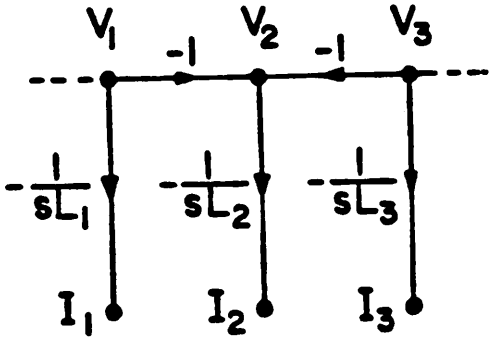


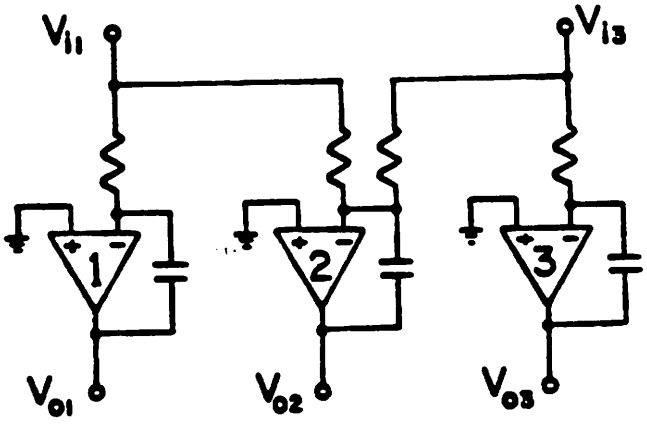
Fig. 2.22 (a) An inductor loop.
(b) An inductor loop simulated by active RC integrators with op. amp. offset voltages.



(a)



(b)



(c)

Fig. 2.23 (a) An inductor loop.
(b) Corresponding signal flow graph.
(c) Realization by active RC integrators.

is given by

$$V_{i1} \approx V_{os1} \tag{2.30}$$

where V_{os1} is the offset voltage in the first op. amp. Likewise,

$$V_{i3} \approx V_{os3} . \tag{2.31}$$

Then

$$V_{o2} = -A_2(V_{os1} + V_{os2} + V_{os3}) \tag{2.32}$$

where A_2 is the gain of the second op. amp. If the offset voltages of the three op. amps. do not add up to zero, then V_{o2} will be driven into saturation.

To solve this problem, the L,C loops in the passive filter can be broken up by Thevenin equivalent circuits using a technique similar to that by Jacobs et al [8] for lowpass filters. Fig. 2.24a shows a doubly-terminated sixth order elliptic bandpass filter. By writing the nodal equation at node 1 , we have

$$I_{R1} - \frac{V_1}{sL_1} - I_{C1} - I_2 - I_{C3} - \frac{V_1 - V_4}{sL_3} = 0 . \tag{2.33}$$

Rewriting (2.33), we have

$$V_1 = (I_{R1} - I_{C1} - I_2 - I_{C3}) \frac{sL_1L_3}{L_1 + L_3} + V_4 \frac{L_1}{L_1 + L_3} . \tag{2.34}$$

Similarly, writing the nodal equation at node 4 will give

$$V_4 = (I_{C3} + I_2 - I_{C4} - I_{R2}) \frac{sL_3L_4}{L_3 + L_4} + V_1 \frac{L_4}{L_3 + L_4} . \tag{2.35}$$

From (2.34) and (2.35), we see that the inductor L_3 can be represented by two voltage-controlled-voltage-sources, with corresponding change in inductor values, as shown in Fig. 2.24b. Likewise, to remove the capacitor C_3, we wrote the nodal equation at node 1 of Fig. 2.24b, we have

$$I_{R1} - I_x - I_2 - V_1(sC_1) - (V_1 - V_4)sC_3 = 0 \tag{2.36}$$

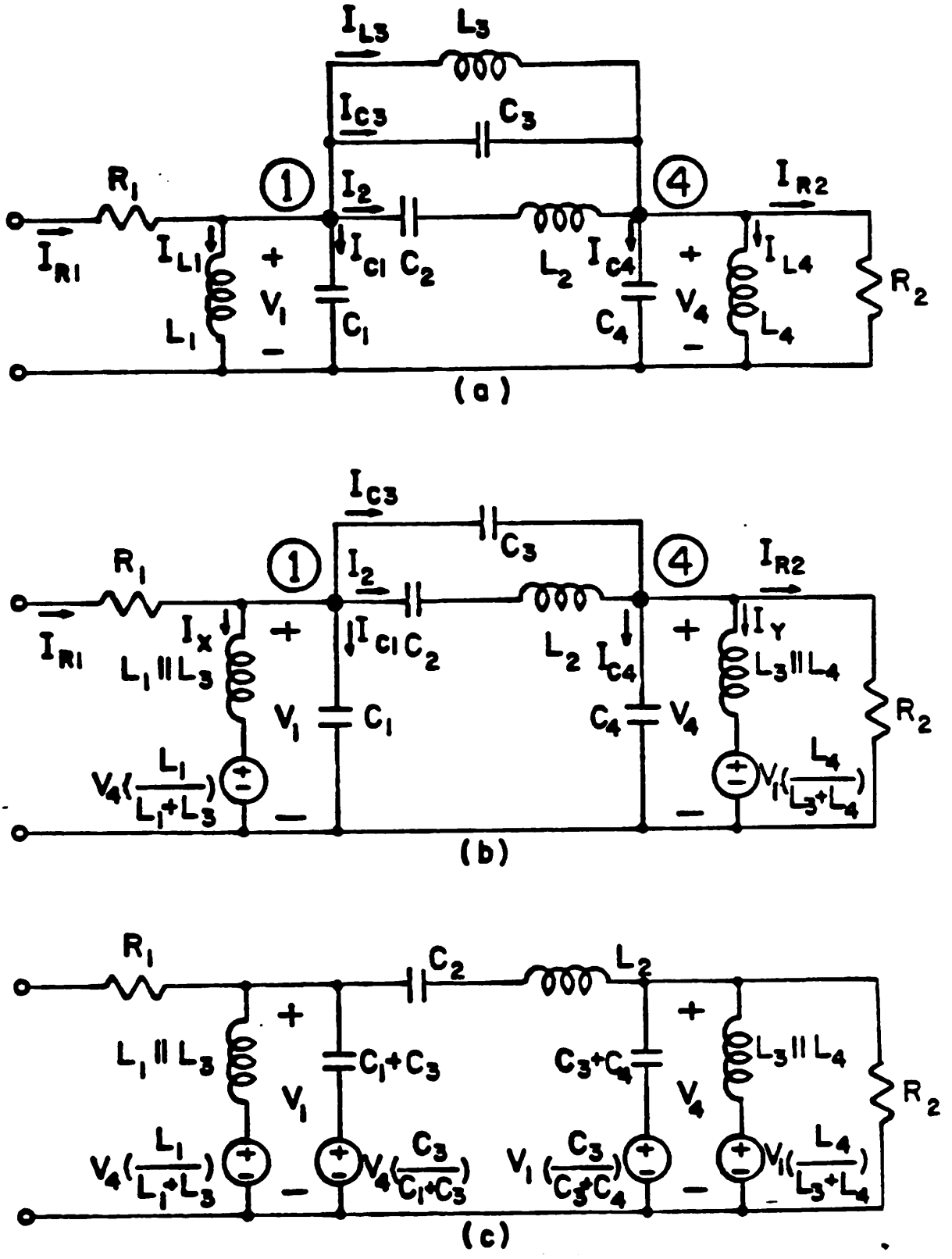


Fig. 2.24 (a) Doubly-terminated sixth order elliptic bandpass LC ladder.
(b) Inductor L_3 represented by voltage-controlled-voltage-source.
(c) Final Thevenin equivalent circuit of Fig. 2.24a.

or

$$V_1 = (I_{R1} - I_x - I_2)\frac{1}{s(C_1+C_3)} + V_4\frac{C_3}{C_1 + C_3}. \tag{2.37}$$

Similarly for node 4

$$V_4 = (I_2 - I_y - I_{R2})\frac{1}{s(C_3+C_4)} + V_1\frac{C_3}{C_3 + C_4}. \tag{2.38}$$

Equations (2.37) and (2.38) results in the final circuit shown in Fig. 2.24c. The shunt inductor L_3 and capacitor C_3 are now replaced by voltage-controlled-voltage-sources. The corresponding switched-capacitor circuit for Fig. 2.24c requires only 6 op. amps. to implement, since the voltage-controlled-voltage-sources can be easily realized by feed-forward paths.

2.5 Sensitivity in High Q Bandpass Ladder Filters

As mentioned earlier, most applications in high frequency communication systems require narrow band filters (Q ~ 40), with a rather tight tolerance in the center frequency accuracy. Compared with voice-band codec filters, these narrow band filters have pole locations much closer to the $j\omega$ axis. Typical codec filters are 5-pole, 4-zero lowpass filters, with largest pole Q approximately equal to 3 (Fig. 2.25a). Whereas a sixth order elliptic bandpass filter with a Q of 40 has individual pole Q's on the order of 100 (Fig. 2.25b). This means that small variations in the pole locations can cause significant variations in the passband response, and may give rise to a large passband ripple or instability in the worst case. Also, since the filters have narrow bandwidths, a small shift in the center frequency can move the passband outside the frequencies of interest, hence the filter design must be able to realize a stable center frequency.

To understand more about the sensitivity problem in a switched-capacitor bandpass ladder, let us analyze one such structure in more detail. For simplicity, we consider a fourth order all-pole bandpass filter. We start with a lowpass prototype filter shown in Fig. 2.26a. This lowpass filter is normalized to have a cutoff frequency at 1 radian per second. After lowpass to bandpass transformation, we have the doubly-terminated LC network shown in Fig. 2.26b. Here,

$$L_1 = \frac{L_{LP}}{B} \tag{2.39a}$$

$$C_1 = \frac{B}{\omega_0^2 L_{LP}} \tag{2.39b}$$

$$L_2 = \frac{B}{\omega_0^2 C_{LP}} \tag{2.39c}$$

$$C_2 = \frac{C_{LP}}{B} \tag{2.39d}$$

where ω_0 is the filter center frequency in radians per second, B is the filter bandwidth, and C_{LP} and L_{LP} are the capacitance and inductance values of the

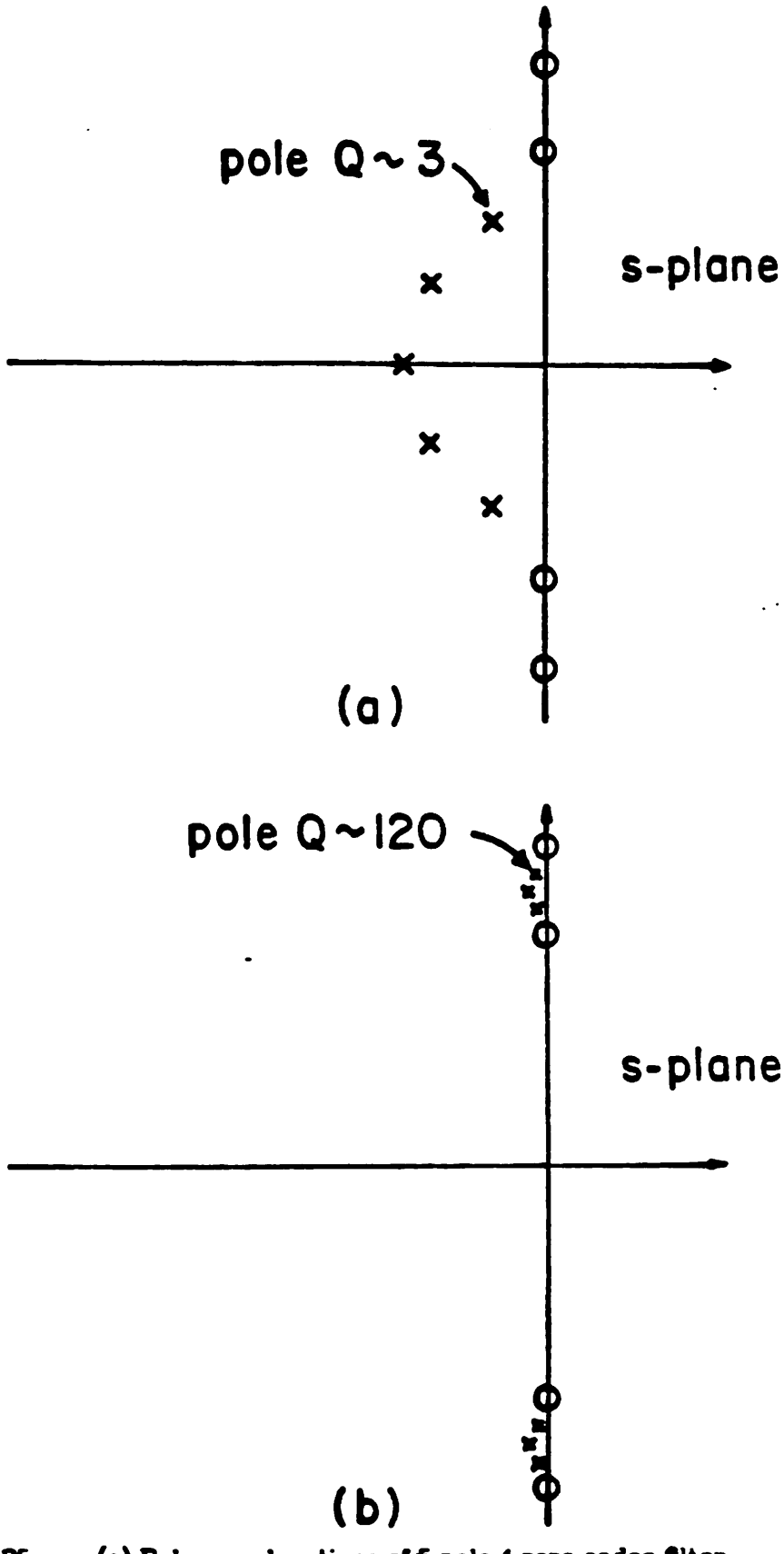


Fig. 2.25 (a) Pole-zero locations of 5 pole 4 zero codec filter.
(b) Pole-zero locations of 6th order elliptic bandpass filter.

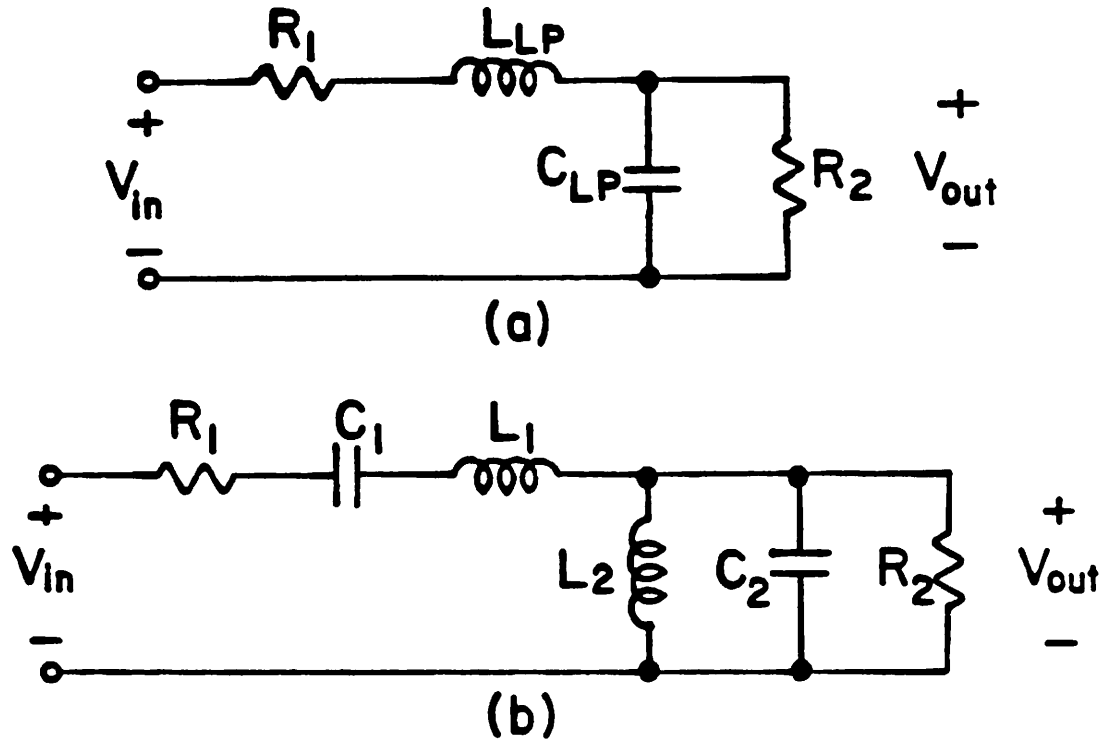


Fig. 2.26 (a) Prototype low pass filter.
(b) Fourth order LC bandpass ladder.

prototype lowpass filter.

The signal flow graph which represents the $i-v$ characteristics of this LC network is shown in Fig. 2.27, where

$$\tau_1 = R_n C_1 \tag{2.40a}$$

$$\tau_2 = \frac{L_1}{R_n} \tag{2.40b}$$

$$\tau_3 = R_n C_2 \tag{2.40c}$$

$$\tau_4 = \frac{L_2}{R_n} \tag{2.40d}$$

$$k_1 = \frac{R_1}{R_n} \tag{2.40e}$$

$$k_2 = \frac{R_n}{R_2} \tag{2.40f}$$

$$a = 1 \tag{2.40g}$$

$$b = 1 \tag{2.40h}$$

with τ_i being the integrating time constants and R_n, the normalizing resistance. A close examination of this signal flow graph shows that it is composed of two resonators coupled together by feedforward paths. This is representative of most bandpass leapfrog realizations, and it will be shown later that elliptic bandpass ladders will not be much different. Each resonator in the graph consists of two integrators in a loop, and resonates at a frequency equal to the center frequency of the filter. This can be seen from (2.39) and (2.40) where

$$\frac{1}{\tau_1 \tau_2} = \frac{1}{L_1 C_1} = \omega_o^2 \tag{2.41a}$$

$$\frac{1}{\tau_3 \tau_4} = \frac{1}{L_2 C_2} = \omega_o^2 . \tag{2.41b}$$

Hence, the filter center frequency is only a function of the resonator capacitor ratios and is independent of the other coupling capacitor ratios.

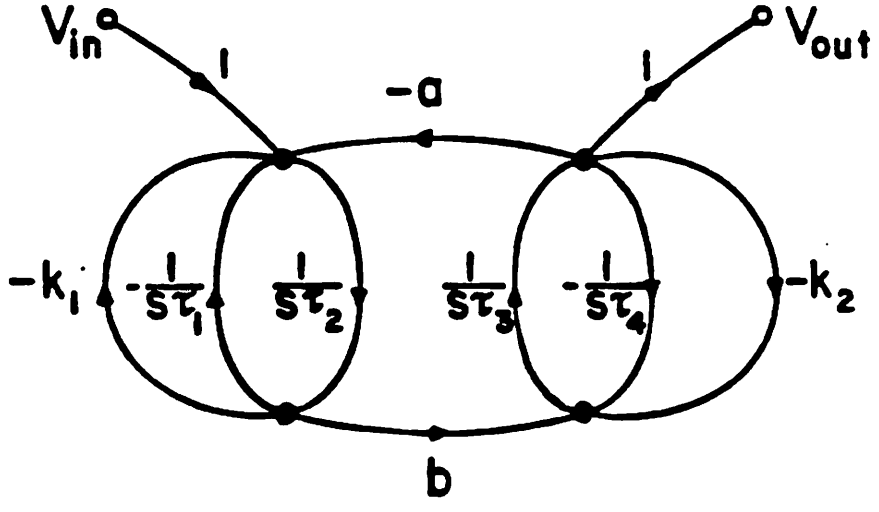


Fig. 2.27 Signal flow graph of fourth order LC bandpass ladder.

To analyze the sensitivity properties of this network, we calculate the incremental sensitivities of the frequency response with respect to the integrating time constants and the coupling ratios. By Mason's theorem, the filter transfer function is given by

$$T(s) = \frac{V_{out}(s)}{V_{in}(s)}$$

$$= \left(\frac{b}{s^2\tau_2\tau_3}\right) \Big/ \left\{ 1 + \frac{1}{s}\left(\frac{k_1}{\tau_2} + \frac{k_2}{\tau_3}\right) + \frac{1}{s^2}\left(\frac{1}{\tau_1\tau_2} + \frac{1}{\tau_3\tau_4} + \frac{ab + k_1k_2}{\tau_2\tau_3}\right) \right.$$

$$\left. + \frac{1}{s^3}\left(\frac{k_1}{\tau_2\tau_3\tau_4} + \frac{k_2}{\tau_1\tau_2\tau_3}\right) + \frac{1}{s^4\tau_1\tau_2\tau_3\tau_4} \right\}. \qquad (2.42)$$

From (2.41a,b), this can be rewritten as

$$T(s) = \frac{\dfrac{b}{s^2\tau_2\tau_3}}{1 + \dfrac{1}{s}\left(\dfrac{k_1}{\tau_2} + \dfrac{k_2}{\tau_3}\right) + \dfrac{1}{s^2}\left(2\omega_o^2 + \dfrac{ab + k_1k_2}{\tau_2\tau_3}\right) + \dfrac{\omega_o^2}{s^3}\left(\dfrac{k_1}{\tau_2} + \dfrac{k_2}{\tau_3}\right) + \dfrac{\omega_o^4}{s^4}}. \qquad (2.43)$$

The incremental sensitivity of $T(s)$ with respect to τ_1 is given by

$$S_{\tau_1}^{T} = \frac{\tau_1}{T(s)} \frac{\partial T(s)}{\partial \tau_1}$$

$$= \frac{\dfrac{\omega_o^2}{s^2} + \dfrac{\omega_o^2}{s^3}\left(\dfrac{k_2}{\tau_3}\right) + \dfrac{\omega_o^4}{s^4}}{1 + \dfrac{1}{s}\left(\dfrac{k_1}{\tau_2} + \dfrac{k_2}{\tau_3}\right) + \dfrac{1}{s^2}\left(2\omega_o^2 + \dfrac{ab + k_1k_2}{\tau_2\tau_3}\right) + \dfrac{\omega_o^2}{s^3}\left(\dfrac{k_1}{\tau_2} + \dfrac{k_2}{\tau_3}\right) + \dfrac{\omega_o^4}{s^4}}. \qquad (2.44)$$

At the center frequency where $s = j\omega_o$,

$$S_{\tau_1}^{T}\Big|_{\omega_o} = -j\frac{k_2\omega_o\tau_2}{ab + k_1k_2}. \qquad (2.45)$$

From (2.39) and (2.40), this can be rewritten as

$$S_{\tau_1}^{T}\Big|_{\omega_o} = -j\left(\frac{\dfrac{\omega_o}{B}L_{LP}}{R_1 + R_2}\right)$$

$$= -j\frac{Q\,L_{LP}}{R_1 + R_2}. \qquad (2.46)$$

R_1, R_2 and L_{LP} are the component values of the normalized lowpass filter of Fig. 2.26a. For $R_1=R_2=1$, values of L_{LP} is on the order of 1, so that

$$S_{\tau_1}^{T}\Big|_{\omega_0} \sim -jQ\,. \tag{2.47}$$

In most cases, we are more interested in the magnitude response of the transfer function rather than the phase response, and the incremental sensitivity of the magnitude response is given by

$$S_{\tau_1}^{|T|} = \mathrm{Re}\{S_{\tau_1}^{T}\} \tag{2.48}$$

so that

$$S_{\tau_1}^{|T|}\Big|_{\omega_0} = 0\,. \tag{2.49}$$

This is expected since the sensitivity of the magnitude of the frequency response with respect to the L,C components in a doubly-terminated LC network is zero at the center frequency [24], and the integrator time constants τ_i are the simulated L,C component values in a leapfrog structure. However, at the passband edges, the sensitivities are markedly different. If we let $s=j\omega_b$ in (2.44) where ω_b is the passband edge given by $\omega_b = \omega_0 - \frac{B}{2}$, then we have

$$S_{\tau_1}^{T}\Big|_{\omega_b} \approx \frac{\frac{B}{\omega_b} + j\frac{k_2}{\tau_3\omega_b}(1+\frac{B}{\omega_b})}{\frac{ab+k_1k_2}{\tau_2\tau_3\omega_b^2} - j(\frac{k_1}{\tau_2}+\frac{k_2}{\tau_3})\frac{B}{\omega_b^2}}\,. \tag{2.50}$$

Here, we have assume that the filter has a high Q so that

$$(1+\frac{B}{\omega_b})^n \approx 1 + n\frac{B}{\omega_b}\,. \tag{2.51}$$

As before, we calculate the incremental sensitivity of the magnitude response. From (2.50) and (2.48), and assuming $\frac{B}{\omega_b} \ll 1$, we have

$$S_{\tau_1}^{|T|}\Big|_{\omega_b} \approx \frac{B\omega_b\left(\frac{ab}{\tau_2\tau_3}-\frac{k_2}{\tau_3^2}\right)}{\left(\frac{ab+k_1k_2}{\tau_2\tau_3}\right)^2+\left(\frac{k_1}{\tau_2}+\frac{k_2}{\tau_3}\right)^2B^2}. \tag{2.52}$$

From (2.39) and (2.40), this can be rewritten as

$$S_{\tau_1}^{|T|}\Big|_{\omega_b} \approx \frac{\frac{\omega_b}{B}\left(\frac{1}{L_{LP}C_{LP}}-\frac{1}{R_2^2C_{LP}^2}\right)}{\left(\frac{1+R_1/R_2}{L_{LP}C_{LP}}\right)^2+\left(\frac{R_1}{L_{LP}}+\frac{1}{R_2C_{LP}}\right)^2}. \tag{2.53}$$

As mentioned earlier, R_1, R_2, L_{LP}, C_{LP} are lowpass prototype components with values on the order of 1, and since $\frac{\omega_b}{B}\approx Q$, we have

$$\left|S_{\tau_1}^{|T|}\right|_{\omega_b} \sim Q\,. \tag{2.54}$$

It can be shown that the same is true for the other τ_i's. Hence at the passband edge, the sensitivities of the magnitude response with respect to the integrator time constants are on the order of Q. On the other hand, if we calculate the same sensitivities with respect to the coupling ratios, we have

$$\begin{aligned} S_a^T &= \frac{a}{T(s)}\frac{\partial T(s)}{\partial a} \\ &= -\frac{\frac{ab}{s^2\tau_2\tau_3}}{1+\frac{1}{s}\left(\frac{k_1}{\tau_2}+\frac{k_2}{\tau_3}\right)+\frac{1}{s^2}\left(2\omega_o^2+\frac{ab+k_1k_2}{\tau_2\tau_3}\right)+\frac{\omega_o^2}{s^3}\left(\frac{k_1}{\tau_2}+\frac{k_2}{\tau_3}\right)+\frac{\omega_o^4}{s^4}} \\ &= -aT(s) \end{aligned} \tag{2.55}$$

or

$$S_a^{|T|} = -a\ \mathrm{Re}\{T(s)\}\,. \tag{2.56}$$

At the passband frequencies, $|T(s)|\approx\frac{1}{2}$, so that

$$\left|S_a^{|T|}\right| \sim \frac{1}{2}. \tag{2.57}$$

The same can be shown to be true for the other coupling ratios. In other words, at the passband edges, the magnitude of the frequency response is about Q times more sensitive to the integrator capacitor ratios than the coupling capacitor ratios.

The significance of this is that the integrator capacitor ratios should be as accurate as possible whereas we can tolerate larger errors in the coupling capacitors. Hence, if we use the component values from Fig. 2.27 without proper scaling, we have from (2.39) and (2.40)

$$\tau_1 = \frac{R_n B}{\omega_o^2 L_{LP}} \tag{2.58}$$

$$\tau_2 = \frac{L_{LP}}{B R_n} \tag{2.59}$$

so that τ_2 is on the order of Q^2 times τ_1. The same is true for other resonators. This would result in some very large capacitor ratios and some small ones, so that in terms of percentage, there will be larger errors involved with the large ratios, resulting in greater distortion in the frequency response than if all the capacitor ratios are on the same order. Hence, the optimal solution to this problem is make all the individual integrators identical. In other words, all the integrators will have identical time constants equal to the reciprocal of the center frequency (in radians per second). In a switched-capacitor integrator shown in Fig. 2.5 or 2.9, the integrating time constant τ is given by

$$\tau = \frac{C_2}{f_s C_1} \tag{2.60}$$

where C_1 is the sampling capacitor, C_2 is the integrating capacitor, and f_s is the sample rate. If all the integrators are made identical with a time constant given by the center frequency f_o, then

$$\frac{1}{2\pi f_o} = \frac{C_2}{f_s C_1} \tag{2.61}$$

or

$$\frac{C_2}{C_1} = \frac{f_s}{2\pi f_o} \tag{2.62}$$

Even if the clock rate to center frequency ratio is 20,

$$\frac{C_2}{C_1} \approx 3 . \tag{2.63}$$

From (2.62), f_o is now only a function of one particular capacitor ratio. This ratio is small (2.63), hence the minimum size capacitor can be on the order of 1pF without jeopardizing the op. amp. settling time. With this kind of capacitor size, accuracies of better than half a percent can be easily achieved, and this is more than enough for most center frequency accuracy requirements. On the other hand, when the individual integrator time constants are scaled to the reciprocal of ω_o by (2.62), the coupling ratios will be on the order of Q, meaning that large capacitor ratios are required for the coupling paths. However, since the sensitivities of the transfer function with respect to these ratios are small (2.57), we can allow a larger error in these capacitors. Typically, a 5% error in these coupling capacitor ratios are tolerable, so that capacitors smaller than 1pF can be used to ease the amplifier settling requirements.

Notice that with this scheme of identical resonators, all the integrators have the same capacitor ratios and hence all the integrator time constants track each other over all conditions of process and temperature variations. As mentioned earlier, the filter center frequency is only a function of the resonator capacitor ratios and is independent of the coupling capacitor ratios. If all the integrator capacitor ratios match each other, then a 0.5% variation of all of the integrator time constants in the same direction will simply shift the overall response by 0.5% without altering its shape. However, a change of one integrator time constant while the others remain the same would grossly distort the passband shape. This is another advantage with this scheme of identical

integrators, for if we have some large integrator capacitor ratios and some small ones, the larger ratios will have large errors whereas the smaller ones will remain sort of constant, resulting in a larger passband distortion.

Another obvious advantage of using this scheme is that it greatly simplifies the layout since all integrators are identical. Nevertheless, there is a disadvantage with this method. With identical resonators, the op. amps. are not properly scaled to have equal maxima, resulting in a reduction of dynamic range. In a narrow band filter with Q on the order of 40, there can be a loss of about 6 dB in dynamic range.

2.6 Implementation of High Q Filters Using Frequency Translation (Decimation Scheme)

As mentioned earlier, high Q filters require large capacitor ratios in the coupling paths between the resonators. Further, increasing the Q of the filter requires the operational amplifier to settle to higher degrees of accuracy within the clock half period to avoid distortion of the response. Hence, the combination of very high clock frequencies and very high Q's presents particularly difficult implementation problems if done directly. Typically, narrow band filters with Q's on the order of 40 require coupling capacitor ratios on the order of 100. In addition, amplifier gains need to be greater than 500 for reasonable accuracies. If operation in the 10 MHz range is required, even with a clock rate of only 5 times the center frequency (say ~ 50 MHz), it is extremely difficult to design an amplifier with such a high gain while at the same time be able to settle within the clock half period with capacitor sizes needed to obtain reasonable accuracies. Unfortunately, filters in the 10 MHz range that require Q's on the order of 40 are not uncommon. One such example is the 10.7 MHz FM intermediate frequency (IF) filter for channel separation in the IF strip of the FM receiver. Hence, we must find ways to relax the filter requirements for this kind of high frequency, high Q filter applications.

In most practical situations, it is possible to reduce the filter requirement by performing the bandpass filtering in two or more stages, each at successively lower frequency. While the overall center frequency to bandwidth ratio can be high, the Q of the individual filtering stages can be much lower. This approach is similar to the multiple IF filter approach used in some communication-receivers. In this case, the idea is to mix (or decimate) the signal in the 10 MHz range to a lower frequency where the critical filtering is performed. This is particularly convenient in the IF application because the signal has to be further demodulated down to the base band in a later stage of the receiver anyway. Hence, it is

not very critical at which intermediate frequency the channel separation is done. In a conventional continuous-time system, modulation of signals is complicated by the need for non-linear circuits such as multipliers. However, in a sampled-data system, such an operation can be realized simply by undersampling the signal, which will then alias it to a lower frequency.

Such a scheme is realized by the block diagram shown in Fig. 2.28. First, the signal from the RF stage (in the 10 MHz range) is sampled at a high clock rate (~50 MHz) where a rough filtering is done. Next, the signal is down sampled at a lower clock rate, say 4 times lower. In so doing, the signal in the 10 MHz range is aliased down to a lower frequency (~2.5 MHz) where the actual filtering for channel separation is performed. This signal will then be further demodulated in a later stage. The purpose of the first filter is to provide antialias for the second filter. This relaxes the requirements for the filtering at the highest sample rates as shown in Fig. 2.29. The only requirement of the first filter is to reject image frequencies while passing the original signal. The second filter which performs the channel separation function has to be precise, but it is at a lower frequency. Since it has the same bandwidth as a direct realization at a higher frequency, the resulting Q is reduced. As an example, the relaxation in filter requirements for a 10.7 MHz FM IF filter is summarized in Table 2.1. Here, the original filter sampled at 54 MHz requires a sixth order Chebyshev filter with a Q of 40 and coupling capacitor ratio on the order of 120. The amplifier gain needed for this application has to be larger than 500. With this decimation scheme, the antialias filter has a Q of only 7 and capacitor ratios around 10. This filter would only require an amplifier gain of about 100 so that it is realizable up to frequencies in the 10 MHz range. In addition, the accuracy requirements of this filter is not critical since its purpose is to reject the image signals only. The second filter has a Q of 10 and capacitor ratios around 20. This filter has to be accurate and would require amplifier gains greater than 200. However,

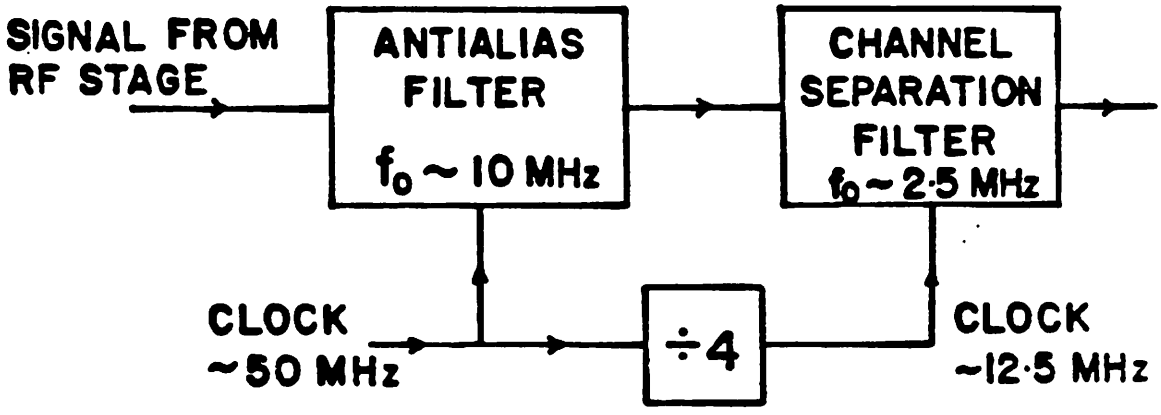


Fig. 2.28 Block diagram of the decimation scheme.

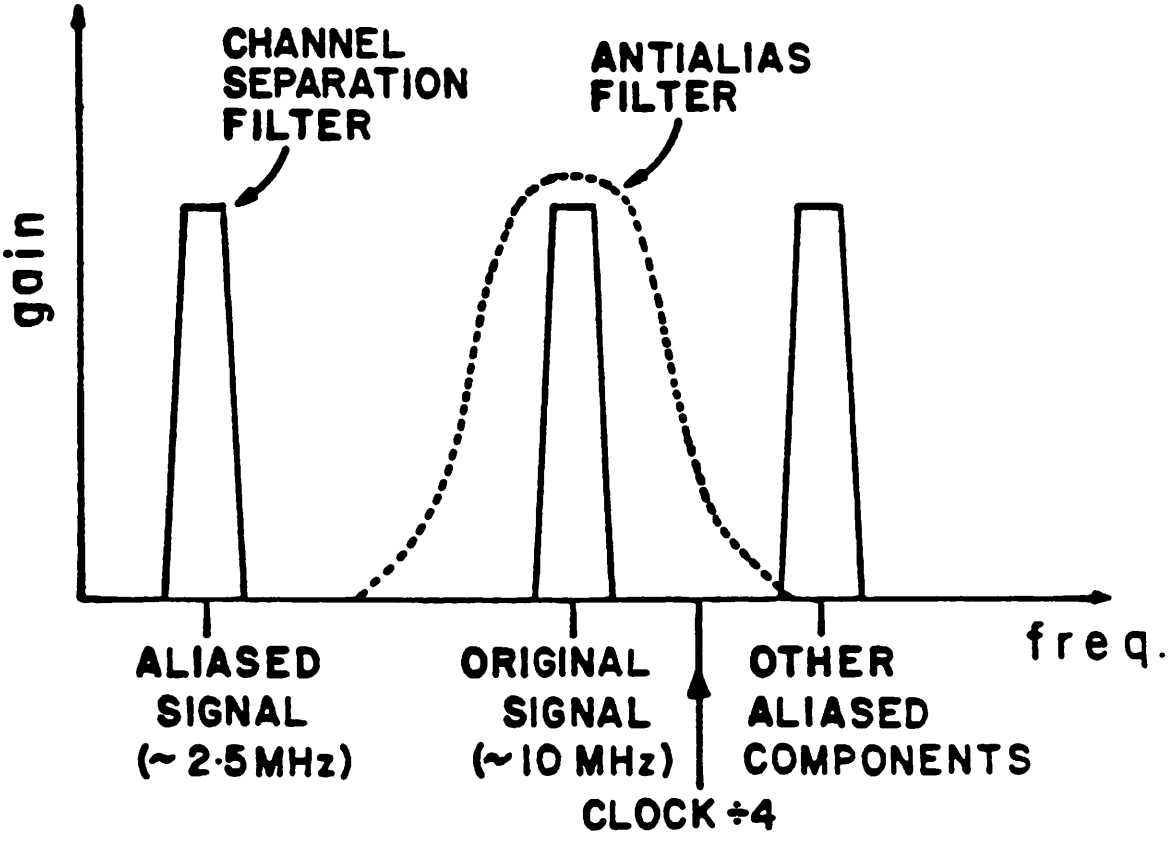


Fig. 2.29 Frequency response of the anti-alias filter and the channel separation filter in the decimation scheme.

	DIRECT REALIZATION	DECIMATION SCHEME	
		ANTIALIAS	CHANNEL SEPARATION
Filter Type	6th order Chebyshev	6th order butterworth	6th order Chebyshev
center frequency	10.7 MHz	10.7 MHz	2.8 MHz
clock rate	54 MHz	54 MHz	13.5 MHz
Q	40	7	10
largest capacitor ratio	120	10	20
amplifier gain required	>500	>100	>200

Table 2.1 Filter requirements for a 10.7 MHz FM IF filter using the direct method and the decimation scheme

this is at 2.8 MHz sampled at 13.5 MHz. Therefore there is a window of 70 ns for the amplifier to settle, which is adequate for the gain requirements.

Chapter 3
Practical Implementation of High Frequency Switched-Capacitor Filters

3.1 Introduction

The most crucial problem in the implementation of high frequency switched-capacitor filter is the settling time of the operational amplifier. Since this places an upper limit on the allowable clock rate, the operating frequency of a switched-capacitor filter is primarily determined by the amplifier transient response. Another problem of practical concern is the component ratios spread in these high Q bandpass filters. As mentioned earlier, filters with Q's on the order of 40 requires coupling capacitor ratios on the order of 100. We must be able to realize these ratios with reasonable accuracies without jeopardizing the amplifier settling time. In addition to these two problems, there are the common issues that plague every switched-capacitor filter design, namely the problems of clock noise and PSRR, and they tend to get worse with higher frequencies.

This chapter is concerned with these practical implementation issues. In Section 3.2, we discuss the use of fully differential operational amplifiers in high frequency switched-capacitor filters. Fully differential op. amps. have been employed in voice band filters [10]. Here, we will concentrate on their advantages in high frequency designs. Section 3.3 deals with the problem of large component ratios spread, and a T-network scheme is used to reduce the capacitor ratios in feedforward coupling paths. A double-sampling LDI integrator is introduced in Section 3.4 which will effectively double the sample rates of conventional LDI integrators. Lastly, in Sections 3.5 and 3.6, detail design examples for an AM IF filter and a FM IF filter are presented.

3.2 Fully Differential Filter Implementation

As mentioned earlier, the selection of an operational amplifier implementation which realizes the required settling time while still achieving sufficient D.C. gain is perhaps the most important consideration in the implementation of high frequency filters. In conventional single-ended NMOS amplifiers, the required gain is typically realized in two gain stages, so that there is a need of a large compensation capacitor to ensure proper operation in feedback configurations. This compensation capacitor tends to limit the bandwidth of the amplifier, which in turn determines the amplifier settling time. In addition, conversion from differential input to single-ended output requires the common mode transients of the differential input stage to settle out before the final output is valid. In most designs, the common mode settling time tends to be longer than the differential mode settling time since the capacitances associated with the common mode nodes are usually larger. Hence, the common mode settling behavior dominates the transient response of the operational amplifier.

To alleviate these speed limitation, fully differential CMOS op. amps. can be used. In a fully differential implementation, the common mode settling behavior no longer determines the transient response of the differential output, hence improving the amplifier bandwidth. Also, the use of CMOS technology can realize the required gain in typically one stage, thereby eliminating the need for a large compensation capacitor.

Another consideration affecting the choice of operational amplifier configuration is the fact that the effect of charge injection into the signal path from the switch transistors becomes much more important as the clock frequency is increased. This occurs because the switches themselves must be larger so as to minimize the channel resistance which characterizes charge redistribution during the clock intervals. As the switches get larger, the channel

charge coupled into the operational amplifier summing node and output node will increase. This larger clock feedthrough can degrade the effective settling time of the amplifier, degrade the power supply rejection if the clock voltage is dependent on a supply, and give rise to large D.C. offsets in the filter.

The power supply rejection ratio (PSRR) of the amplifier is particularly important in high-frequency, high-Q filters, since power supply variations can be coupled to some internal nodes of the op. amp. and will eventually appear at the output. Because high-Q filters typically take the form of an array of resonators which are weakly coupled to each other, parasitic coupling paths through the power supply can have a strong effect on the response even if they are small in magnitude.

These considerations point strongly to the use of fully differential signal paths and a differential output operational amplifier. Because it is fully differential, clock noise and power supply variations appear as common mode signals, and therefore will not affect filter response. The design of such a differential high speed CMOS amplifier for high frequency switched-capacitor filter applications is considered in detail in [26]. This design uses a high swing folded-cascode configuration shown in Fig. 3.1, with common mode feedback circuitry to stablize the output D.C. conditions. Typical characteristics for this kind of amplifier is shown in Table 3.1, which is sufficient for most high frequency applications.

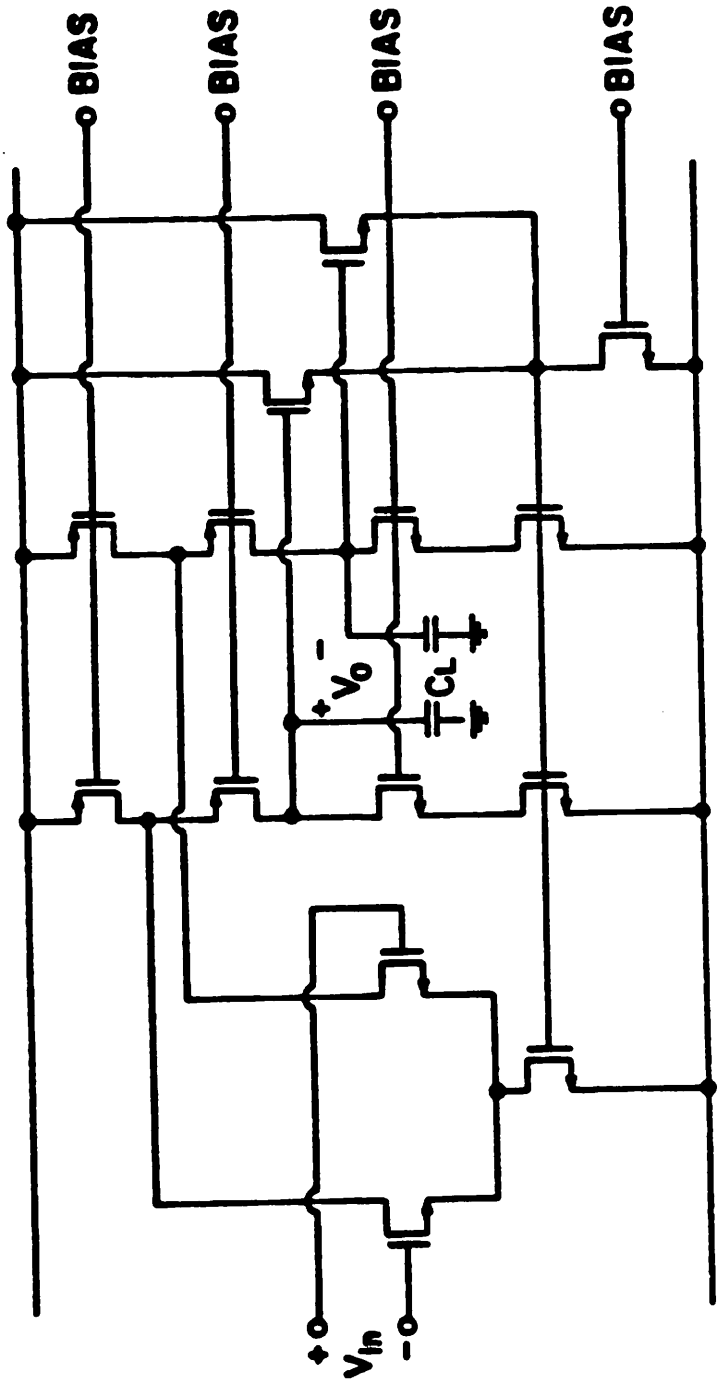


Fig. 3.1 Fully differential folded-cascode amplifier.

Unity gain frequency (2 pF load)	80 MHz
Setting time to 0.1% of final value (2 pF load 2.5v steps)	40 ns
Power dissipation	10 mW
Die area	200 mil^2
Open loop gain	1500

Table 3.1 Operational Amplifier Performance (± 5v supply)

3.3 Component Ratio Spread in High Q Bandpass Ladder Filters

In addition to the tight amplifier requirements, the fact that high frequency filters have high Q's also give rise to large spreads in the required capacitor ratios. As an example, let us consider the case of the sixth order elliptic bandpass filter shown in Fig. 2.24. The signal flow graph which represents the $i-v$ characteristics of Fig. 2.24c is shown in Fig. 3.2. Here, the solid lines in the graph constitutes the all-pole sixth order filter, whereas the dash lines are the feedforward paths which realize the voltage-controlled-voltage-sources. Like the one shown in Fig. 2.27, this signal flow graph is composed of resonators coupled together by feedforward paths, and the resonating frequency of each resonator is equal to the filter center frequency. If we choose to use the identical resonator scheme in Section 2.5 to improve the sensitivity performance of the filter, then the coupling paths between these resonators have capacitor ratios which are only determined by the Q of the filter as well as the nature of the band edges. Typically, a high Q or a sharp roll-off would mean a large ratio, and a Q of 40 in this sixth order elliptic bandpass filter would require capacitor ratios on the order of 100. In order to maintain adequate ratio accuracy, the minimum size capacitor cannot be too small ($\geq$ 0.1 pF), and thus the large ratios would require large capacitors which in turn slow down the amplifier response.

There are two kinds of coupling paths between the resonators. A realization of the signal flow graph in Fig. 3.2 using switched capacitor integrators is shown in Fig. 3.3. Here single-ended op. amps. are used for illustration, the same is true for the fully differential case. It is seen that the coupling paths which feed into the outputs of integrators can be realized by feed-forward capacitors, whereas those which feed into the inputs of integrators has to be realized via sampling capacitors. For the feed-forward capacitors, a T-network scheme can

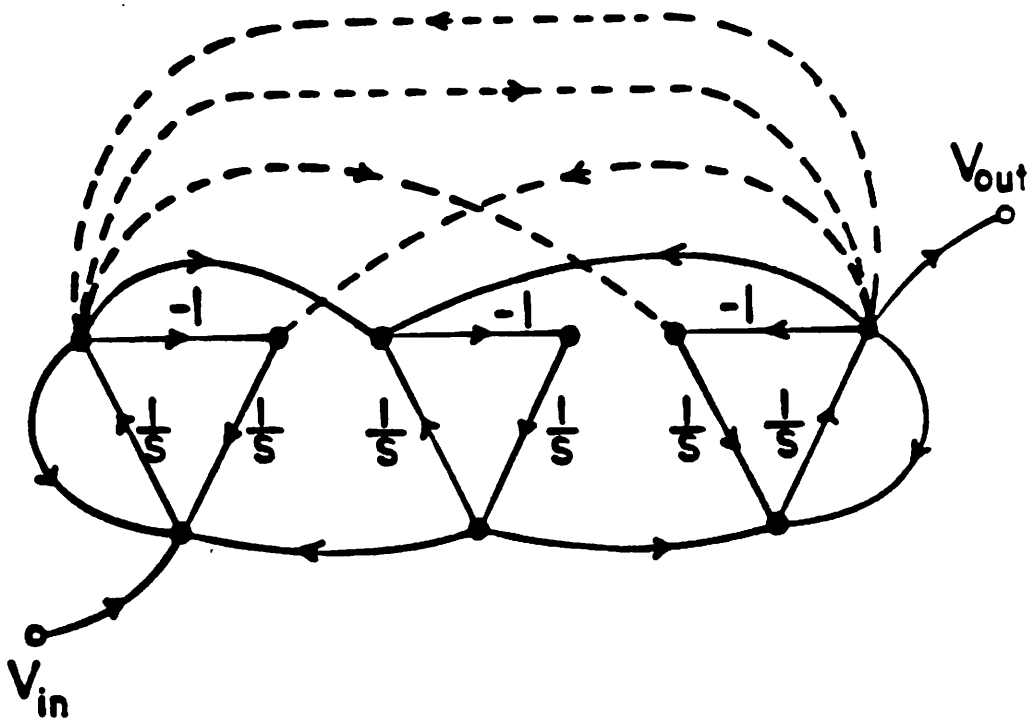


Fig. 3.2 Signal flow graph of sixth order elliptic bandpass filter shown in Fig. 2.24c.

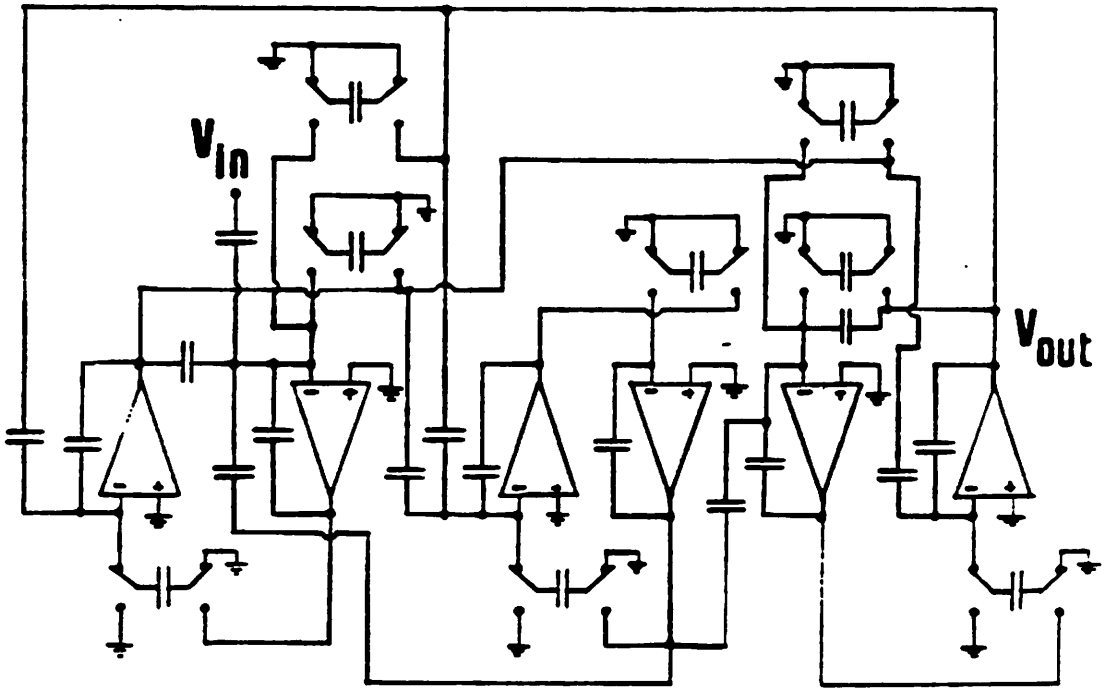


Fig. 3.3 Six op. amp. realization of sixth order elliptic bandpass filter.

be used to reduce the large coupling ratios required. A straight forward realization of a gain of 0.01 is shown in Fig. 3.4a. Here, the gain of the circuit is given by

$$\frac{V_2(s)}{V_1(s)} = -\frac{C_1}{C_2} \tag{3.1}$$

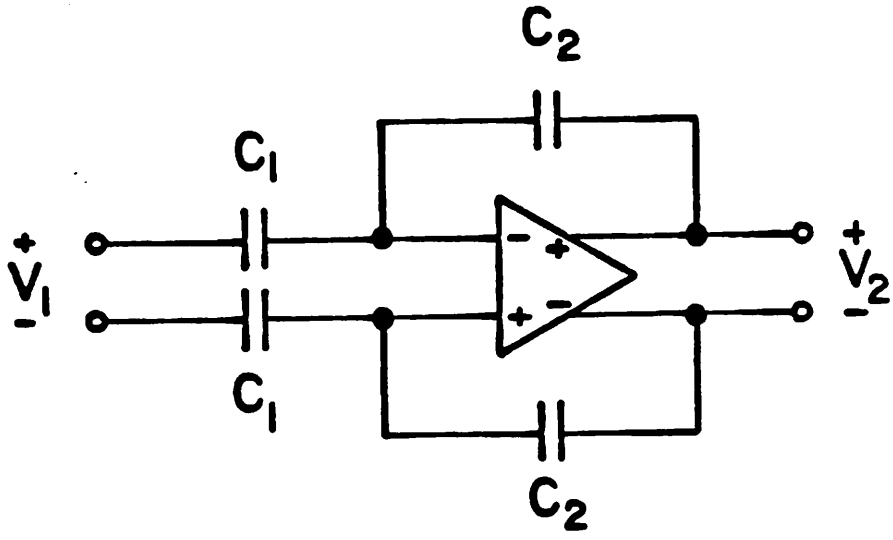
and the ratio C_2:C_1 is equal to 100:1. A T-network scheme used to realize a similar gain of 0.01 is shown in Fig. 3.4b. Here, the transfer function of the circuit is given by

$$\frac{V_2(s)}{V_1(s)} = -\frac{C_1}{C_1 + C_3 + C_4} \cdot \frac{C_4}{C_2}. \tag{3.2}$$

The ratio C_2:C_3:C_1:C_4 is equal to 10:8:1:1 for the same gain of 0.01. Thus, a maximum capacitor ratio of 100 is reduced to 10 by this method.

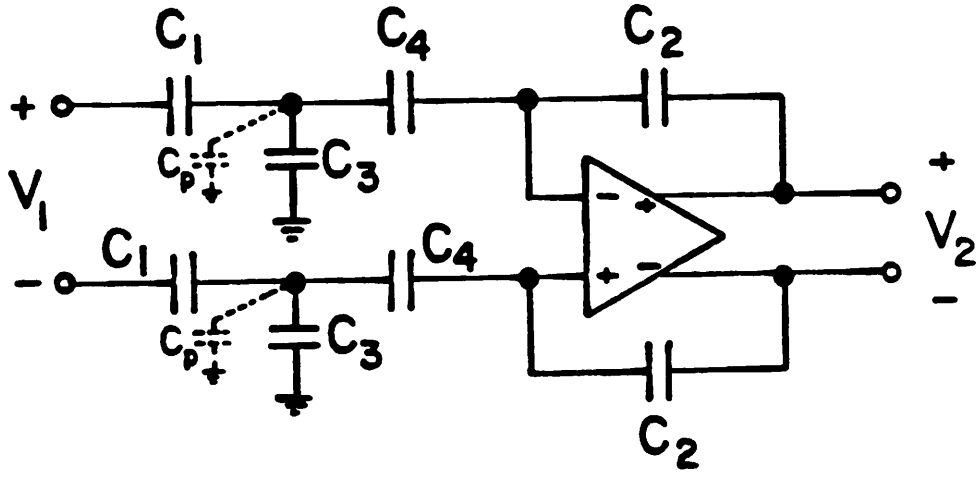
Nevertheless, the circuit in Fig. 3.4b is sensitive to parasitic capacitances C_p at the center node of the T-network. However, this capacitance can be reduced to a negligible amount by means of proper layout. Fig. 3.5 shows such a layout of the T-network. Here, the solid lines are the thin oxide regions, and the dash lines are the polysilicon of the capacitors (polysilicon to substrate capacitors are assumed). The polysilicon forms the center node of the T-network. Thus, parasitic capacitance C_p consists of only 2 squares of minimum feature size over the field oxide. For the case of Fig. 3.4b, if the minimum size capacitors (C_1 and C_4) are equal to to 0.1 pF, then C_3 is equal to 0.8 pF. Assuming a 4 X 4 μm minimum feature over a 0.8 μm field oxide, C_p is approximately equal to 1.4 fF. This C_p adds to C_3, the biggest capacitor of the T-network, and 1.4 fF amounts to less than 0.2% of 0.8 pF. Hence, the parasitic capacitances are entirely negligible.

This T-network scheme is not applicable towards the coupling paths realized via sampling capacitors, because of the additional switching involved and the



$C_2 : C_1 = 100 : 1$

(a)



$C_4 : C_3 : C_2 : C_1 = 1 : 8 : 10 : 1$

(b)

Fig. 3.4 (a) Direct realization of a gain of 0.01.
(b) T-network scheme to realize a gain of 0.01.

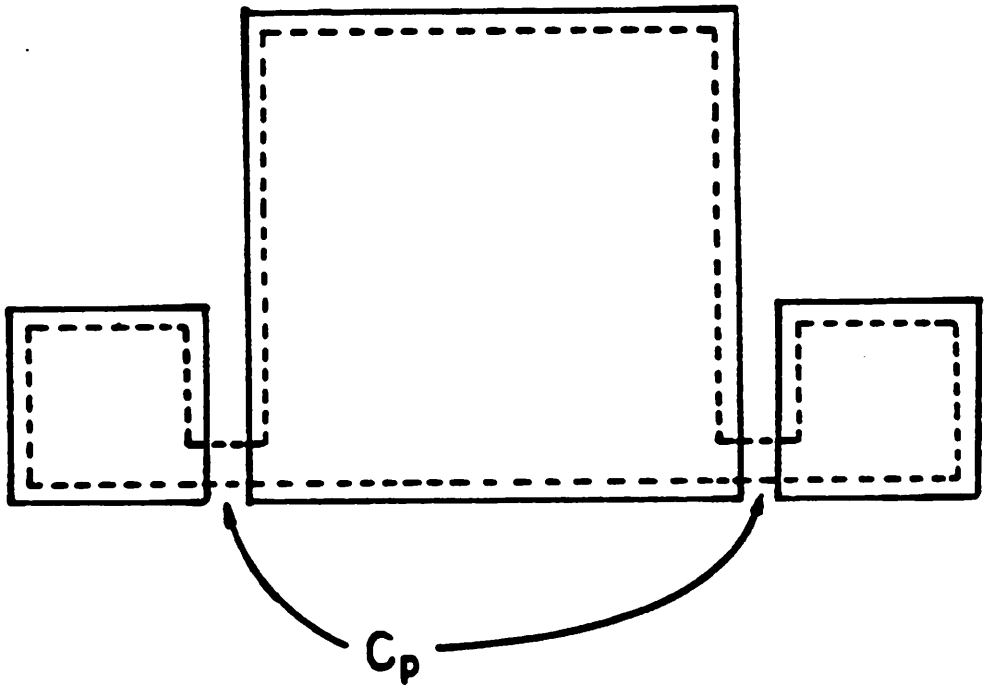


Fig. 3.5 T-network layout strategy to minimize capacitances.

added clock noise. Thus, it is desirable to convert this type of coupling to feed-forward capacitors. This is done by redirecting the signal paths which go into an integrator input to the output of the other integrator of the same resonator. This is shown in Figs. 3.6a and b. Here, the signal path α which feed into the input of an integrator in Fig. 3.6a is redirected as shown in Fig. 3.6b. However, this integrator output may feed other parts of the circuit as well. Hence, an equal amount of the signal has to be subtracted from those nodes fed by this output. In Fig. 3.7b, this is shown by the signal path $\alpha\beta$. Using this method, the signal flow graph in Fig. 3.2 can be redrawn in Fig. 3.7. It should be noted that Fig. 3.7 has been simplified by ignoring signal paths which have a gain factor of 10^{-4} or less. The corresponding switched-capacitor circuit for this flow graph is shown in Fig. 3.8. This circuit has 7 op. amps. instead of 6 because the filter output node has been modified and a signal has to be subtracted from it to regain the proper output. An extra op. amp. is therefore needed to do the subtraction. In this circuit, all the coupling paths are now realized by feed-forward capacitors, and thus T-networks are applicable for reducing the capacitor ratios.

One drawback of the use of continuous coupling paths such as those represented by T-networks is the fact that the settling behavior of the operational amplifiers within the filter is more complex than before, since continuous-time paths exist which link the output voltage of one amplifier to the summing node of another during the settling interval. Fortunately, in the case of high Q filters, the coupling paths between operational amplifiers is sufficiently weak that the loop gain around any continuous-time path loops is small compared to unity. Under these conditions, the settling time to a given accuracy is lengthened by a factor on the order of 2 for operational amplifiers whose inputs and outputs are interconnected by continuous-time paths.

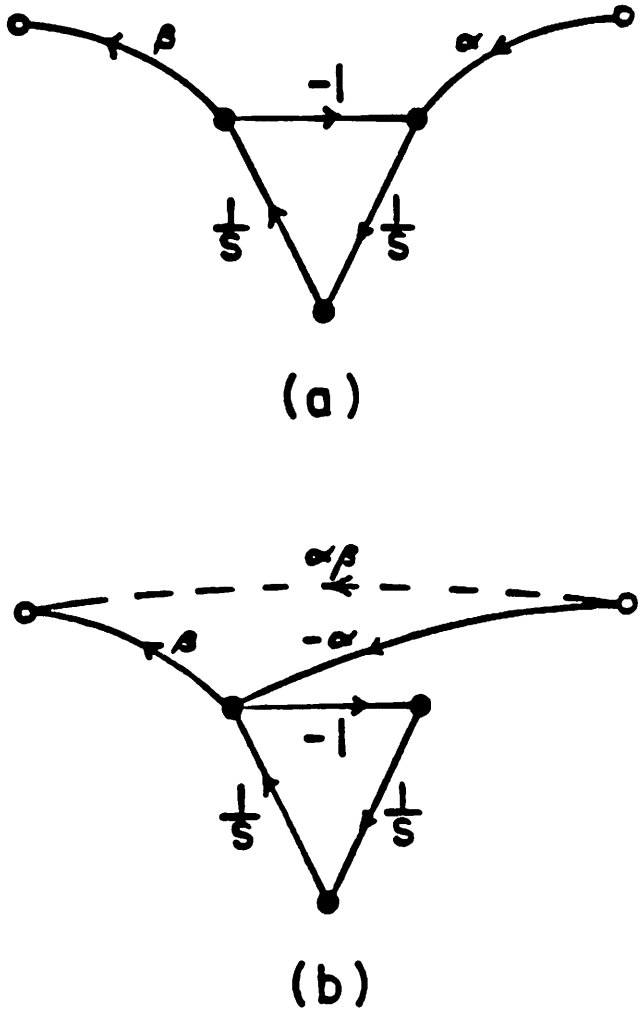


Fig. 3.6 Conversion of coupling paths via sampling capacitors to coupling paths via feed-forward capacitors.
(a) Original signal flow diagram.
(b) Signal flow diagram after conversion.

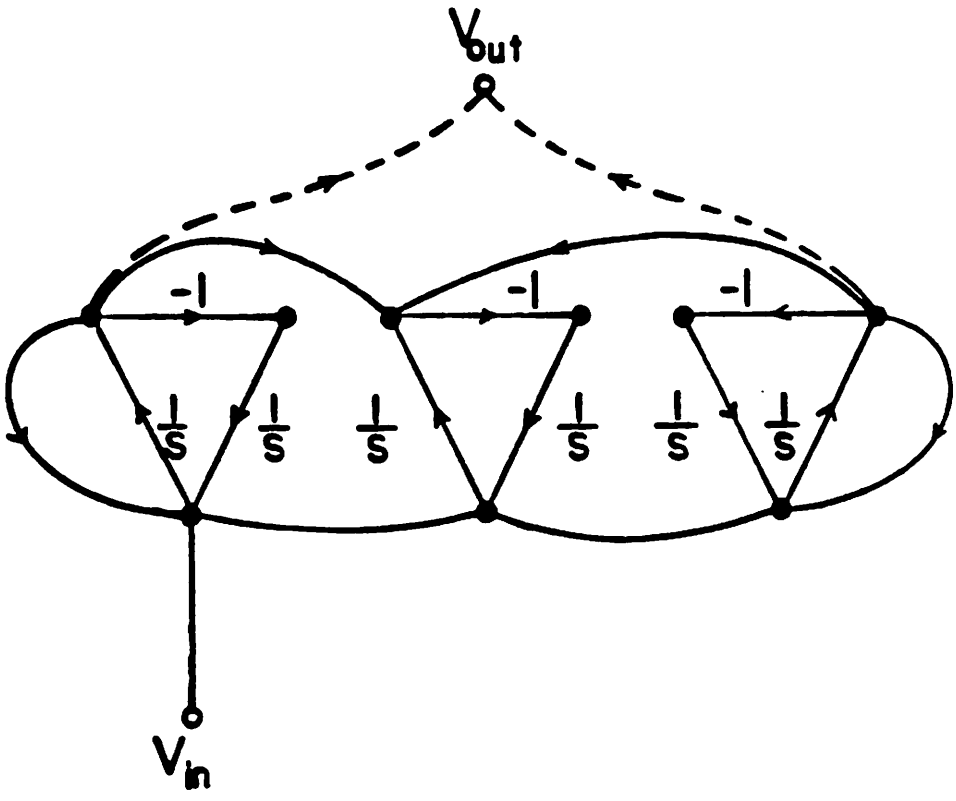


Fig. 3.7 Signal flow graph for sixth order elliptic bandpass filter after converting coupling paths via sampling capacitors to coupling paths via feed-forward capacitors.

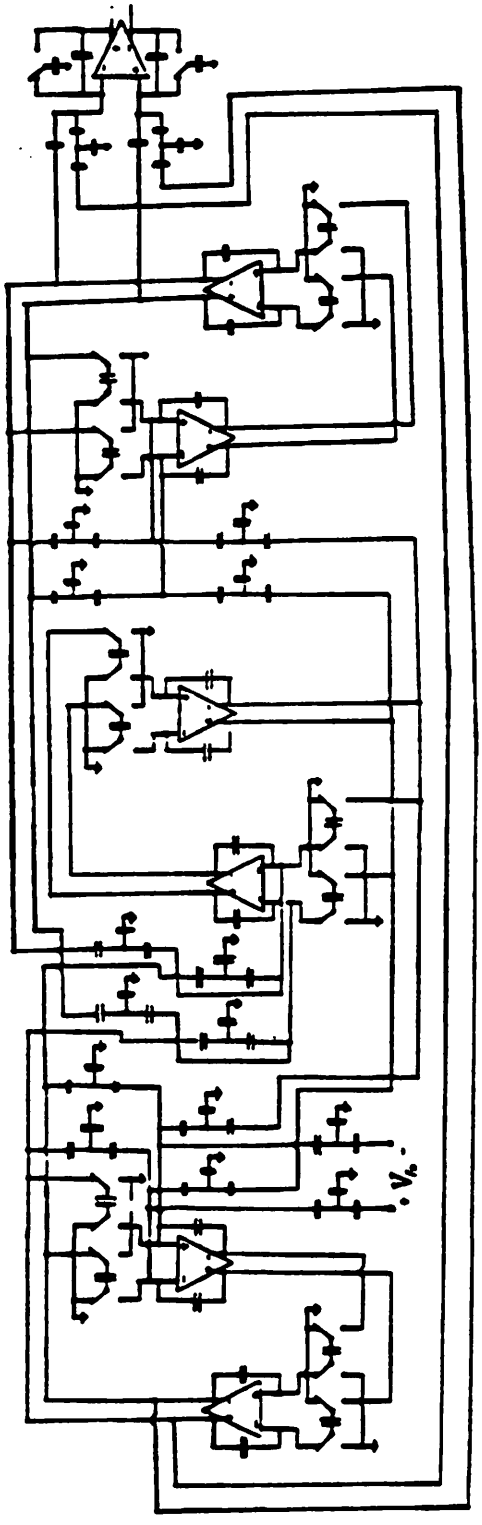


Fig. 3.8 Seven op. amp. realization of sixth order elliptic bandpass filter.

3.4 Double-Sampling Integrators and Their Implementation in LDI Ladders

To obtain the highest possible frequency operation, it is desirable to have the op. amps. being used at all times. The LDI integrators of Figs. 2.6 and 2.10, however, have the property that the outputs of the op. amps. are only sampled during one phase and are not sampled during the other one. That is, they are only sampled once every clock cycle, T_c, and the effective sample rate, f_s, is therefore, equal to the clock frequency $f_c = \frac{1}{T_c}$. If these outputs were sampled on both clock phases, then the effective sample rate could be twice the clock frequency.

A single-ended integrator to implement the LDI transformation which has a sample rate equal to twice the clock rate, $f_s = 2f_c$, is shown in Fig. 3.9a, where V_1 is the non-inverting integrator input, and V_2, the inverting input. Doubling the effective sample rate is accomplished by putting two sampling capacitors with opposite clock phases in parallel. Thus, charge is fed into the feedback capacitor on both clock phases. In so doing, each clock phase is considered as one full delay period. This circuit is describe by the following difference equation:

$$C_2 V_3[(n+1)T_s] = C_2 V_3(nT_s) + C_1 V_1(nT_s) - C_1 V_2[(n+1)T_s] \tag{3.3}$$

in which

$$T_s = \frac{T_c}{2}.$$

In terms of z-transform, we can write

$$V_3 = \frac{C_1}{C_2} \frac{1}{1-z^{-1}}(z^{-1}V_1 - V_2). \tag{3.4}$$

This can be put into block diagram form as shown in Fig. 3.9b. From this, we see

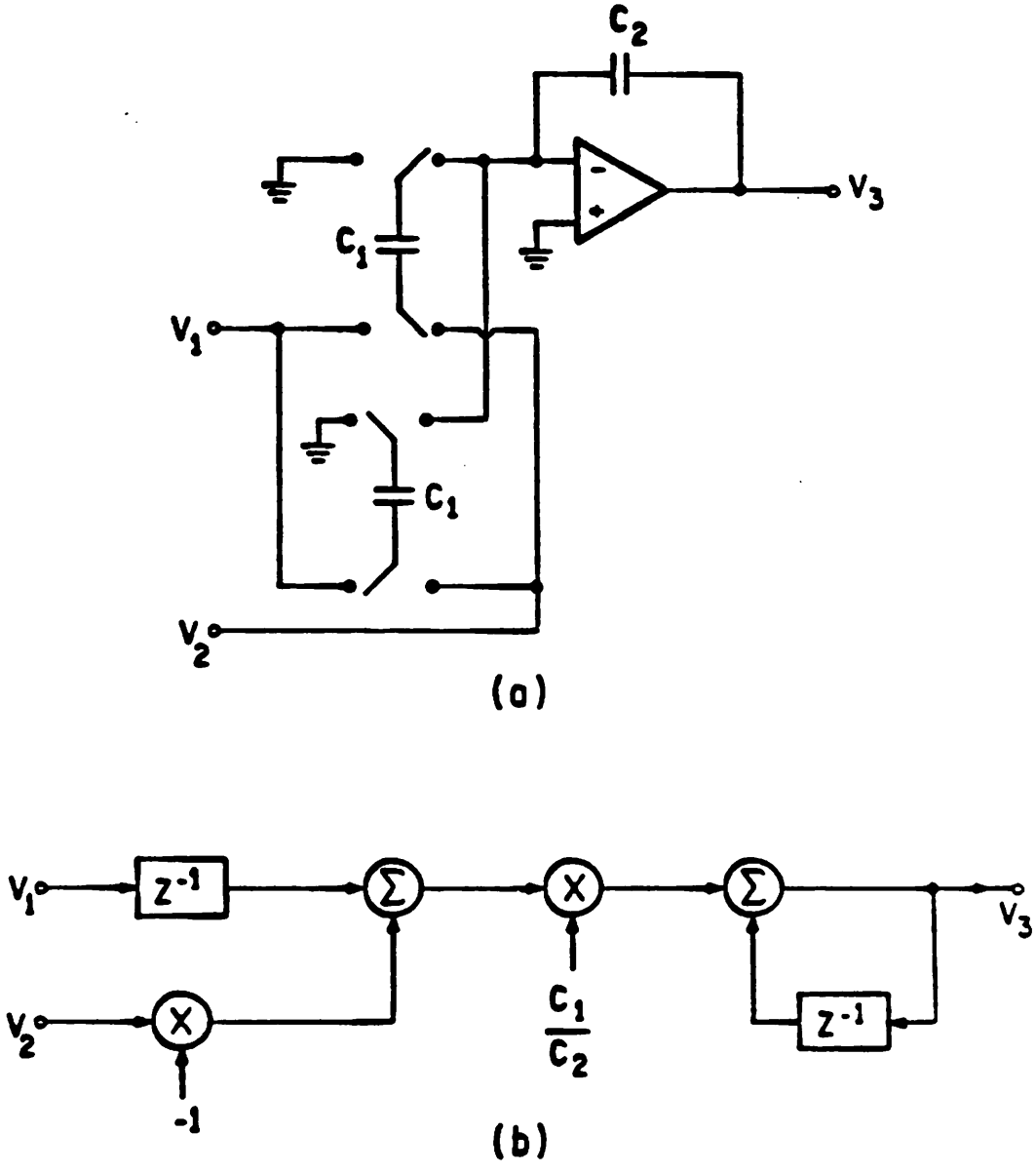


Fig. 3.9 (a) Single-ended LDI integrators which utilize both clock phases.
(b) Corresponding z-domain block diagram.

that it takes one delay for the signal to go from the non-inverting input to the output whereas there is a delay-free path for the inverting input.

For the case of a fully differential integrator, the circuit which utilizes both clock phases is shown in Fig. 3.10a. Since inversion of signal can be realized by simply flipping the differential signal lines, there is no distinction between an inverting or a non-inverting input. This circuit is described by the following difference equation:

$$C_2 V_3[(n+1)T_s] = C_2 V_3(nT_s) + C_1 V_1(nT_s) + C_1 V_2[(n+1)T_s] . \tag{3.5}$$

As before

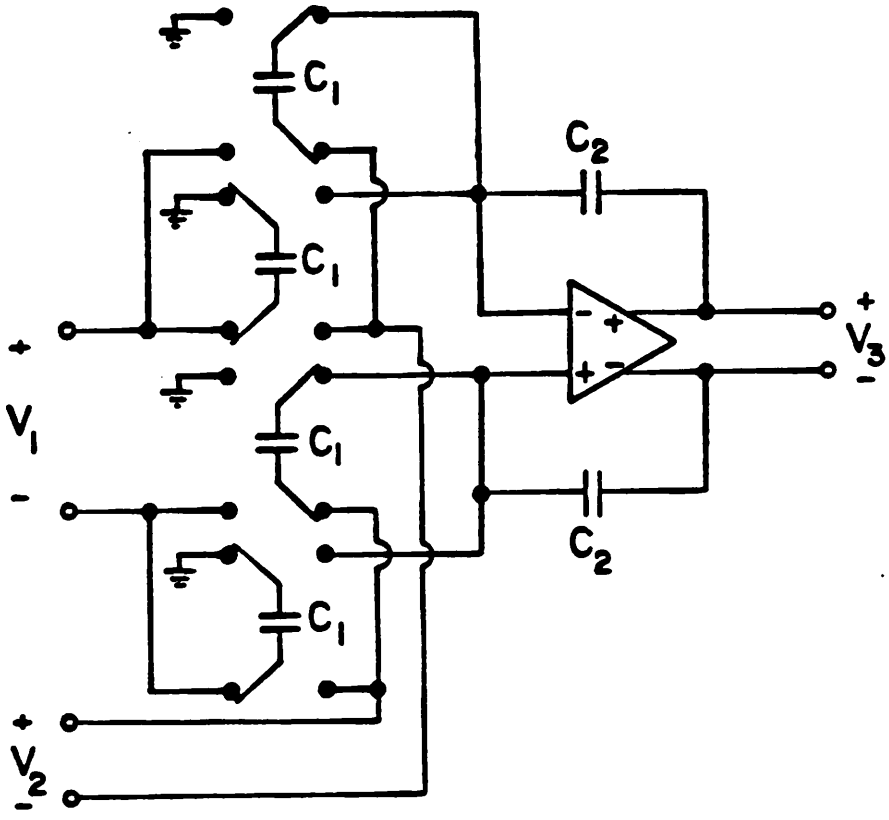
$$T_s = \frac{T_c}{2},$$

and in z-transform notation, we can write

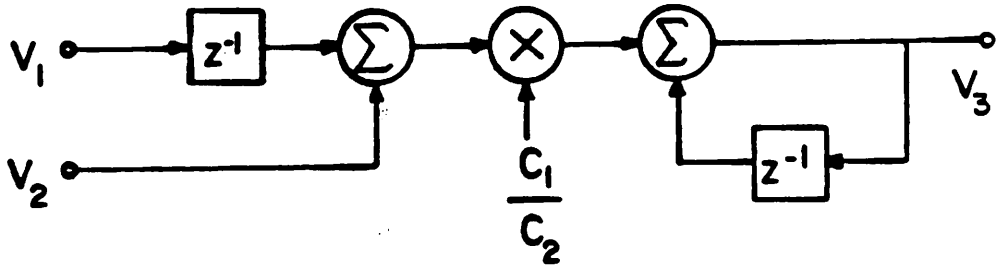
$$V_3 = \frac{C_1}{C_2} \frac{1}{1-z^{-1}} (z^{-1} V_1 + V_2) . \tag{3.6}$$

The z-domain block diagram for (3.6) is shown in Fig. 3.10b. This circuit is more flexible than the single-ended case due to the availability of both inverting and non-inverting inputs which can be either delay-free or takes one delay to propagate to the output.

To realize the LDI ladder structure with these double-sampling integrators, we first notice that for leapfrog filters, each loop consists of an inverting and a non-inverting integrator. For the single-ended case, the inverting integrator is delay free whereas the non-inverting integrator needs one delay. Thus, there is one delay in each loop, as required by the LDI configuration. Such a two integrator loop is shown in Fig. 3.11, with corresponding z-domain block diagram shown in Fig. 3.12. This block diagram is similar to the one shown in Fig. 2.7, and if we divide the delay in the loop equally between the two integrators, they will be exactly identical. Likewise, for the fully differential case, we need one integrator



(a)



(b)

Fig. 3.10 (a) Fully differential LDI integrator which utilizes both clock phases.
(b) Corresponding z-domain block diagram.

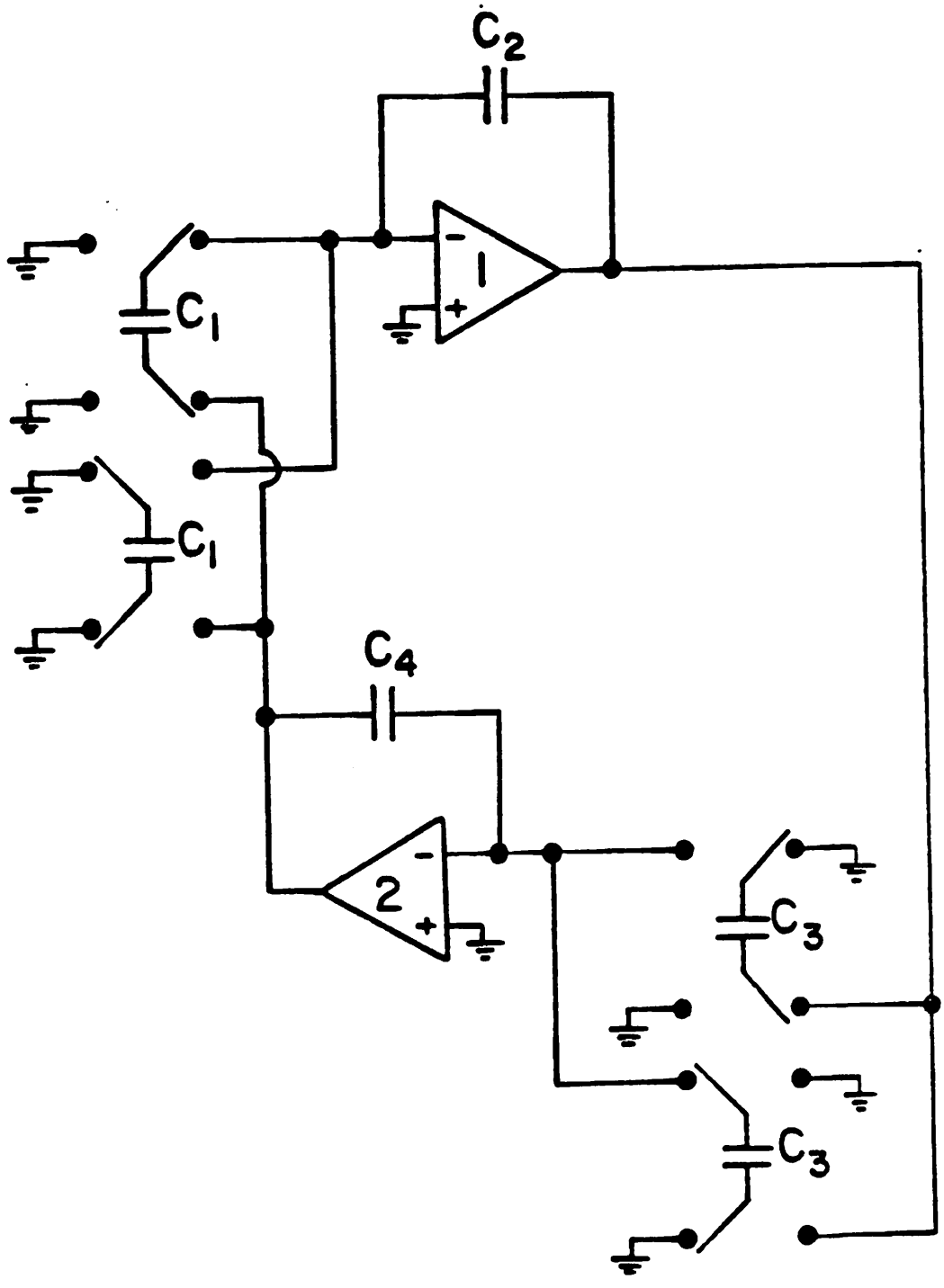


Fig. 3.11 A two-integrator loop using single-ended double-sampling integrators.

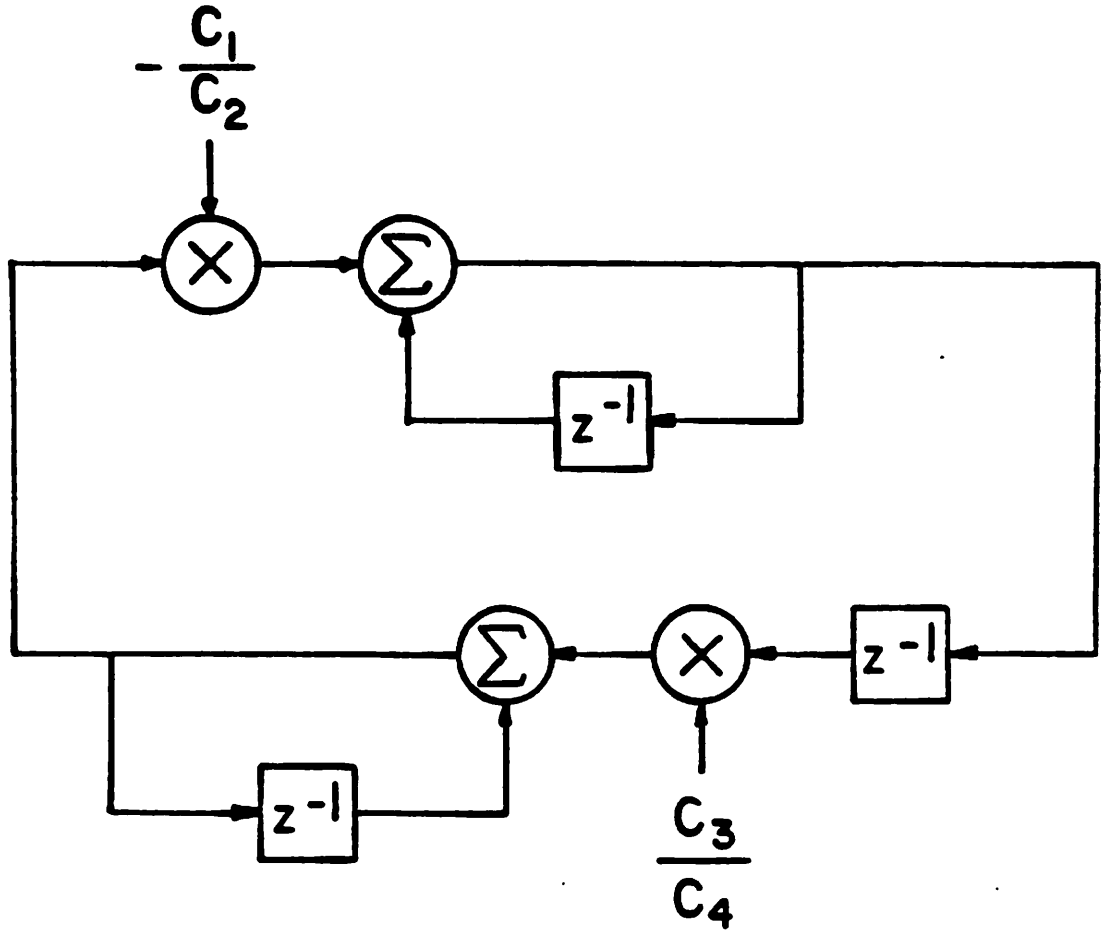


Fig. 3.12 z-domain block diagram of the loop shown in Fig. 3.11.

with a delay-free input and one which takes one delay to transfer the signal. Here, we can choose either the inverting or non-inverting integrator as delay-free since we can easily realize a sign change by inverting the differential signal. The circuit which realizes a two integrator loop for this case is shown in Fig. 3.13. The z-domain block diagram for this circuit is also represented by Fig. 3.12.

We have shown that both the single-ended and fully differential double-sampling integrators can implement the LDI transformation throughout the ladder. However, the fully differential case is more flexible in realizing the resistive terminations. This is because a termination loop requires negative feedback, so that the termination must be fed back to the inverting input, which is delay-free for the single-ended case. Thus the termination for the single-ended double-sampling case is necessarily of the minus one half delay type. This is shown in Fig. 3.14a and b. For the fully differential case, we can realize either minus one half delay terminations (Fig. 3.15a and b) or extra half delay delay terminations (Fig. 3.16a and b). Thus, complex-conjugate configurations can be implemented with fully differential double-sampling integrators. It should be pointed out that for bandpass filters, the single-ended double-sampling integrator can also realize extra half delay terminations. This is because bandpass leapfrog structures have localized feedback loops which are not shared by the other parts of the circuit, thus the extra half delay termination can be accomplished by inverting the signal through the local loop and fed back via the non-inverting input, as shown in Fig. 3.17.

In using double-sampling integrators, one must be careful to make sure that both clock phases are exactly equally spaced out, i.e., the time from the falling edge of one clock phase to that of the other must be constant. In other words, we need $\tau_1 = \tau_2$ in Fig. 3.18. This is because we have assumed the delay in both clock phases are identical in this configuration. Nevertheless, in high

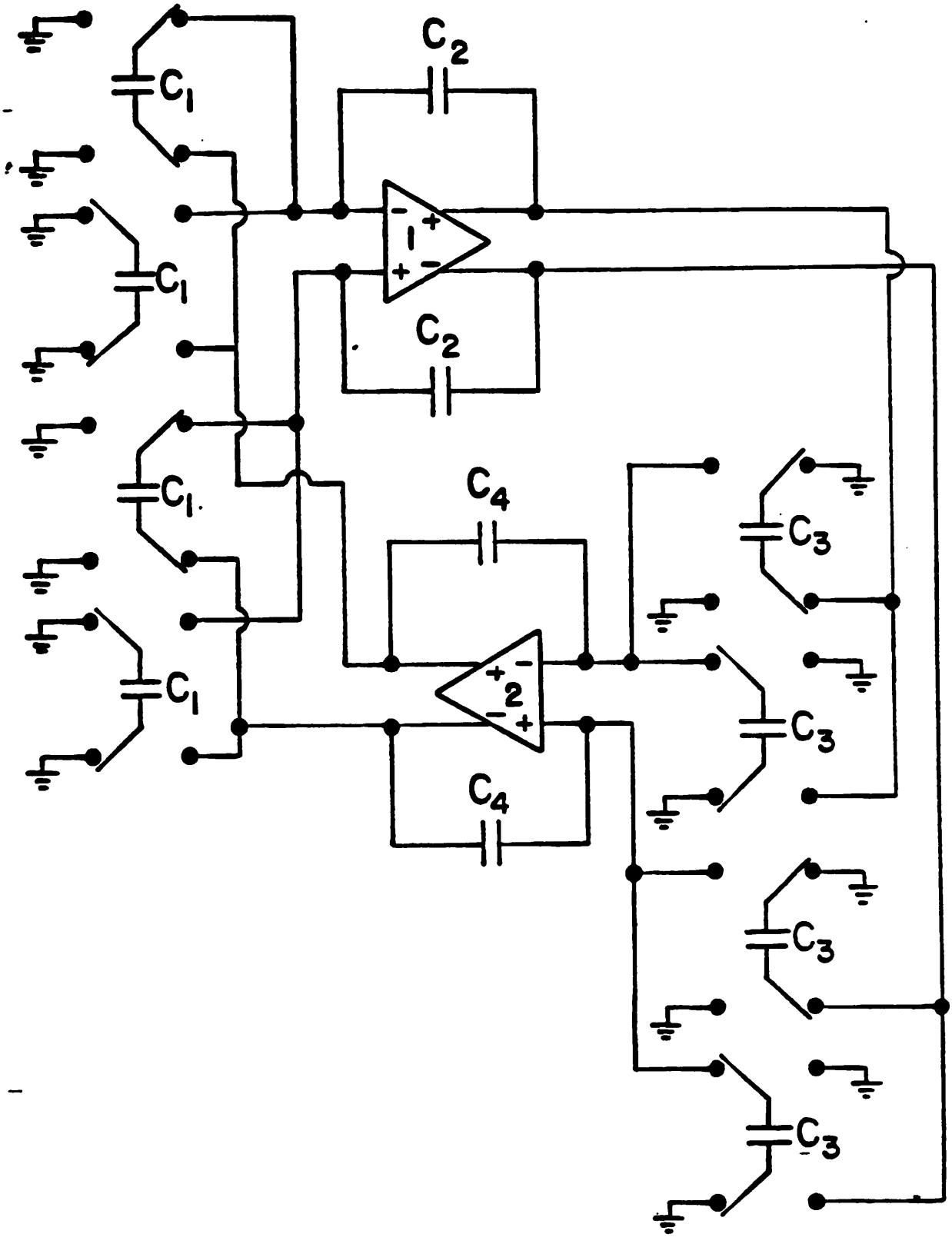


Fig. 3.13 A two-integrator loop using fully differential double-sampling integrators.

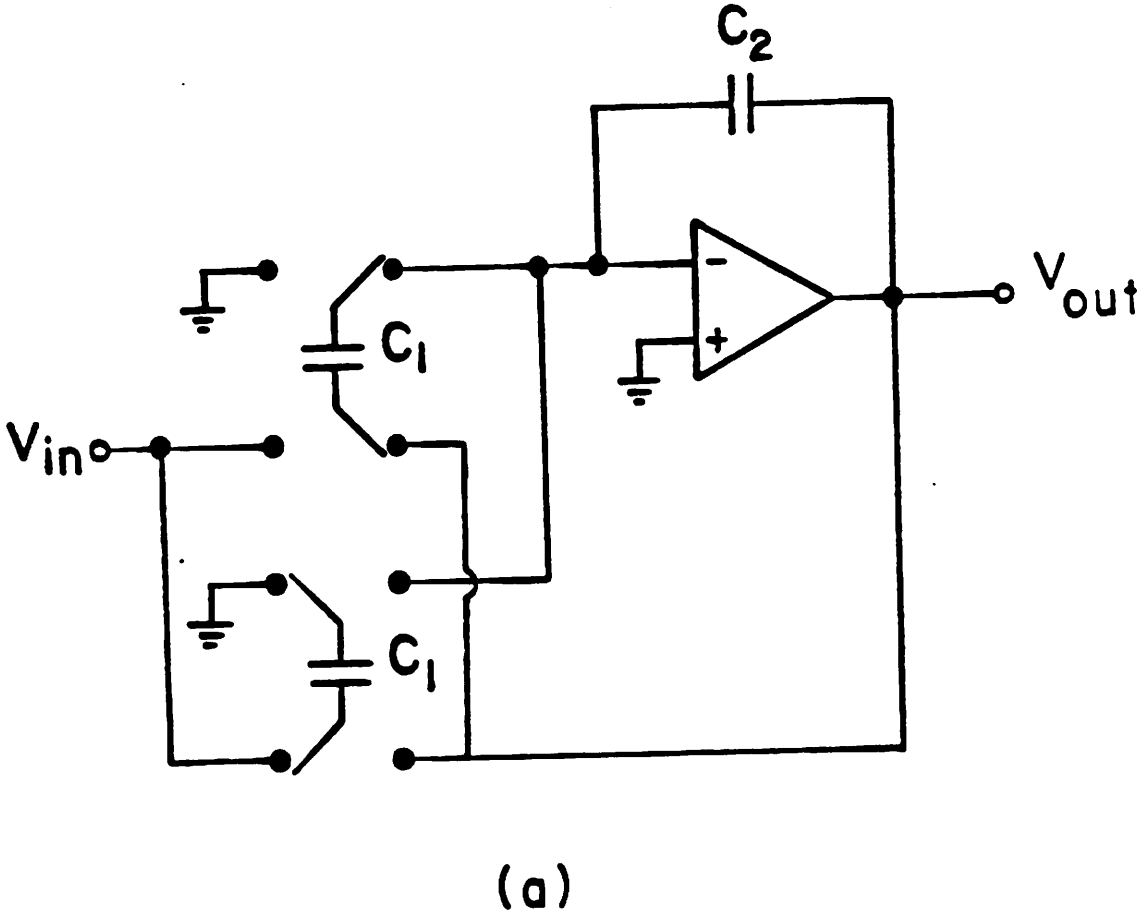
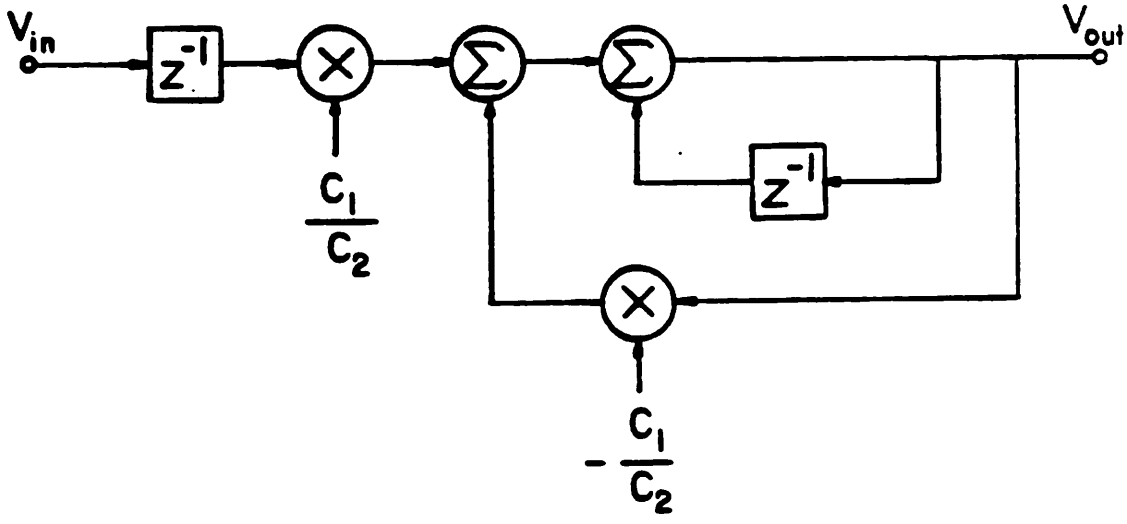
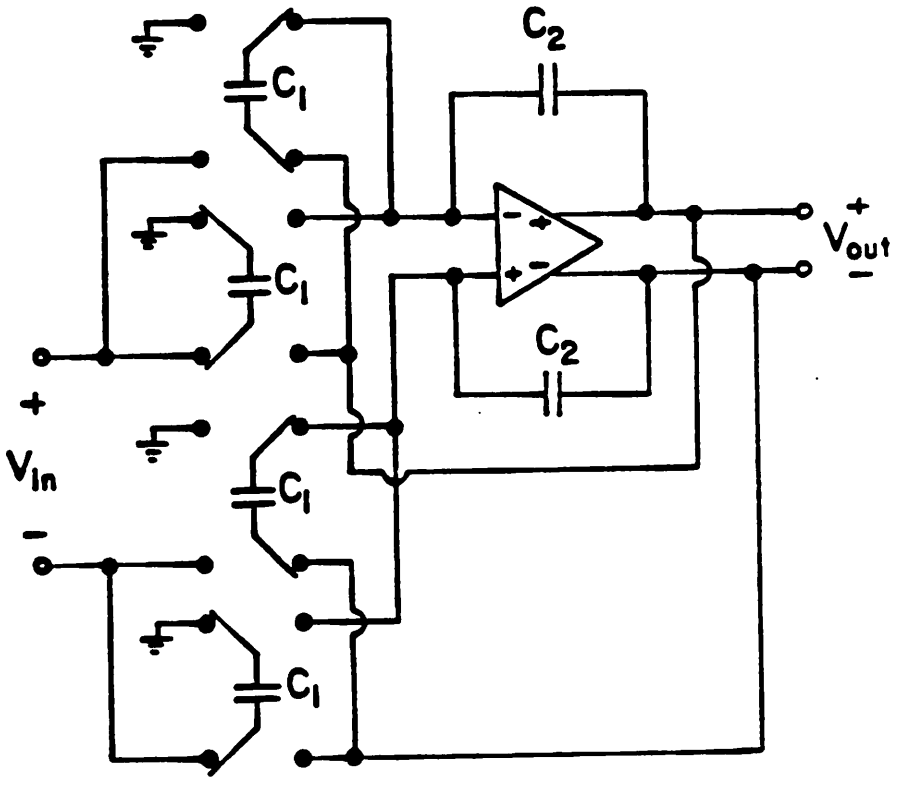


Fig. 3.14 (a) Single-ended double-sampling integrator with a minus one half delay termination.



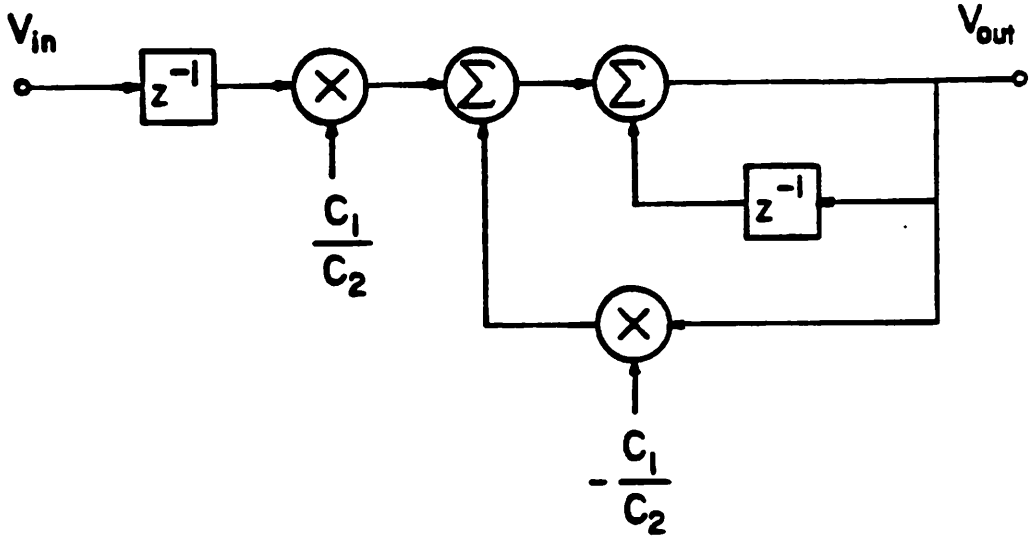
(b)

Fig. 3.14 (b) s-domain block diagram representation.



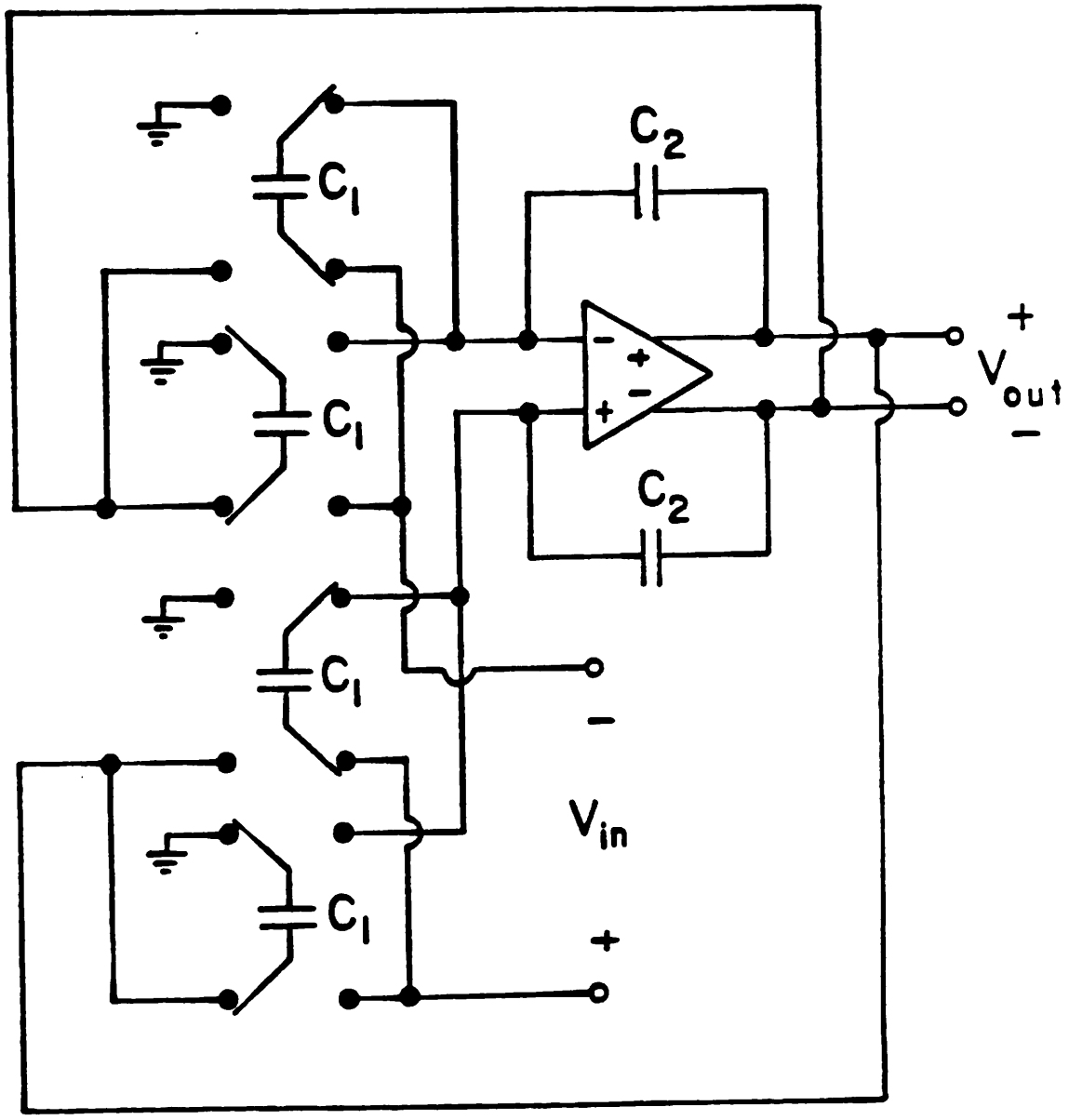
(a)

Fig. 3.15 (a) Fully differential double-sampling integrator with a minus one half delay termination.



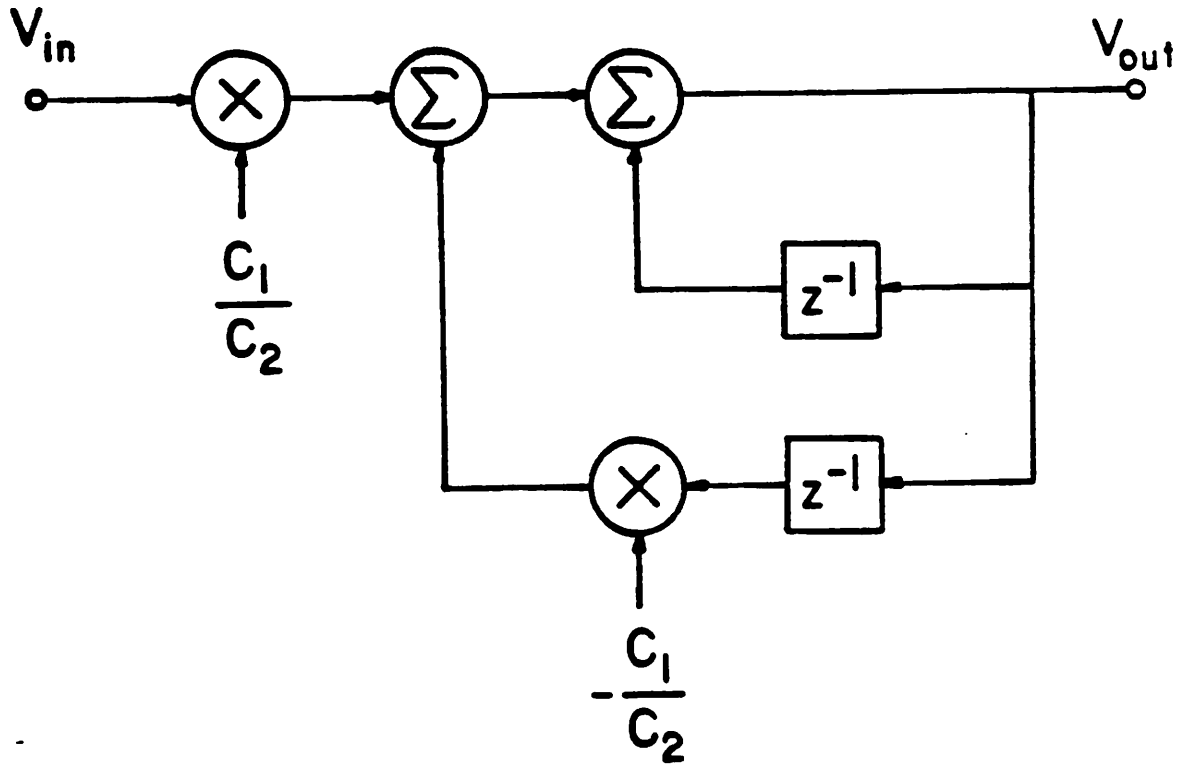
(b)

Fig. 3.15 (b) z-domain block diagram representation.



(a)

Fig. 3.16 (a) Fully differential double-sampling integrator with an extra half delay termination.



(b)

Fig. 3.16 (b) z-domain block diagram.

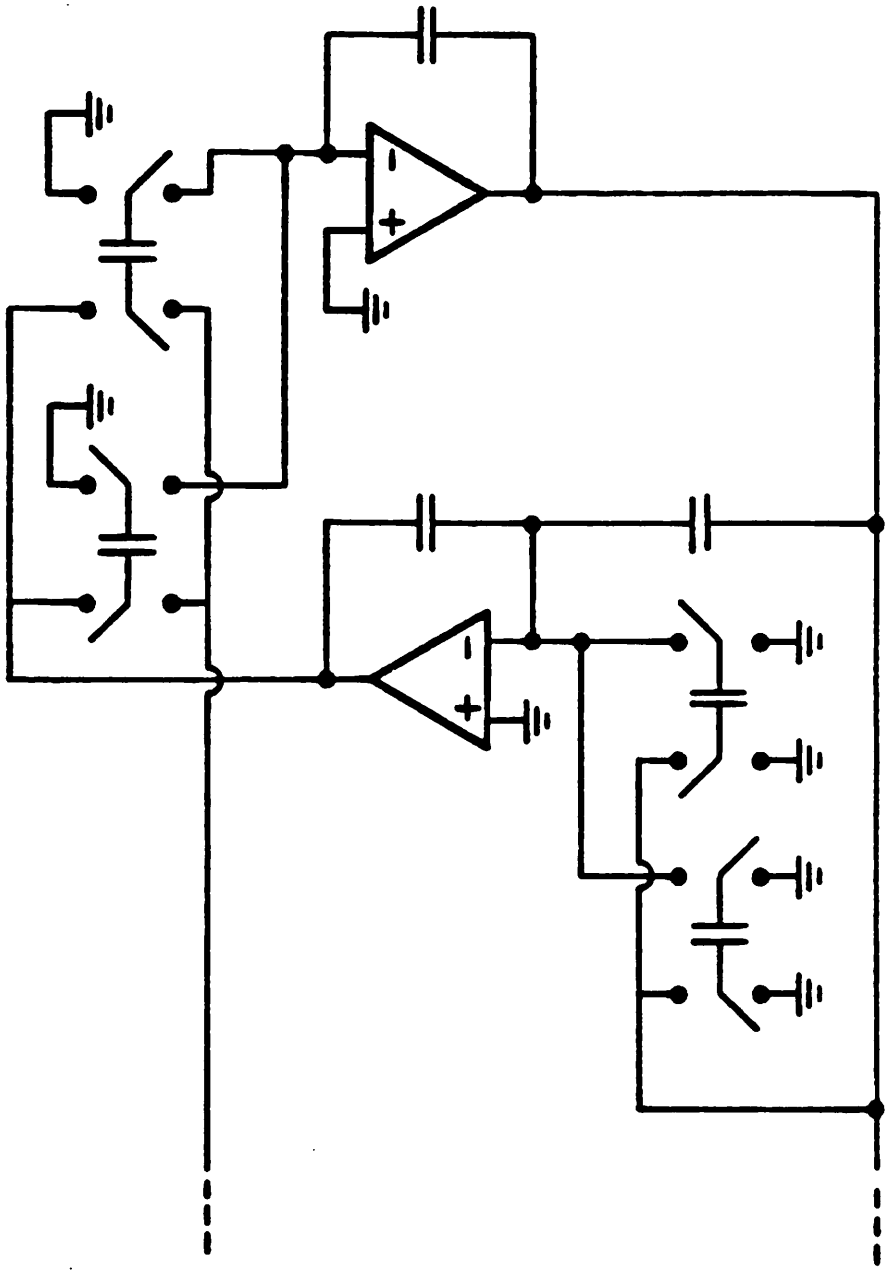


Fig. 3.17 Termination with an extra half delay in terms of LDI transformation in single-ended double-sampling bandpass filter.

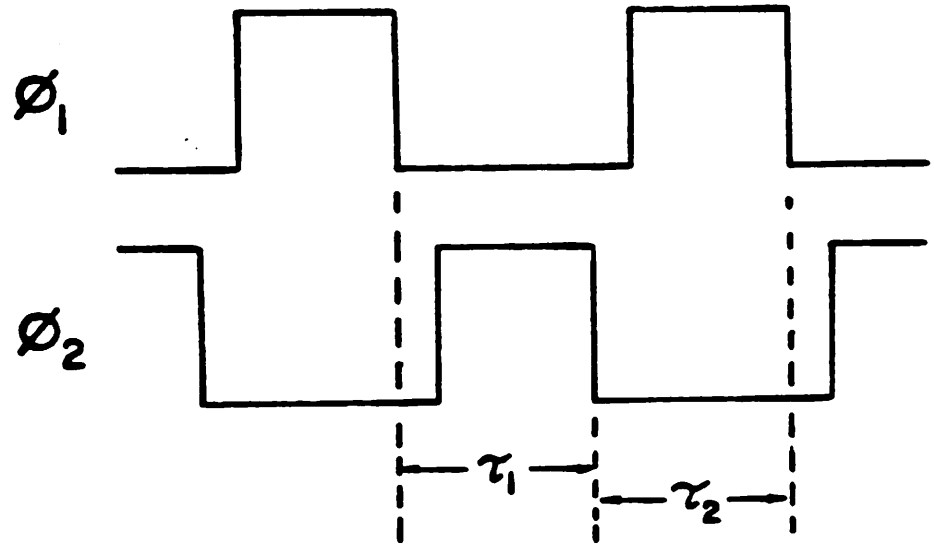


Fig. 3.18 Non-overlapping clock phases used in switched-capacitor filters.

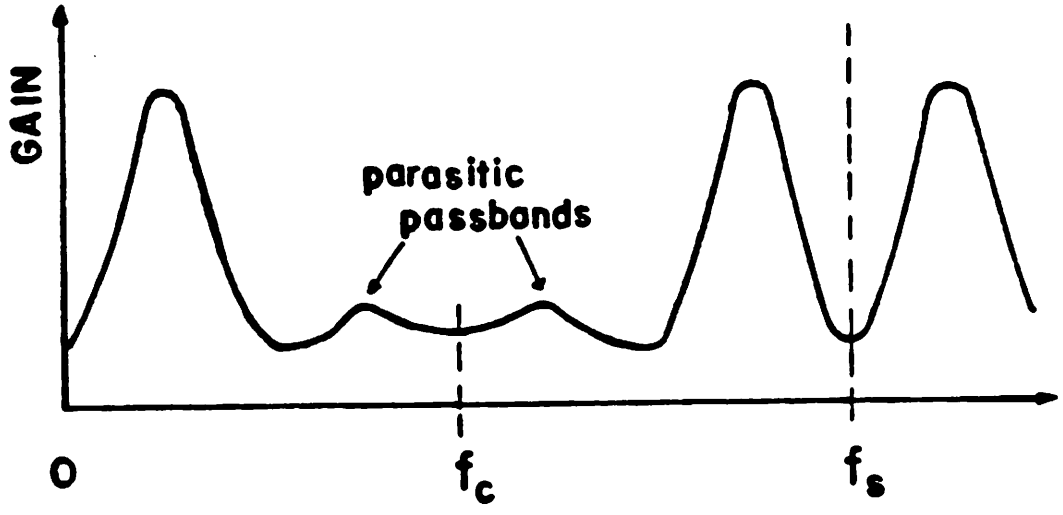


Fig. 3.19 Parasitic passbands due to non-ideal clock phases.

frequency operation, it is difficult to achieve this due to the finite fall times as well as unavoidable phase jitter. This will result in two problems. First, there will be parasitic passbands around the clock frequency f_c, which is half the effective sample rate f_s (Fig. 3.19). This will degrade the stopband rejection, but is usually negligible since the parasitic passbands are typically small and are way out in the stopband. Second, signals around f_c may feed through and alias back inside the passband. This feedthrough is usually small but will affect the filter dynamic range. To solve this problem, we must make sure that the response of the anti-alias filter must have rolled off by 20 dB or so at f_c so that signals around f_c are first attenuated before feeding into the switched-capacitor filter. Most anti-alias filters would have rolled off by 40 to 60 dB at f_s so that this additional requirement is not a tight restriction.

3.5 Design of 260 KHz AM IF Filter

This section describes the design details for a 260 KHz AM IF filter. This filter is for channel separation in the IF stage of the receiver. As mentioned earlier, the conventional IF frequency for the AM band is 455 KHz, but for fully integrated car radios, the IF can be chosen to be at 260 KHz. The specifications for such a filter is shown in Table 3.2, with typical values shown in Fig. 3.20. This is a highly selective filter, with Q on the order of 40. For this filter, we use an elliptic approach with identical resonator architecture. This results in a sixth order filter with coupling capacitor ratios on the order of 100, which is reduced to around 10 by T-network schemes.

The design of such a filter starts from the LC prototype network. Since the LDI transformation warps the frequency axis, we have to predistort the frequency specifications by equation (2.6). A summary of these numbers is shown in Table 3.3. We notice that the frequency values change by approximately 1%, which is rather significant in such a high Q filter. Thus, predistortion is necessary for this case of a 4 MHz sample rate.

Next, these specifications are transformed to those for a normalized lowpass prototype filter. This is done by classical bandpass to lowpass transformation techniques [27], with normalizing frequency ω_n given by

$$\begin{aligned}\omega_n &= 2\pi\left[\sqrt{(254.77)(261.62)}\ KHz\right] \\ &= 2\pi\left[258.17 KHz\right]\end{aligned} \tag{3.7}$$

and normalizing bandwidth B given by

$$\begin{aligned}B &= 2\pi\left[(261.62 - 254.77)\ KHz\right] \\ &= 2\pi\left[6.85\ KHz\right].\end{aligned} \tag{3.8}$$

The obtained lowpass specifications are shown in Fig. 3.21. Here, the frequency axis is in radians per second. The LC ladder which will satisfy these

center frequency	260 KHz ±1%
ripple content	3 dB
-3 dB bandwidth	5 KHz min.
rejection at ±10 KHz	30 dB min.
stopband rejection	55 dB min.
clock frequency	4 MHz
gain	-6 dB

Table 3.2 Specifications for 260 KHz AM IF filter

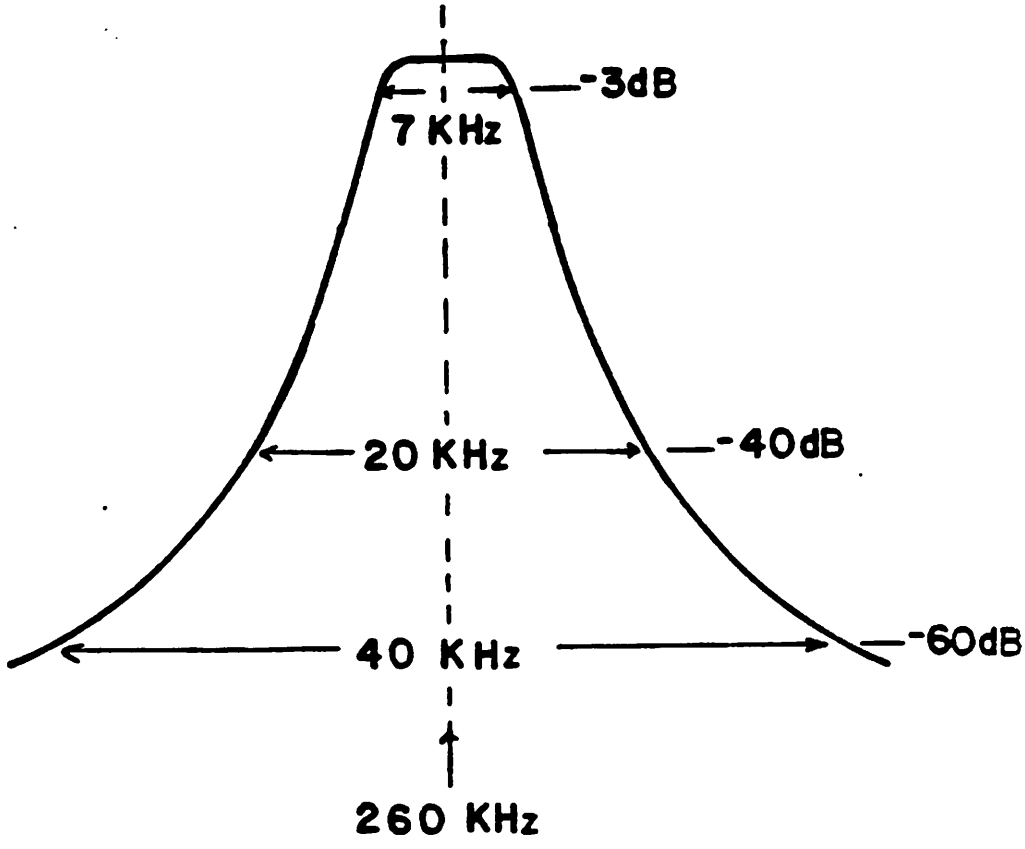


Fig. 3.20 Typical frequency response of 260 KHz AM IF filter.

	Typical Values	Predistorted Values (4 MHz)
center frequency	260 KHz	258.2 KHz
-3 dB bandwidth from to	 256.5 KHz 263.5 KHz	 254.77 KHz 261.62 KHz
-40 dB bandwidth from to	 250 KHz 270 KHz	 248.40 KHz 267.98 KHz
-60 dB bandwidth from to	 240 KHz 280 KHz	 238.58 KHz 277.75 KHz

Table 3.3 Typical frequency specifications and the predistorted values for 260 KHz AM IF filter

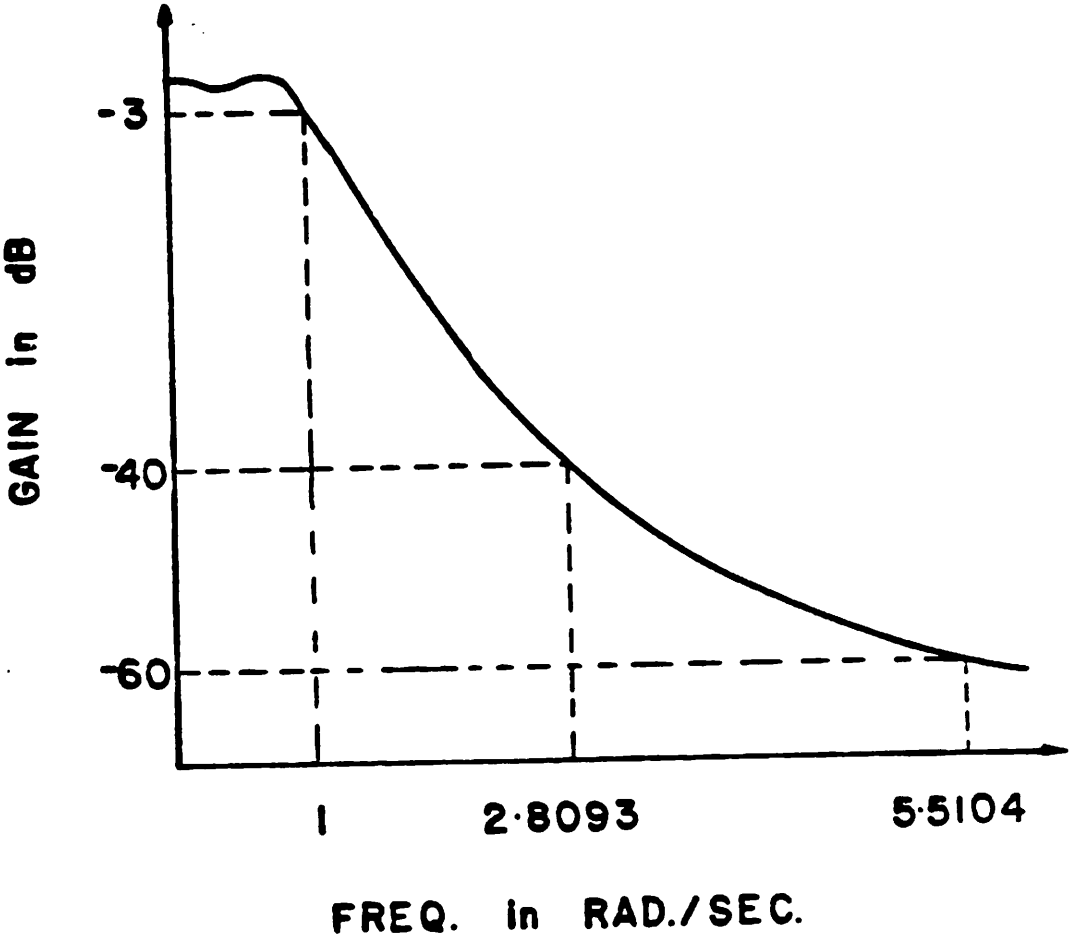


Fig. 3.21 Lowpass prototype specifications.

requirements is shown in Fig. 3.22. These component values are obtained from the filter synthesis program FILSYN [28]. To obtain the denormalized bandpass LC ladder, the lowpass L,C elements are transformed into the bandpass L,C elements. This is shown in Fig. 3.23, where the lowpass inductor L_{LP} is transformed into a series inductor L_1 and capacitor C_1 given by

$$L_1 = \frac{L_{LP}}{B} \tag{3.9a}$$

$$C_1 = \frac{B}{\omega_n^2 L_{LP}}. \tag{3.9b}$$

Likewise, the lowpass capacitor C_{LP} is transformed into a parallel inductor L_2 and capacitor C_2 given by

$$L_2 = \frac{B}{\omega_n^2 C_{LP}} \tag{3.10a}$$

$$C_2 = \frac{C_{LP}}{B}. \tag{3.10b}$$

The bandpass ladder obtained from this transformation is shown in Fig. 3.24. As mentioned earlier in Section 2.4, elliptic bandpass ladders have D.C. stability problems in active realizations. Using the Thevenin equivalent circuit technique discussed, we arrive at the final LC circuit shown in Fig. 3.25.

To implement this doubly-terminated LC ladder in switched-capacitor form, we used the identical resonator filer architecture (Section 2.6) to minimize the sensitivity problems. The resulting signal flow graph which realize this circuit is shown in Fig. 3.26. We notice that large ratios are required for the coupling paths between the resonators. Therefore, the T-network scheme discussed in Section 3.3 is used. This results in the signal flow graph of Fig. 3.27, with corresponding fully differential switched-capacitor realization shown in Fig. 3.28.

Here, double-sampling integrators are not needed because the clock frequency is not very high (4 MHz). The switch phasing is identical to that of Fig.

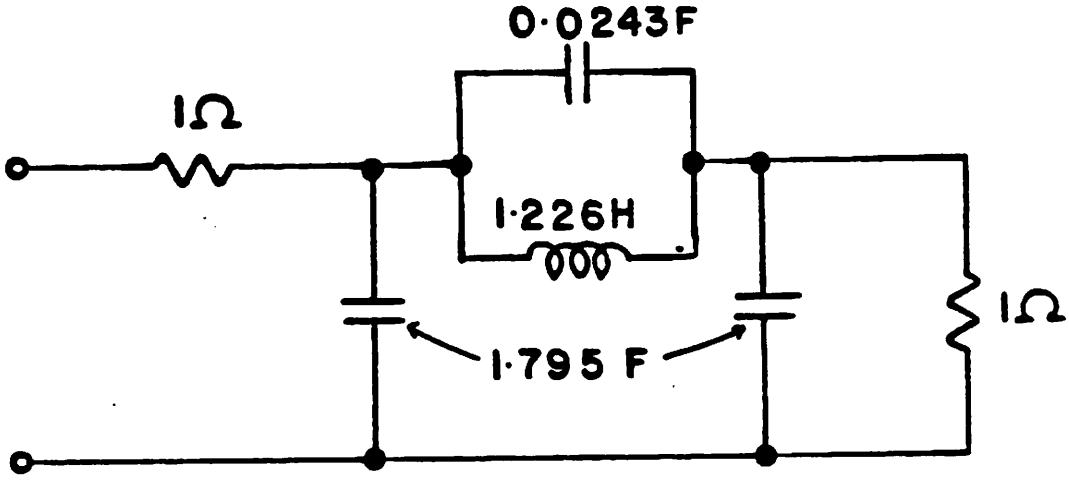


Fig. 3.22 Lowpass LC ladder.

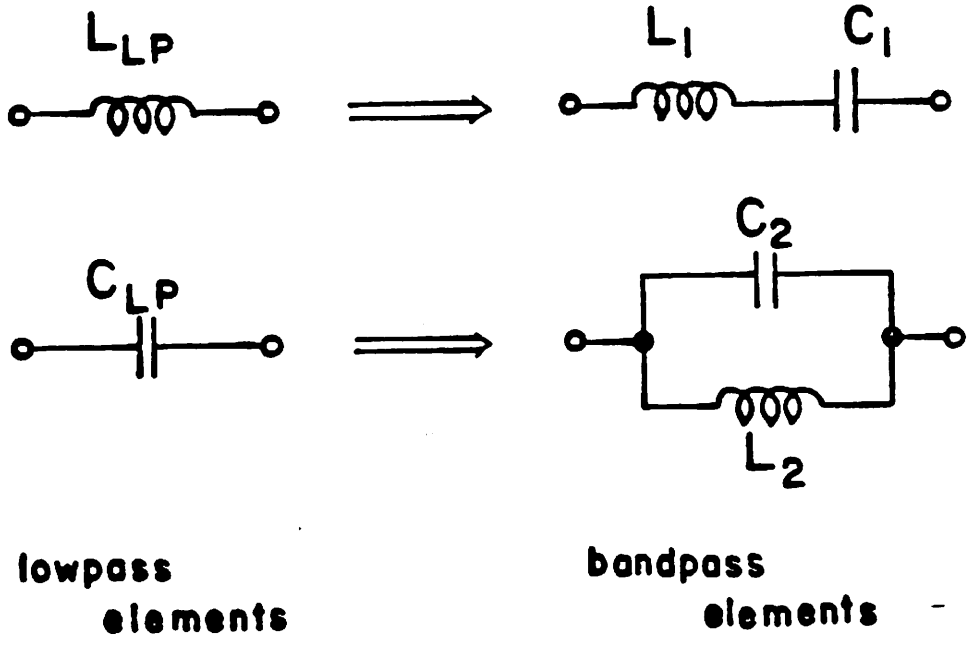
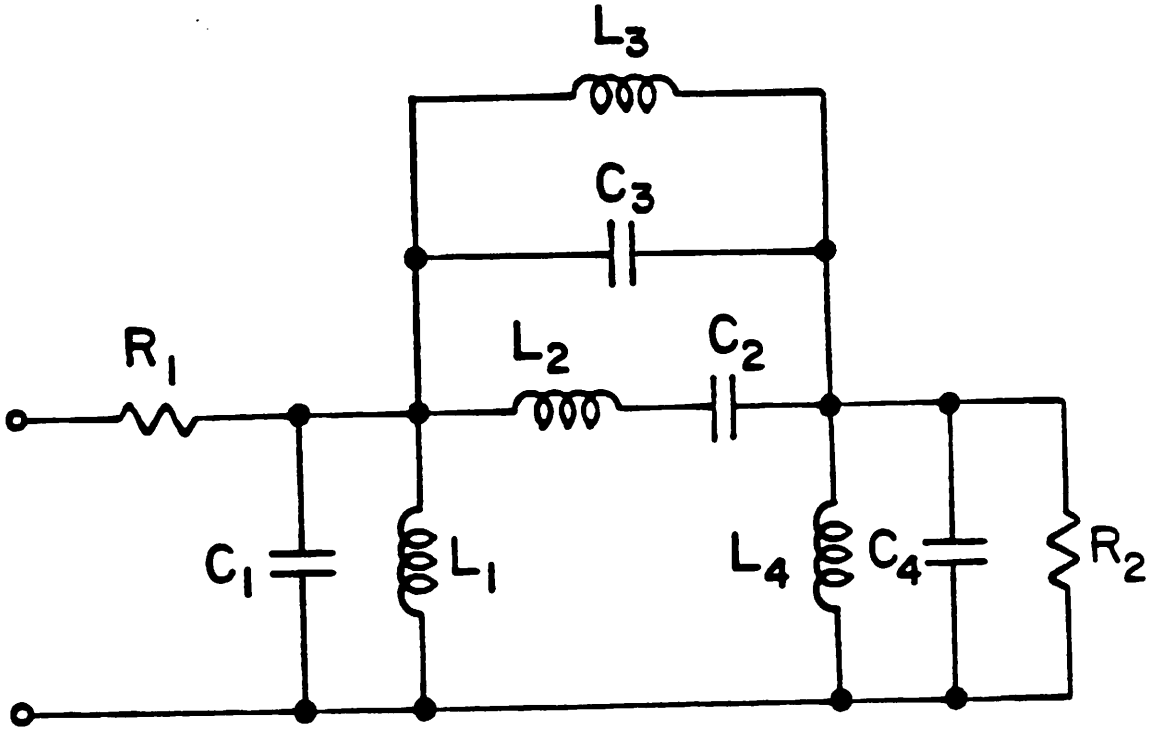


Fig. 3.23 Lowpass to bandpass transformation



$C_1 = C_4 = 41.681\ \mu F$

$L_1 = L_4 = 9.1304\ nH$

$C_2 = 13.368\ nF$

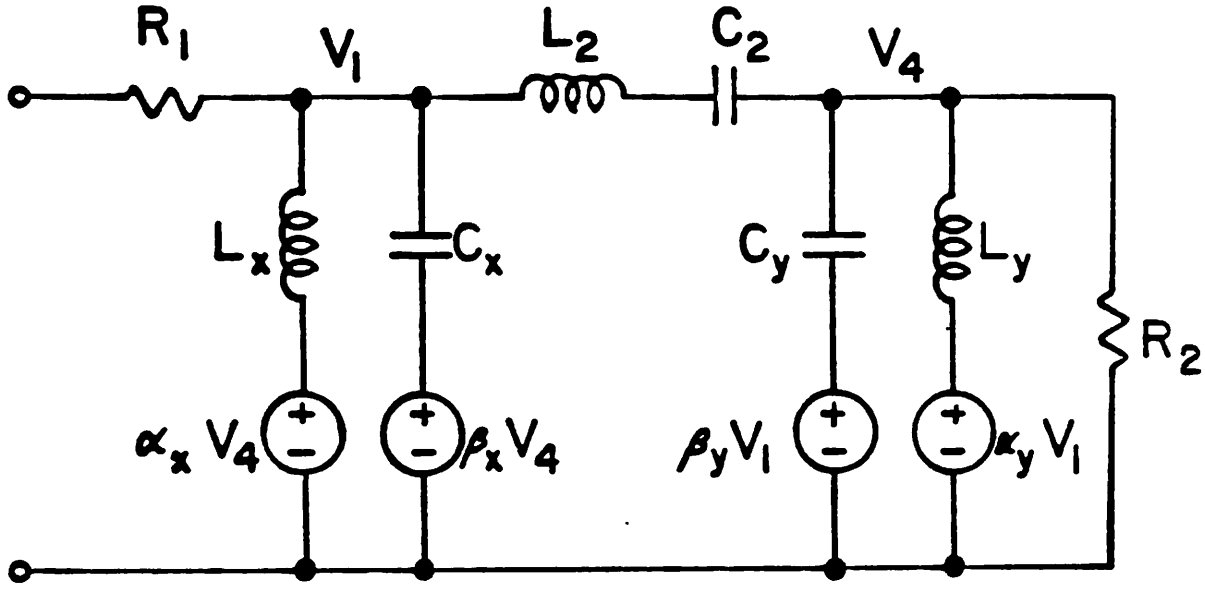
$L_2 = 28.468\ \mu H$

$C_3 = 0.56431\ \mu F$

$L_3 = 0.67439\ \mu F$

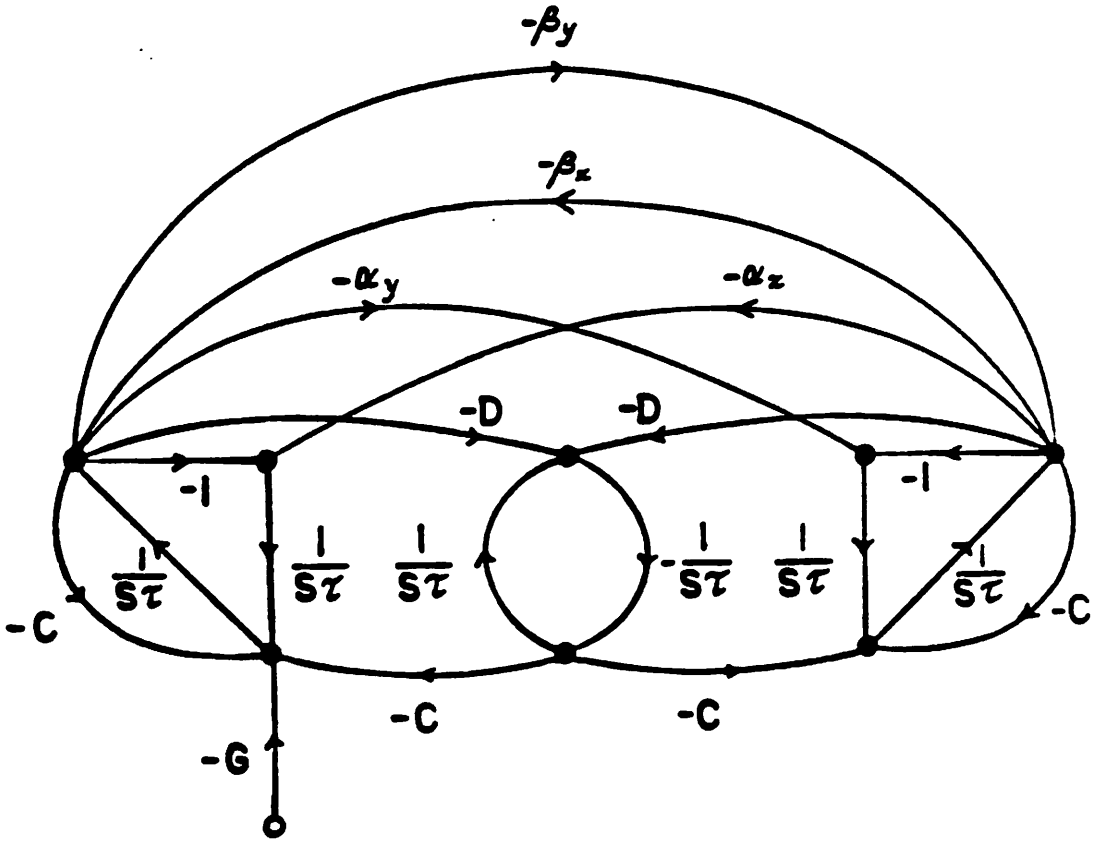
$R_1 = R_2 = 1\ ohm$

Fig. 3.24 **Doubly-terminated LC elliptic bandpass ladder.**



$$L_x = L_y = 9.0084 \text{ nH}$$
$$C_x = C_y = 42.246\ \mu\text{F}$$
$$C_2 = 13.368 \text{ nF}$$
$$L_2 = 28.468\ \mu\text{H}$$
$$\alpha_x = \alpha_y = \beta_x = \beta_y = 0.013358$$
$$R_1 = R_2 = 1 \text{ ohm}$$

Fig. 3.25 **Thevenin equivalent circuit of Fig. 3.24.**



$$\alpha_x = \alpha_y = \beta_x = \beta_y = \gamma = 0.013358$$

$$\frac{1}{f_s T} = 0.40525$$

$$C = 0.0146$$

$$D = 0.02167$$

$$G = 0.014115$$

Fig. 3.26 Signal flow graph of Fig. 3.25.

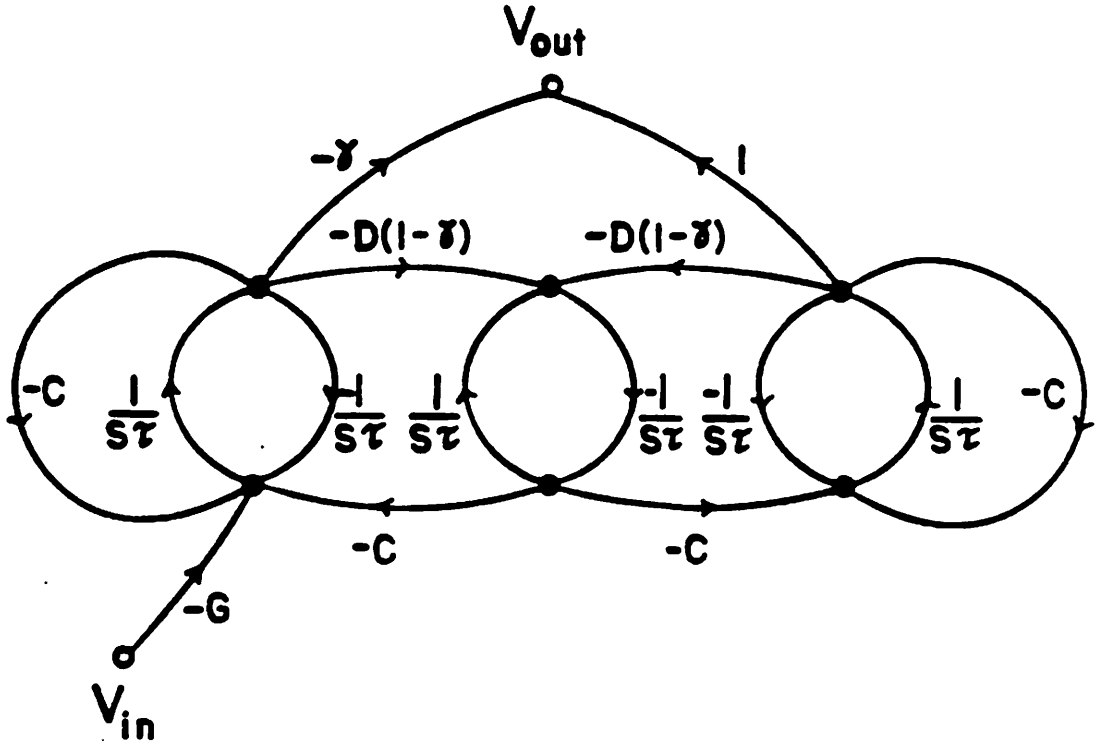


Fig. 3.27 Signal flow graph for T-network scheme

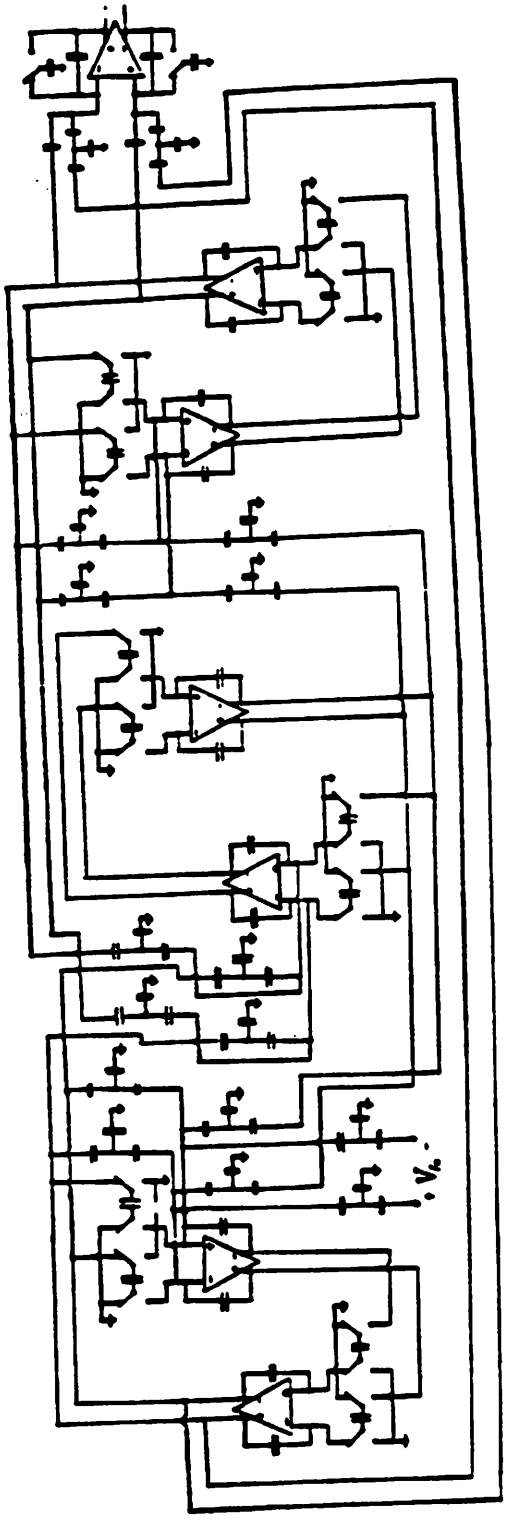


Fig. 3.28 Fully differential switched-capacitor filter for AM IF applications.

2.10, with a half cycle delay per integrator. Also, extra half delay terminations are used in both the source and the load. The complex-conjugate configuration is not used to avoid delay-free paths, which may complicate the op. amp. settling behavior. Also, the clock to center frequency ratio is large (~16), so that complex-conjugate terminations are not needed in the first place. This circuit is simulated on the computer using the digital filter analysis program DINAP [29], where we assume ideal op. amps. and perfect settling in modeling the filter. The simulated frequency response is shown in Figs. 3.29a and b, and the performance characteristics are summarized in Table 3.4. Comparing Tables 3.2 and 3.4, we notice that the simulated results meet all the specifications, with adequate margin for experimental deviations. Several sensitivity simulations are also done using DINAP by varying the capacitor ratios randomly. These results are summarized in Table 3.5. We notice that that variations in the integrator capacitor ratios directly affect the center frequency without changing the overall shape, as expected. Hence, the integrator capacitors have to be accurate, since a 1% change will shift the center frequency by 1%. On the other hand, we found that varying the coupling capacitor ratios by say 5% is still acceptable and within specifications. Hence, the accuracy requirements for these coupling capacitors are not stringent. In addition to sensitivity analysis, simulations are also done to evaluate the effects of finite amplifier gain. Again, DINAP is used in the simulations where we assume the op. amps. have perfect settling. These results are summarized in Table 3.6. We notice that an amplifier gain of 500 is adequate for this application. The passband response with a gain of 500 is plotted in Fig. 3.30, together with the ideal curve for comparison.

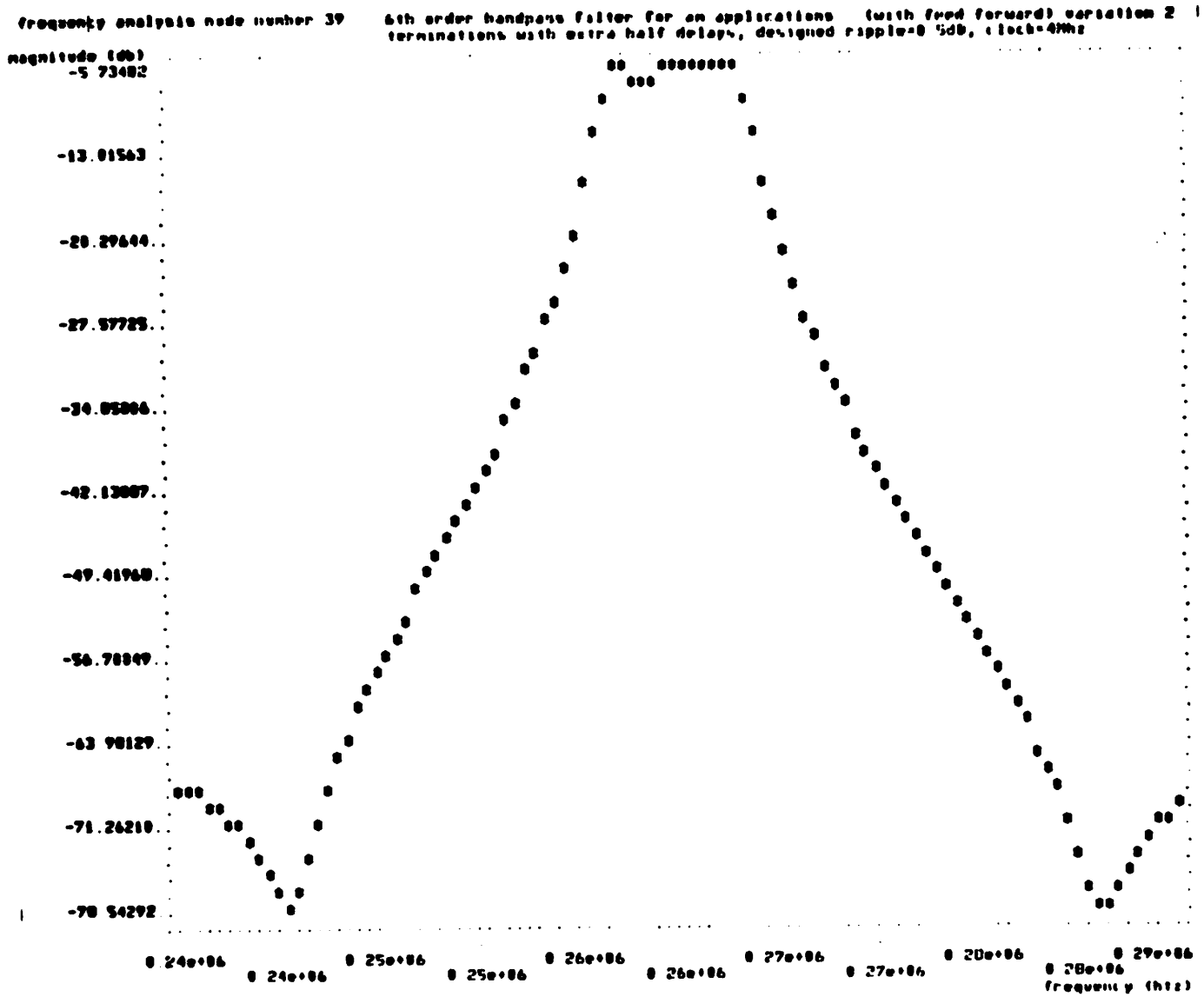


Fig. 3.29 (a) Simulated overall response of AM IF filter.

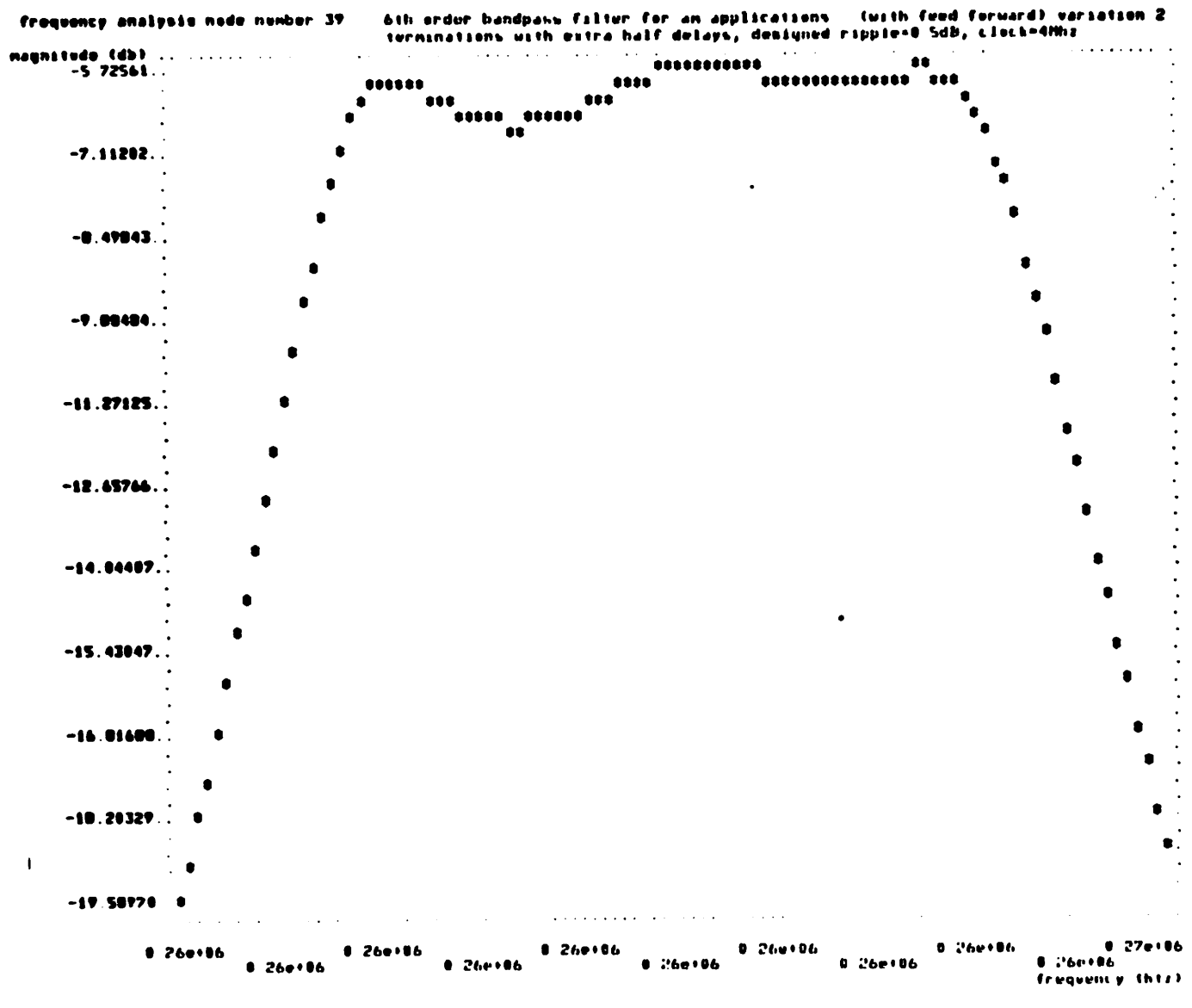


Fig. 3.29 (b) Simulated detailed passband response of AM IF filter.

center frequency	260 kHz
ripple content	0.97 dB
-3 dB BW	from 256.4 KHz to 263.6 KHz = 7.2 KHz
-30 dB BW	from 251.7 KHz to 268.6 KHz = 16.9 KHz
-40 dB BW	from 248.9 KHz to 271.7 KHz = 22.8 KHz
-60 dB BW	from 243.0 KHz to 278.2 KHz = 35.2 KHz
± 10 KHz BW rejection	34.7 dB
min stopband rejection	60.4 dB
Gain	-6 dB

Table 3.4 Simulated performance of the sixth order elliptic bandpass filter circuit of Fig. 3.28

	Specification	Nominal Reponse	Decrease integrator capacitor ratio by 1 %	Decrease coupling capacitor ratio by 5 %	Decrease smaller summing capacitor ratio by 5 %	Decrease all ratios at the same time
f_0	260 KHz ± 1 %	260 KHz	257.4 KHz	260 KHz	260 KHz	257.4 KHz
Ripple content	3 dB	0.97 dB	0.97 dB	0.98 dB	0.97 dB	0.97 dB
-3 dB bandwidth	5 KHz	7.2 KHz	7 KHz	6.8 KHz	7.2 KHz	6.7 KHz
± 10 KHz rejecton	30 dB	34.7 dB	35.2 dB	36.5 dB	34.6 dB	36.6 dB
Stop band rejecton	55 dB	60.4 dB	60.4 dB	60.4 dB	61.1 dB	61.1 dB

Note:
Increasing the ratio changes the reponse by approximately the same amount in the opposite direction

Table 3.5 Sensitivity analysis of sixth order elliptic bandpass filter

	specs	gain =∞	1000	500
ripple	3 dB	0.97 dB		
-3 dB BW	5 KHz min	7.2 KHz	6.85 KHz	6 KHz
f_0	260 KHz ± 1 %	260 KHz	259.68 KHz	251.35 KHz
± 10 KHz BW rejection	30 dB min	34.7 dB	33.4 dB	32 dB
min stopband rejection	55 dB min	60.4 dB	58.8 dB	57.3 dB
overall gain	-6 dB	-6 dB	-7.6 dB	-9.0 dB

Table 3.6 Effect of finite amplifier gain in sixth order elliptic bandpass filter

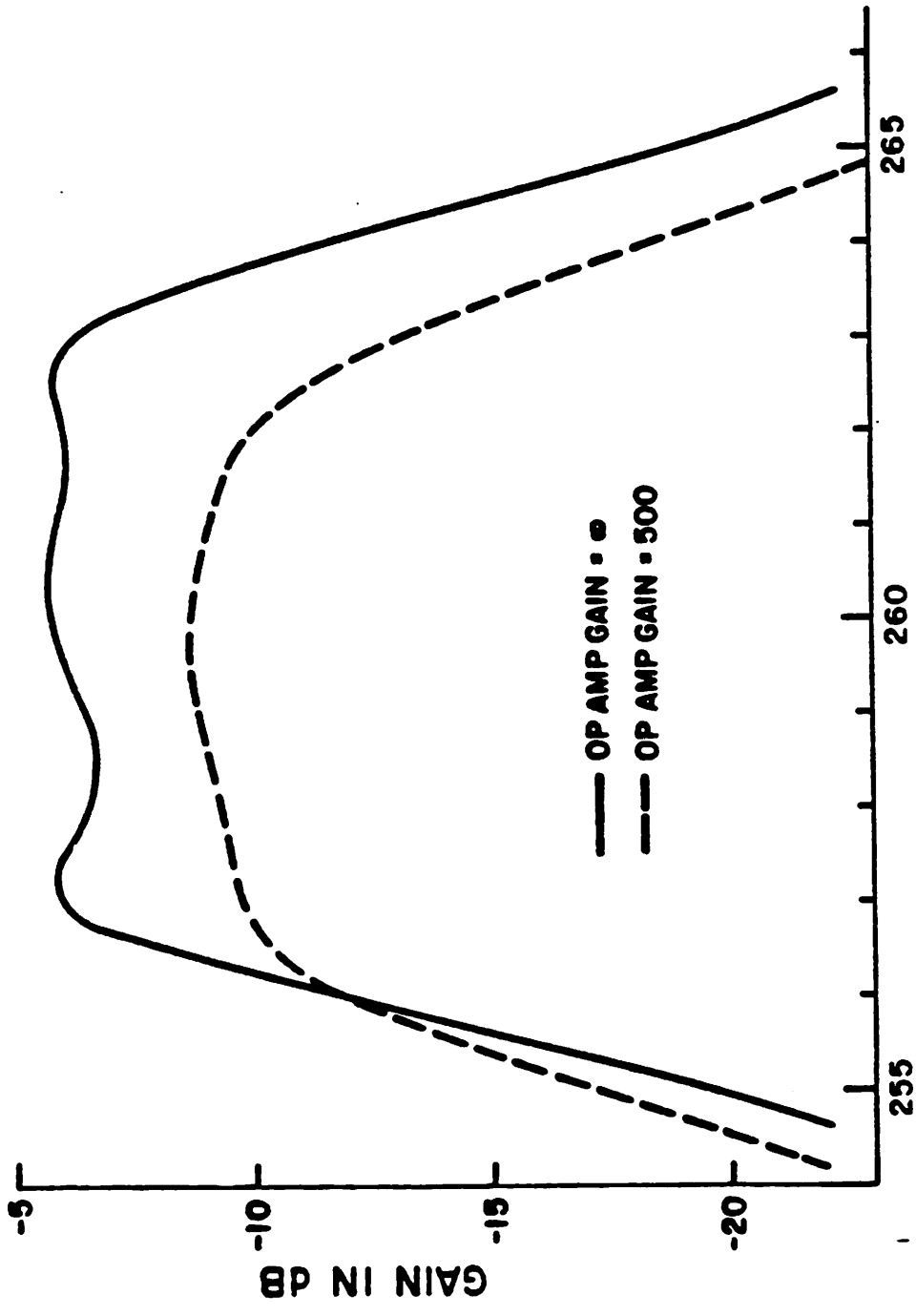


Fig. 3.30 Simulated detail passband response of AM IF filter with different op. amp. gains.

3.6 Design of 10.7 MHz FM IF Filter

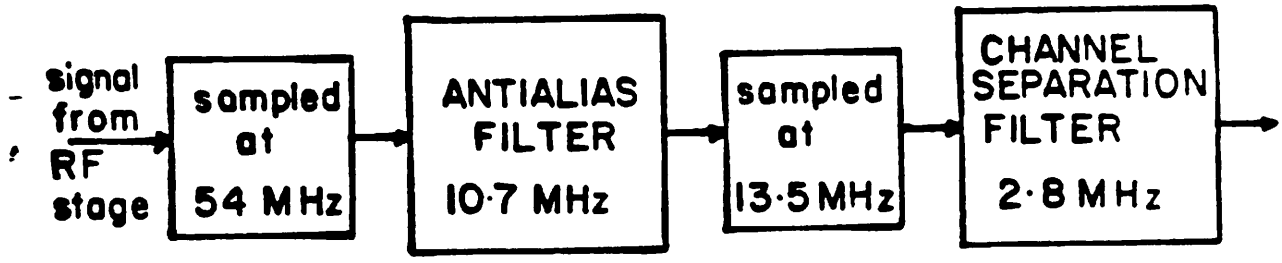
The FM IF filter is similar to the AM IF filter in many respects, except that it is at a much higher frequency (10.7 MHz), so that a completely different design approach is needed. Like the AM IF filter, it is used for channel separation in the intermediate frequency stage, and some typical specifications for such a filter is shown in Table 3.7. As discussed in Section 2.6, it is difficult to realize such a high Q filter at 10.7 MHz, so that a two stage filtering technique is employed to relax the filter requirements. This decimation technique has been described in Section 2.6, and the block diagram of the overall system for the 10.7 MHz FM filter is shown in Fig. 3.31a, with frequency response characteristics depicted in Fig. 3.31b. The choice of the 13.5 MHz sample rate for the channel separation filter is based on the following reason. If this sample rate is too high, the signal aliased back at $(f_s - 10.7\ MHz)$ will be too close to the signal at 10.7 MHz, thus imposing a tight roll-off requirement for the antialias filter. On the other hand, if this sample rate is too small, the aliased signal in the second principle range after f_s will be at $f_s + (f_s - 10.7\ MHz)$, which again will be close to the original 10.7 MHz signal, thus making it difficult for the antialias filter.

The filters which meet these requirements are sixth order all-pole filters. The choice of non-elliptical design is to preserve the phase characteristics to be as linear as possible in the passband. The antialias filter is a sixth order Butterworth filter, with a Q of about 7, whereas the channel separation filter is a sixth order Chebyshev filter with a Q of about 10. These filter requirements have been described earlier in Table 2.1.

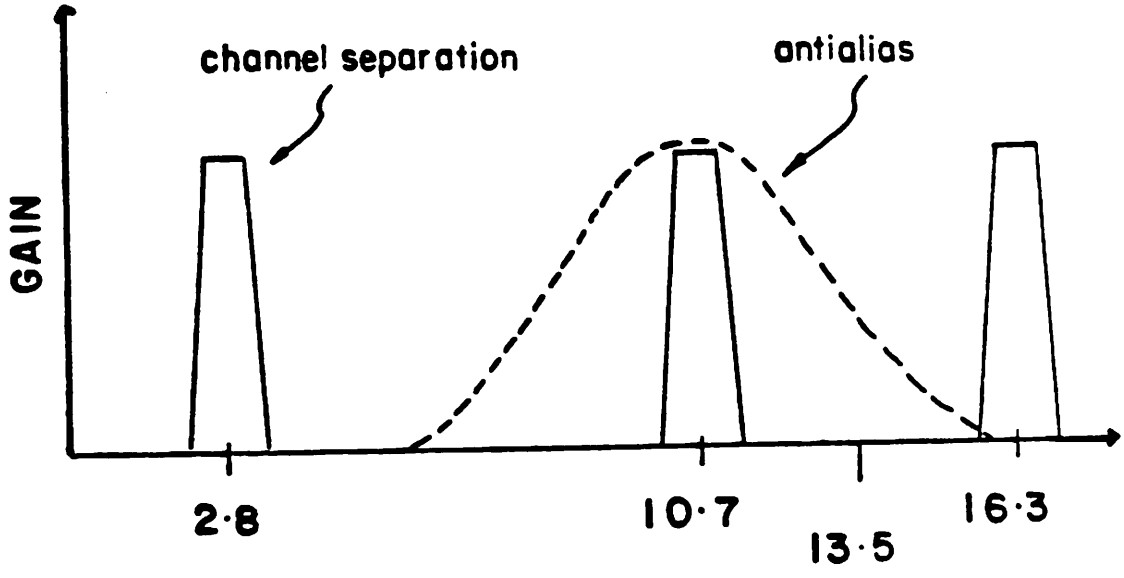
Like the AM case, the design of both the antialias and channel separation filters for the FM case starts with a normalized lowpass LC ladder shown in Fig. 3.32. Here, both filters are all-pole filters, so that their structures are identical but with different element values. The LC ladder after lowpass to bandpass

center frequency	10.7 MHz
ripple	3 dB
-3 dB bandwidth	230 KHz min
-20 dB bandwidth	600 KHz max

Table 3.7 Typical specifications for 10.7 MHz FM IF filter

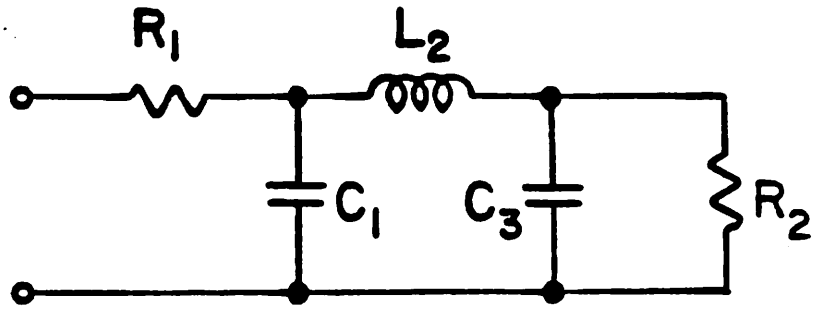


(a)



(b)

Fig. 3.31 (a) Decimation scheme for 10.7 MHz FM IF filter.
(b) Gain characteristics of decimation scheme.



Antialias

$R_1 = R_2 = 1$ ohm

$C_1 = C_3 = .999215$ F

$L_2 = 1.99843$ H

Channel Separation

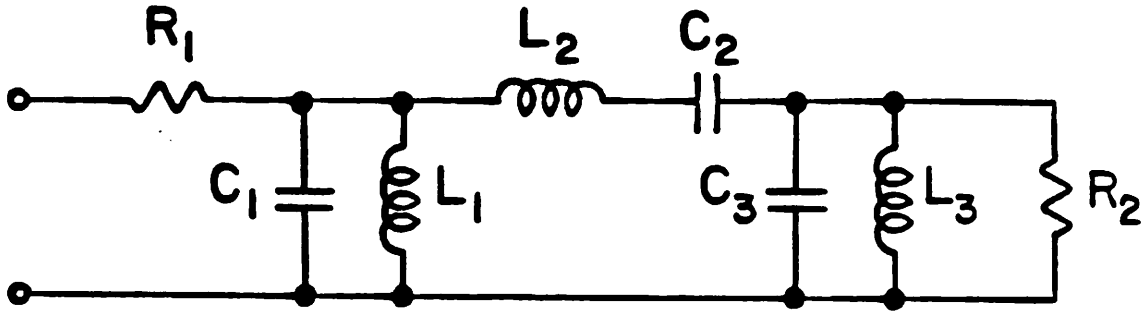
$R_1 = R_2 = 1$ ohm

$C_1 = C_3 = 1.59631$ F

$L_2 = 1.09667$ H

Fig. 3.32 **Normalized LC lowpass ladder for FM IF composite filter.**

transformation is shown in Fig. 3.33, with corresponding signal flow diagram shown in Fig. 3.34. The final fully-differential switched-capacitor circuit is shown in Fig. 3.35. Here, double-sampling integrators are used to provide a larger window for the op. amps. to settle. T-networks are not used since the coupling ratios are not large with relaxed filter requirements. In addition, they tend to complicate amplifier transient response. Also, complex-conjugate terminations are used and the simulated responses are shown in Figs. 3.36a-d. Also shown are the simulated results with different amplifier gains. A simple sensitivity analysis is also performed where we vary the integrator capacitor ratios by 0.5% and coupling capacitor ratios by 5%. The results are summarized in Table 3.8a and b. We notice that for both antialias and channel separation filters, this kind of capacitor ratio errors is still acceptable. An amplifier gain of 100 is adequate for the antialias case but marginal for the channel separation filter, where we would prefer a gain of 200. The results of the composite filter with the response of the channel separation filter translated back up to 10.7 MHz is summarized in Table 3.9, where we have excellent center frequency accuracy and very good overall characteristics, even with an amplifier gain of only 100.



Antialias

$R_1 = R_2 = 1$ ohm

$C_1 = C_3 = 0.114712\ \mu F$

$L_1 = L_3 = 2.19838$ nH

$C_2 = 1.71313\ \mu F$

$L_2 = 0.229425\ \mu F$

Channel Separation

$R_1 = R_2 = 1$ ohm

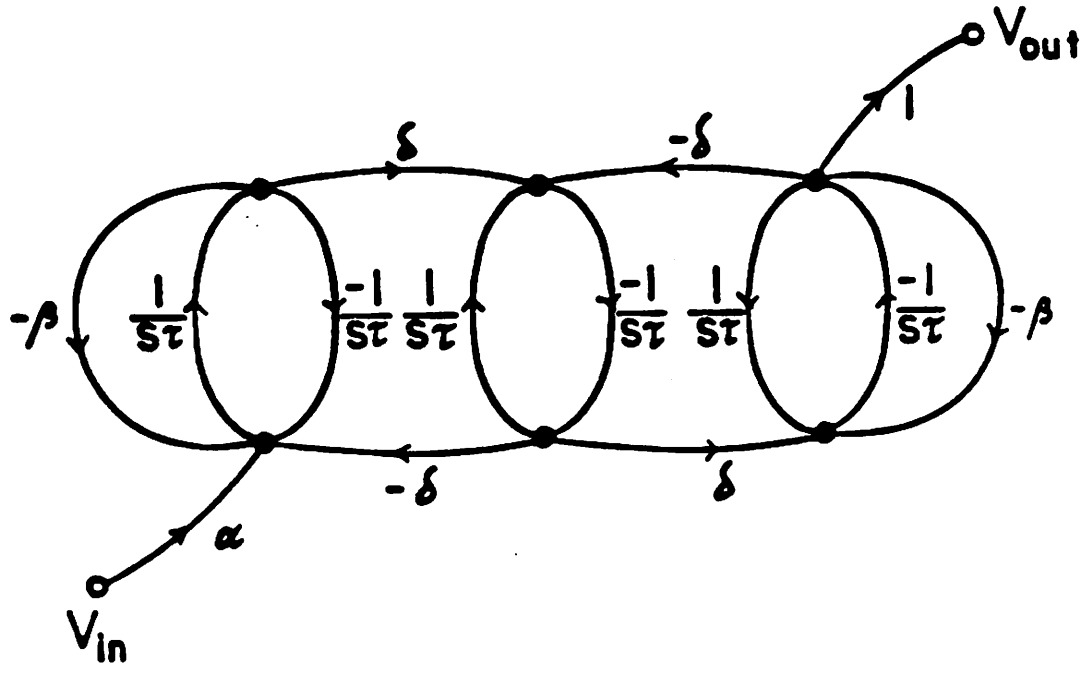
$C_1 = C_3 = 1.3894\ \mu F$

$L_1 = L_3 = 2.68967$ nH

$C_2 = 6.10182\ \mu F$

$L_2 = 0.954525\ \mu H$

Fig. 3.33 **Bandpass LC ladder for FM IF composite filter.**



Antialias

$\alpha = 0.26231$

$\beta = .16144$

$\delta = .11415$

$\tau = 1.1661$

Channel Separation

$\alpha = 0.08491$

$\beta = 0.05331$

$\delta = 0.06432$

$\tau = 1.2117$

Fig. 3.34 **Signal flow graph for FM IF composite filter.**

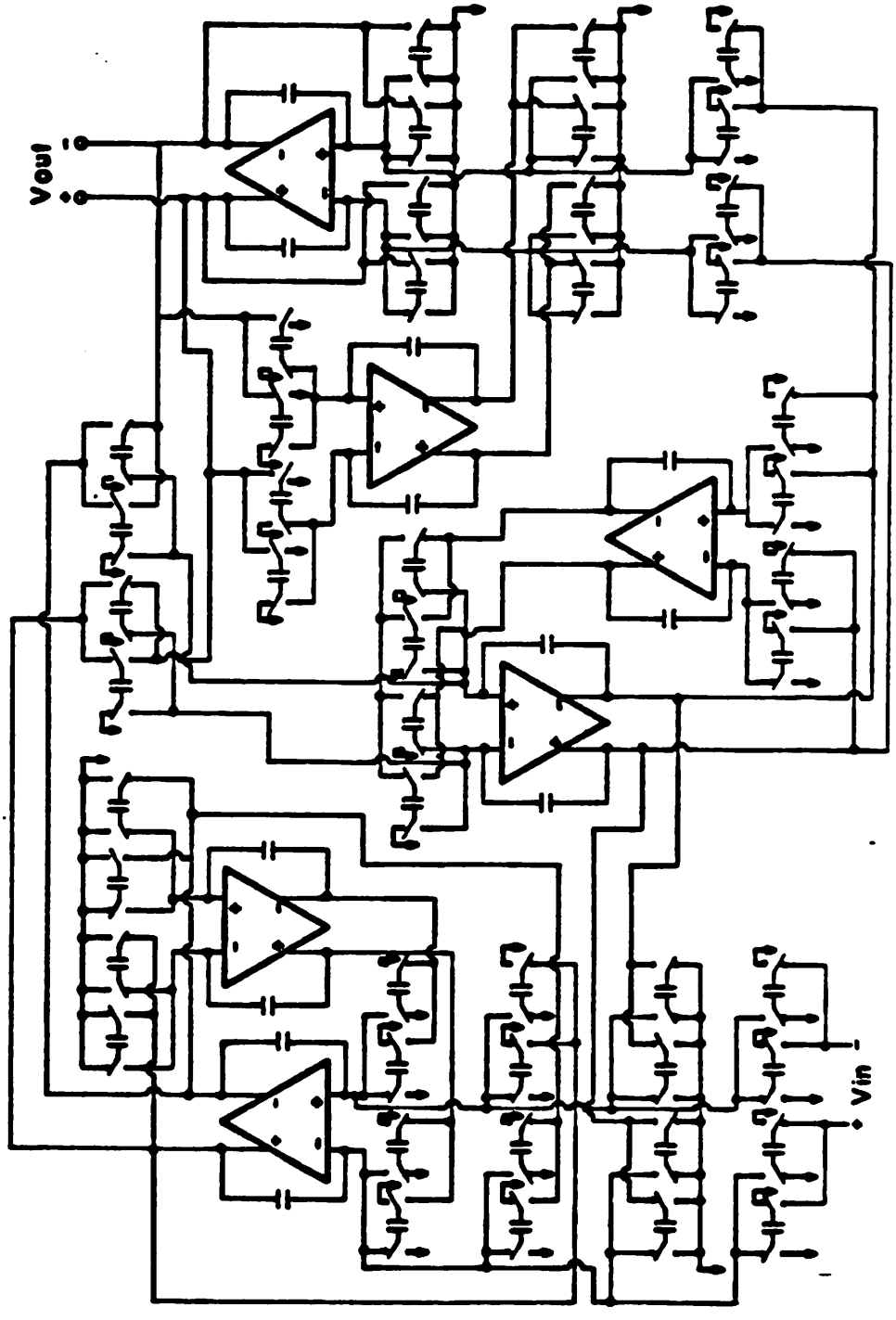


Fig. 3.35 Fully differential switched-capacitor filter for FM IF applications

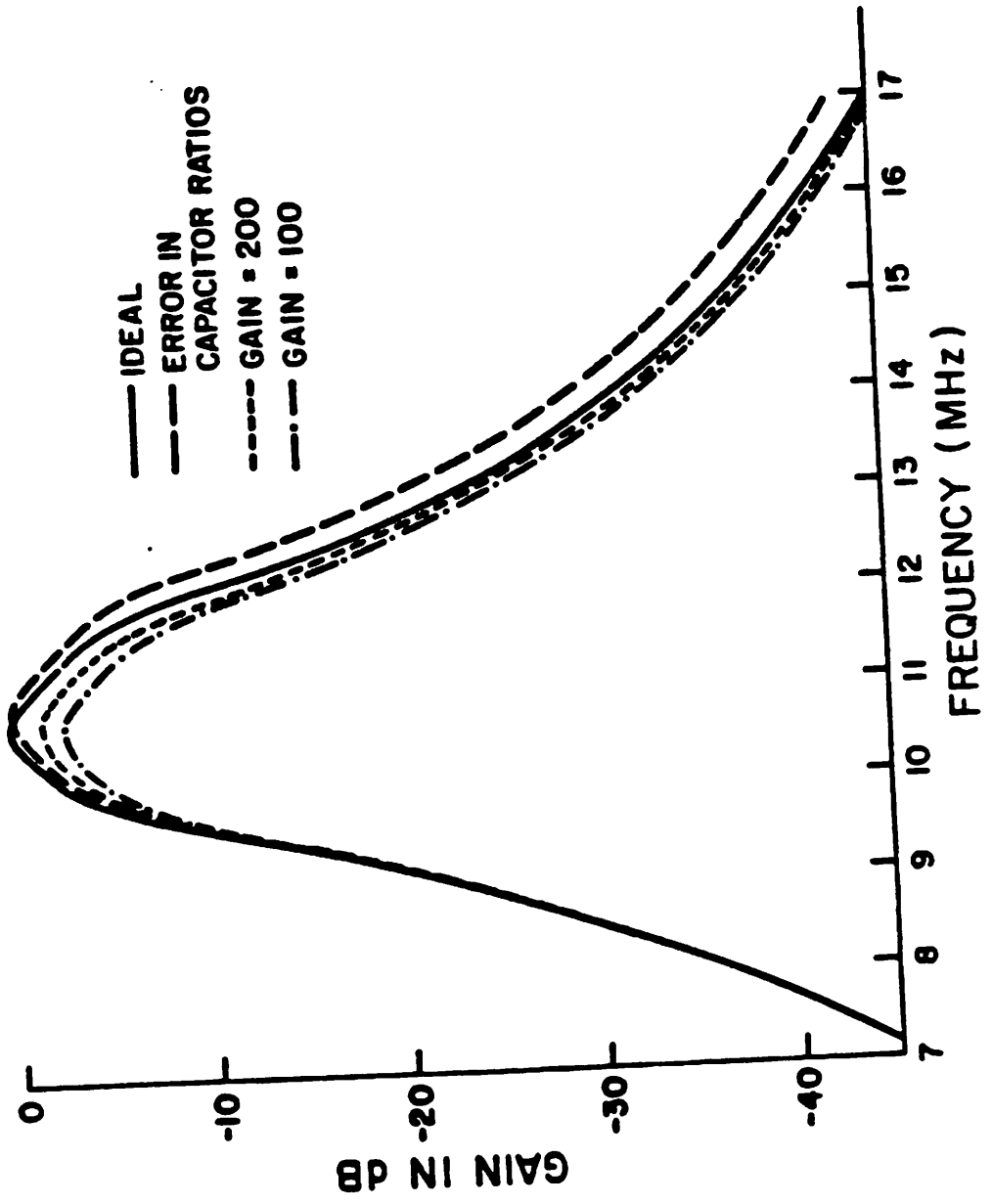


Fig. 3.38 (a) Simulated overall response of FM antialias filter.

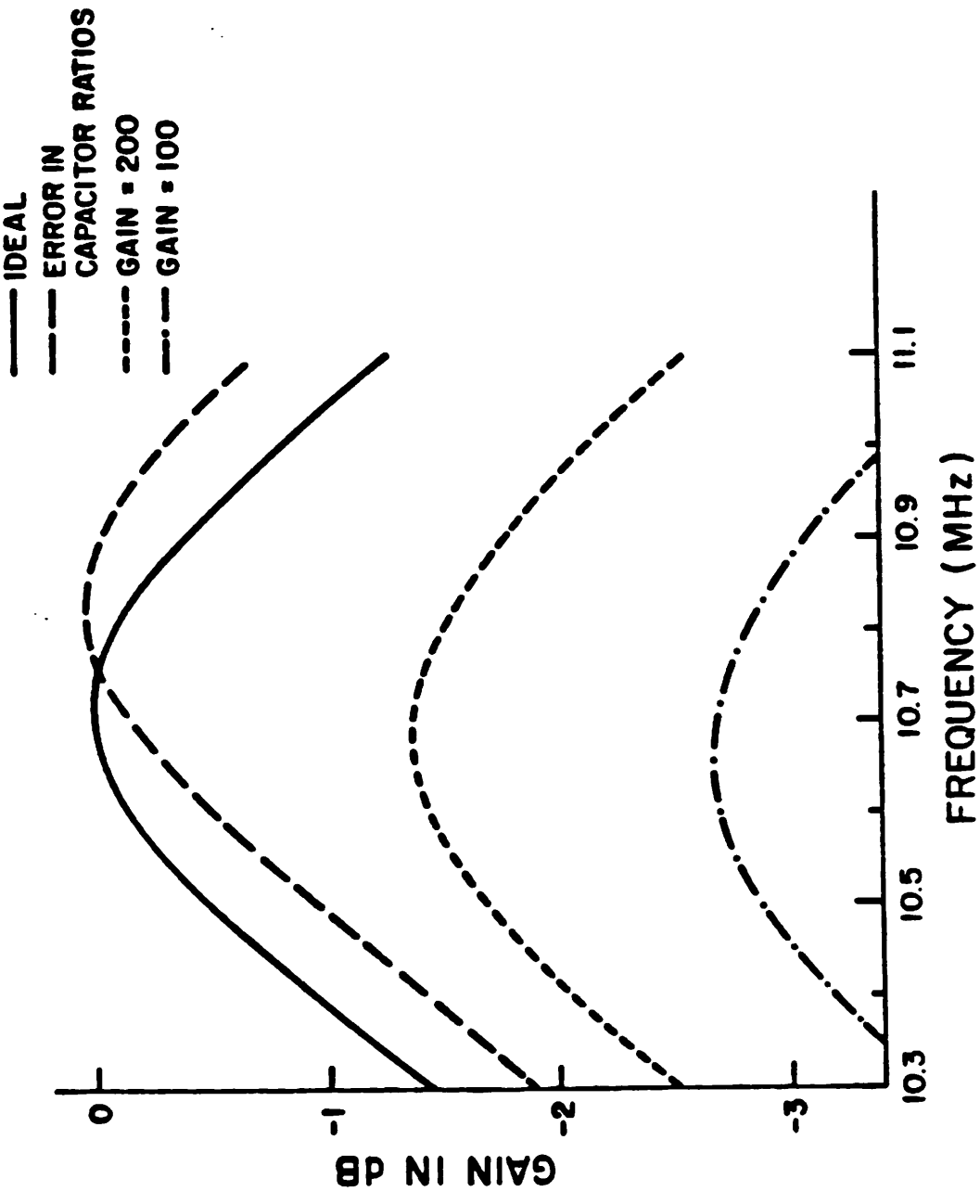


Fig. 3.36 (b) Simulated detailed passband response of FM antialias filter.

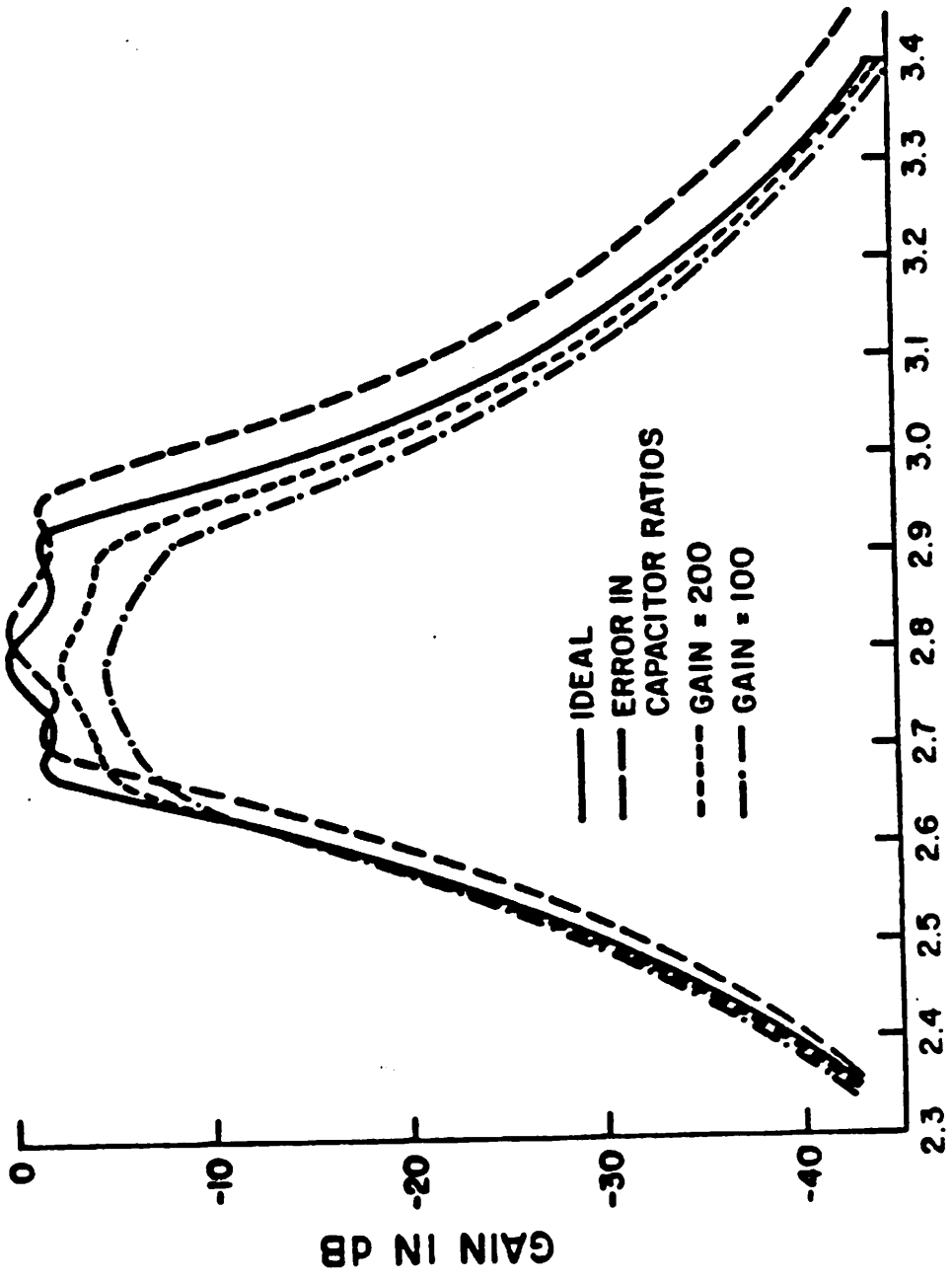


Fig. 3.38 (c) Simulated overall response of FM channel-separation filter.

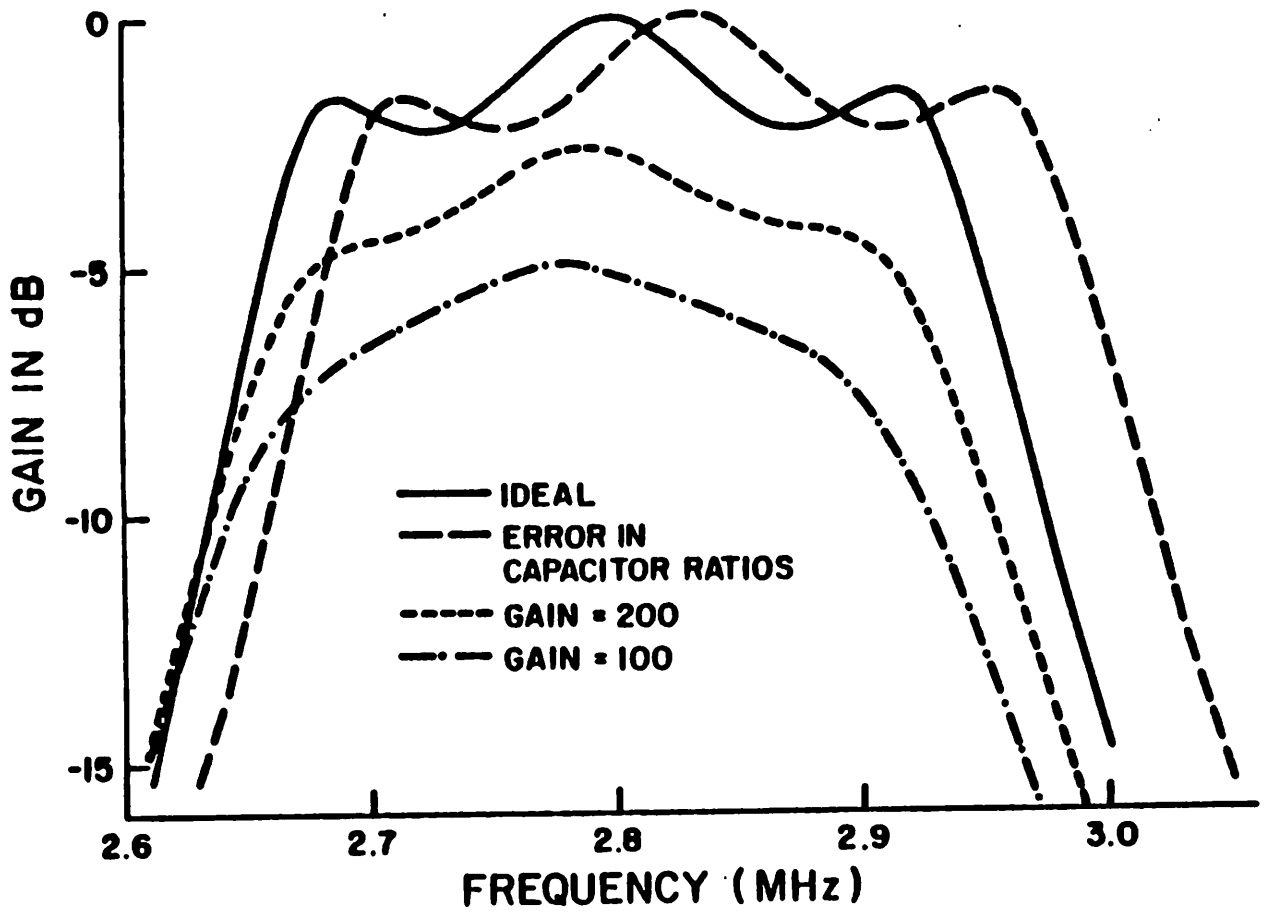


Fig. 3.36 (d) Simulated detailed passband response of FM channel-separation filter.

	nominal	error in capacitor ratios	finite op. amp. gain = 100	finte op. amp. gain = 200
center freq. (MHz)	10.72	10.84	10.66	10.68
passband variation about 10.7 MHz (dB)	<0.3	<0.7	<0.3	<0.2
rejection at 16.1 MHz (dB)	-42	-40	-40	-41

Table 3.8a Simulated response of FM antialias filter

	nominal	error in capacitor ratios	finite op. amp. gain = 100	finte op. amp. gain = 200
center freq. (MHz)	2.8	2.83	2.78	2.79
ripple (dB)	2.25	2.3		
-2.3 dB bandwidth (KHz)	256	270	210	230
-20 dB bandwidth (KHz)	460	485	515	490
-40 dB bandwidth (KHz)	920	1040	1090	1000

Table 3.8b Simulated response of FM channel-separation filter

	nominal	error in capacitor ratios	finite op. amp. gain = 100	finite op. amp. gain = 200
center freq. (MHz)	10.7	10.67	10.72	10.71
ripple (dB)	2.55	3.0		
-3 dB bandwidth (KHz)	260	270	230	250
-20 dB bandwidth (KHz)	450	470	510	475
-40 dB bandwidth (KHz)	870	910	1030	940

Table 3.9 Simulated composite filter response of FM IF filter (center freq. of actual filter translated to 10.7 MHz)

Chapter 4

CMOS Technology in High Speed Analog Applications

4.1 Comparison of NMOS and CMOS Technologies in High Speed Analog Applications

Due to its lower power consumption, CMOS technology is gaining widespread acceptance in digital integrated circuits. Since most analog IC's act as interface circuits between the analog world and the digital processors, an increasing number of analog IC's are designed in CMOS. There are certain advantages in designing analog circuits in CMOS [34], and this section will compare CMOS and NMOS technologies in high speed analog applications.

Since we are mainly concerned with high speed circuits, the kind of MOS technology that will be employed in the near future will be a multi-level silicon gate technology with very thin gate dielectrics (~ 200 Å), very short channel lengths (down to 1 μm electrical), shallow junctions (~ 0.25 μm), and tight alignment tolerances (down to $\frac{1}{4}\mu m$). With this kind of technology, the speed of the circuit will be dominated by junction and interconnect capacitances, so that the circuit performance is extremely layout dependent, and should be analyzed accordingly.

There are roughly two categories of analog IC's. One is A/D, D/A converters which, in high speed applications, will be typically using latch-comparators in flash or pipeline configurations. The speed of this type of circuits will be limited by the time required by the comparators to latch. In very high speed applications, we will be talking about clocking rates on the order of 1 GHz and in these cases, the precision required are usually not high, say 4 bits. The other type is switched-capacitor circuits which require high gain operational amplifiers. The

speed of such circuits will be determined by the amplifier settling time. In general, NMOS technology will be faster where little or no gain is required whereas CMOS technology will out-perform NMOS where high gain amplifiers are needed. Hence, NMOS technology is usually used in very high speed latch-comparators for A/D, D/A applications where low precision is acceptable. On the other hand, CMOS technology is employed in high frequency, highly selective switched-capacitor filters where accuracy is required.

To see why this is true, we have to analyze the layout of an inverter in both technologies. Fig. 4.1a shows a typical NMOS inverter with depletion load. The transient response of this inverter is determined by the capacitance at the output node which is dominated by the diffusion capacitance to substrate. A typical layout for such an inverter is shown in Fig. 4.1b. Here, for simplicity, we are assuming that both the driver and the load have the same channel width, which can be done with proper tailoring of the threshold voltages. Using 1.5 μm drawn design rules, the diffusing length at the output node is 3 μm. The disadvantage with this layout is that the circuit is sensitive to misalignment between the polysilicon and the buried contact since the channel length of the depletion load device can be altered by the misalignment. A different layout which is insensitive to misalignment error is shown in Fig. 4.1c. With the same design rules, the diffusion length at the output node is now 4.5 μm. In CMOS technology, the layout for an inverter is much more cumbersome. Fig. 4.2a shows a CMOS inverter circuit, with corresponding layout shown in Fig. 4.2b. The disadvantage with CMOS inverters is that the diffusions of the load and driver transistors are of different types so that they cannot share a common area. In addition, they have to be linked together by an interconnect polysilicon or metal which runs over a long distance between the n-diffusion in the substrate and the p-diffusion in the well. Even if we neglect this capacitance from the interconnect to the substrate, the diffusion capacitance alone at the output node is 6 μm long using the same

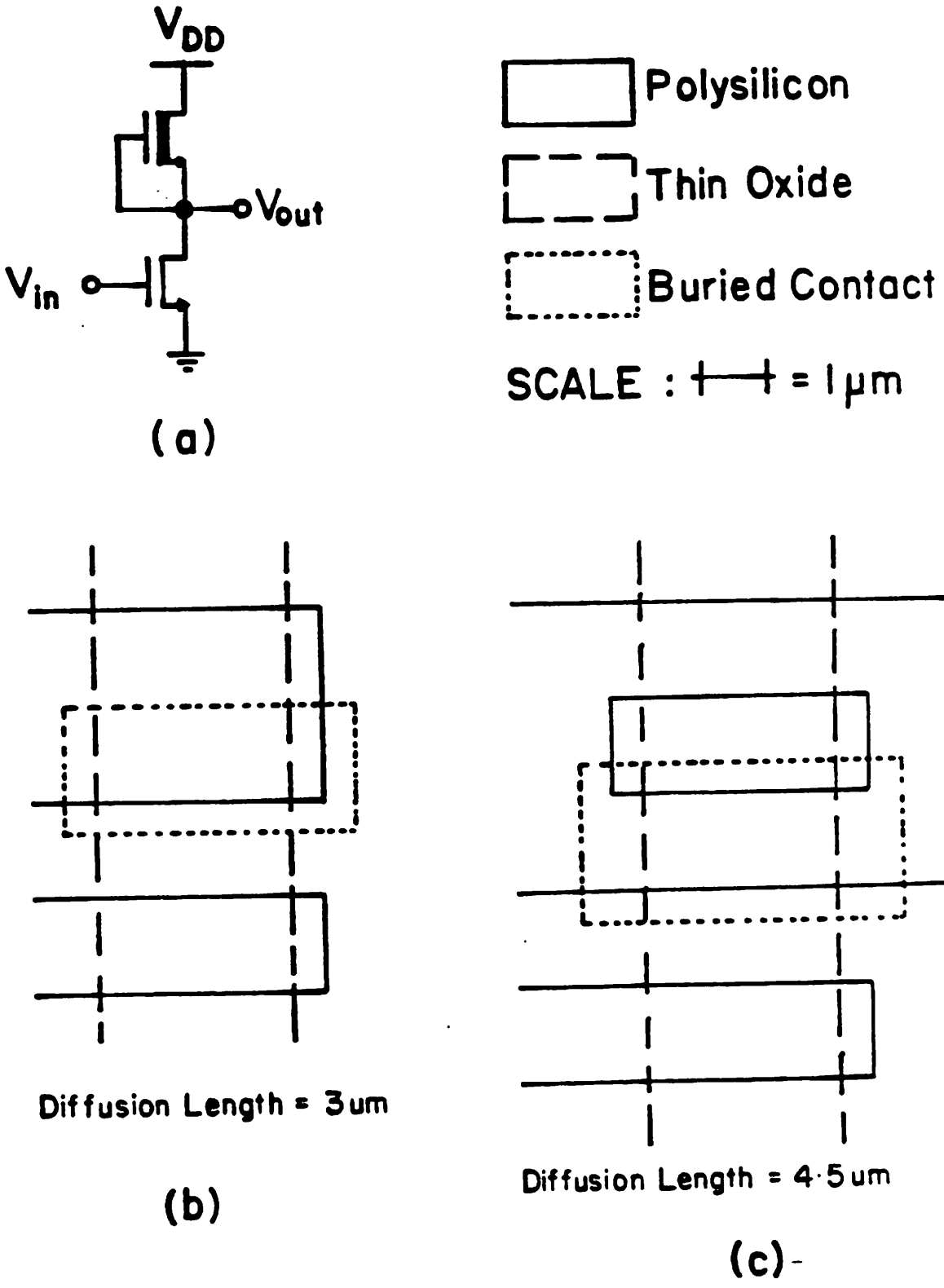


Fig. 4.1 (a) NMOS inverter with depletion load.
(b) Typical NMOS inverter layout.
(c) NMOS inverter layout which is insensitive to alignment error.

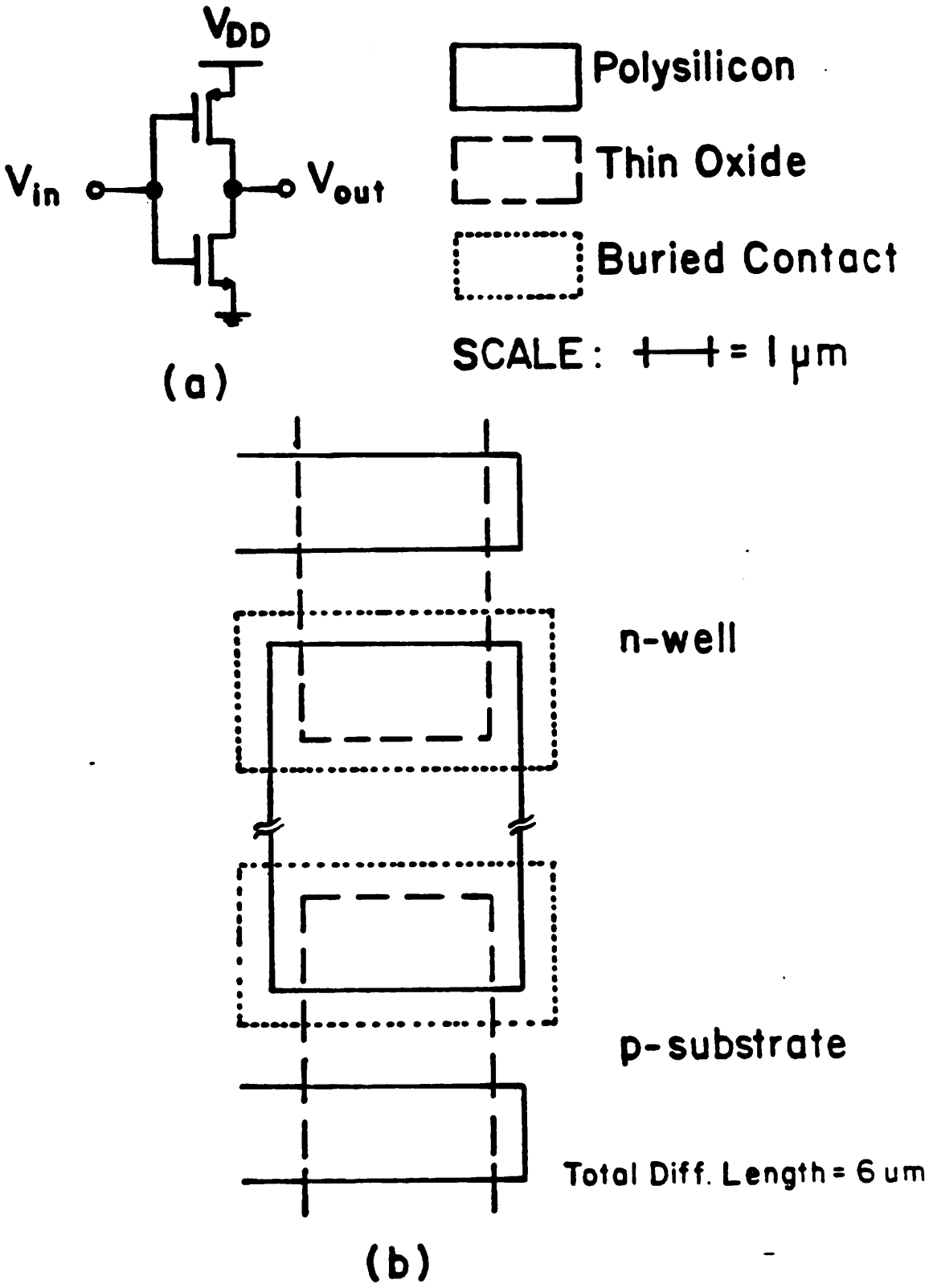


Fig. 4.2 (a) CMOS inverter circuit.
(b) CMOS inverter layout.

design rules as in the NMOS case. Also, there are additional edge capacitances at the ends of the diffusions which is equal to twice the channel width. Hence the capacitance at the output node of a CMOS inverter is roughly twice that of an NMOS inverter, so that the response of the CMOS inverter is slower. Of course, the CMOS inverter has a lot more gain than the NMOS case, but this may not be needed in low precision applications where speed is the primary issue.

On the other hand, if we consider the design of a high gain operational amplifier, many other considerations come into the picture which strongly favor the choice of a CMOS technology. For the NMOS inverter shown in Fig. 4.1a, the small signal gain is given by

$$\frac{V_{out}}{V_{in}} = g_m r_{eq} \tag{4.1}$$

where g_m is the transconductance of the driver and r_{eq} is the equivalent resistance at the output node. For transistors with 3 to 4 μm channel lengths, r_{eq} is dominated by the body effect associated with the depletion load transistor. However, for the CMOS inverter of Fig. 4.2a, there is no body effect since the source of the PMOS load is always tied to the n-well and its threshold voltage will not change. Hence CMOS inverters can have significantly higher gain than the NMOS case.

With channel lengths down to the 1 μm range, the channel length modulation effects will become important, so that the gain advantage of simple CMOS inverters is somewhat reduced. However, the ability for CMOS circuits to realize a much better current source than NMOS circuits makes high gain CMOS amplifiers much faster than the NMOS case. To understand why this is true, let us first consider the inherent gain-speed tradeoff of a MOS transistor. The first order MOS current equation in the saturation region is given by

$$I_D = \frac{\mu C_{ox}}{2}\frac{W}{L}(V_{GS} - V_T)^2 \tag{4.2}$$

where I_D is the drain current, V_{GS} is the gate to source voltage, V_T is the threshold voltage, μ is the electron or hole mobility, C_{ox} is the oxide capacitance, and W and L are the channel width and length respectively. The transconductance g_m is given by

$$g_m = \frac{\partial I_D}{\partial V_{GS}} = \mu C_{ox}\frac{W}{L}(V_{GS} - V_T)\,. \tag{4.3}$$

If we include the channel length modulation effects, the current equation can be rewritten as

$$I_D = \frac{\mu C_{ox}}{2}\frac{W}{L}(V_{GS} - V_T)^2(1 + \lambda V_{DS}) \tag{4.4}$$

where λ is the channel length modulation parameter and V_{DS} is the drain to source voltage. The small signal output resistance r_{out} is given by

$$\frac{1}{r_{out}} = \frac{\partial I_D}{\partial V_{DS}} = \lambda\frac{\mu C_{ox}}{2}\frac{W}{L}(V_{GS} - V_T)^2 \tag{4.5}$$

so that the $g_m r_{out}$ product is given by

$$g_m r_{out} = \frac{2}{\lambda(V_{GS} - V_T)}\,. \tag{4.6}$$

Notice that equation (4.6) is derived from different equations, (4.2) and (4.4). In addition, these current equations are derived from first order MOS models so that equation (4.6) is not entirely accurate. However, it points out the important parameters which determine the $g_m r_{out}$ product of a transistor, and they are λ and $(V_{GS} - V_T)$. The significance of the $g_m r_{out}$ product is that it gives an upper bound of the small signal gain of the simple inverter circuits of Figs. 4.1a and 4.2a. The actual gain is given by equation (4.1) where r_{eq} is the parallel output resistances of the driver and the load, so that the actual gain of an inverter is typically half the $g_m r_{out}$ product of the driver. Notice that the $g_m r_{out}$ product

is basically a function of λ and $(V_{GS} - V_T)$. For a given channel length, λ is constant, so that $g_m r_{out}$ is determined by $(V_{GS} - V_T)$. From equation (4.6), we see that when $(V_{GS} - V_T)$ gets smaller, the $g_m r_{out}$ product increases and the gain is therefore larger. However, the current gets smaller with $(V_{GS} - V_T)$ so that the speed is slower. In other words, we can always trade gain for speed. For optimal performance, we typically choose $(V_{GS} - V_T)$ at around 1.5 V where the $g_m r_{out}$ product is roughly 15 for transistors with 1 μm channel lengths. This is a small gain and to increase this number, we can either increase the channel length or use a cascode type of circuit. Increasing the channel length can decrease λ and therefore increase the $g_m r_{out}$ product. However, this is not a good strategy because the larger the channel length, the bigger the capacitances, and this makes the transistor a lot slower. Hence, we typically use a cascode type of circuitry as shown in Fig. 4.3. Assuming we have an ideal current source in the load, the gain of this circuit is roughly given by

$$\frac{V_{out}}{V_{in}} = g_{m1} r_{out1} \, g_{m2} r_{out2} \, . \tag{4.7}$$

Hence, the gain is increased by another factor of $g_m r_{out}$ using this technique. The capacitance associated with the output node is roughly the same as that of the simple inverter circuit, but the resistance associated with this node is also increase by a factor of $g_m r_{out}$. Since both the gain and the output resistance increase by the same factor while the capacitance stays the same, the unity gain frequency of this cascode circuit will be approximately equal to that of the simple inverter. Using this cascode technique, we can increase the D.C. gain of the circuit without sacrificing the bandwidth. However, we have been assuming that we have an ideal current source in the load. In CMOS technology, we can realize this by cascoding the load device as shown in Fig. 4.4, so that the output resistance of the load cascode is on the same order as the driver cascode. In the NMOS case, no such technique is available. For the load device, the output is

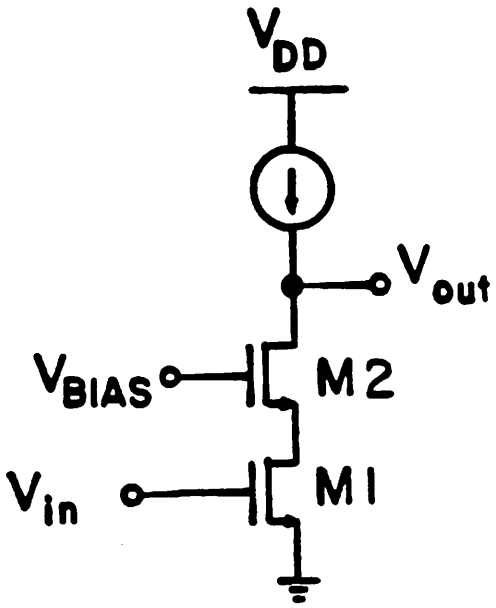


Fig. 4.3 Typical cascode circuit.

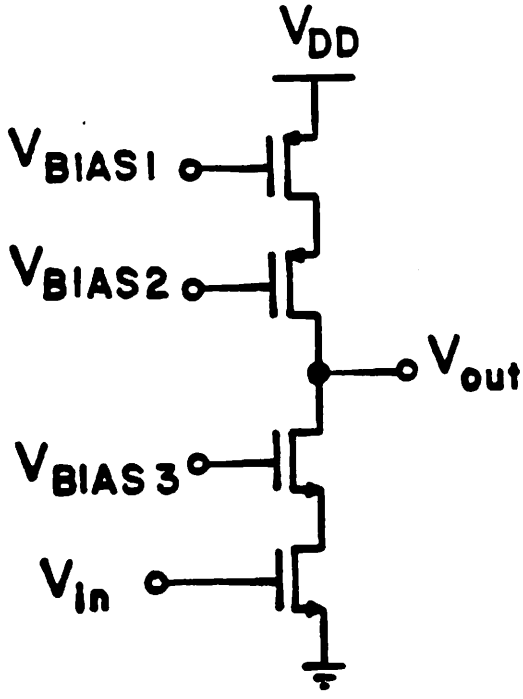


Fig. 4.4 CMOS cascode circuit.

always tied to the source of NMOS transistor and cascode circuitry cannot be used. In this case, the gain is still dominated by the load output resistance, which remains the same. To increase the gain for NMOS circuits, we have to cascade two inverters in series, as shown in Fig. 4.5. The gain for this circuit will be

$$\frac{V_{out}}{V_{in}} = g_{m1} r_{eq1} g_{m2} r_{eq2} . \tag{4.8}$$

The disadvantage with this two-stage amplifier is that we have two dominant poles in the circuit and a compensation capacitance C_c is required. This compensation capacitance can greatly reduce the bandwidth of the amplifier so that the NMOS design in Fig. 4.5 is significantly inferior to that of the CMOS design in Fig. 4.4. The cascode technique in the CMOS case can be further extended to triple cascode if more gain is required. In the extreme case where a two-stage CMOS amplifier is needed, the compensation for the CMOS amplifier can be easier than the NMOS case. This is because we can put most of the gain in one stage so that the dominant pole is far away from the second pole before compensation. This is shown in Fig. 4.6. The dominant pole of this circuit is at the output node of the first stage since the output resistance is extremely high. The second pole at the output node of the second stage will be far away from the dominant pole since the output resistance is much smaller, even though this second stage may have to drive a larger load capacitance. Since these two poles are far apart to begin with, the pole-splitting compensation capacitance can be very small. In the NMOS case where no cascode technique is available, the two poles of the two-stage amplifier in Fig. 4.5 cannot be made far apart before compensation, so that C_c has to be large. In addition, in order to achieve the same gain as the CMOS circuit in Fig. 4.6 using NMOS techniques, we have to use four inverters in series, and it will be a nightmare to compensate a four-stage amplifier will all four poles close together.

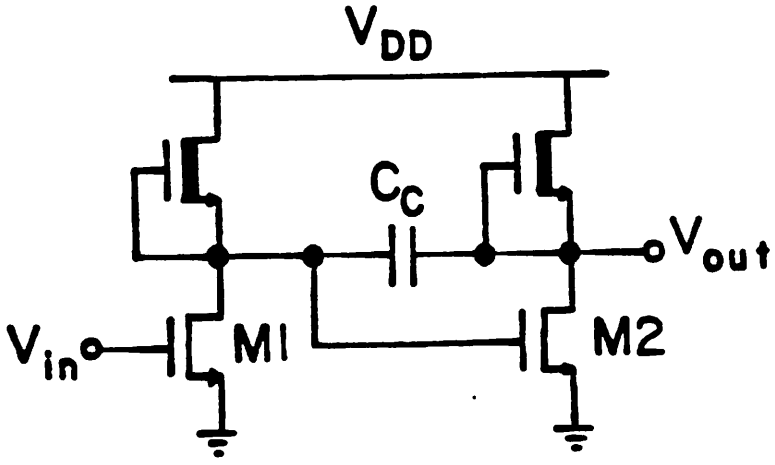


Fig. 4.5 NMOS cascade circuit.

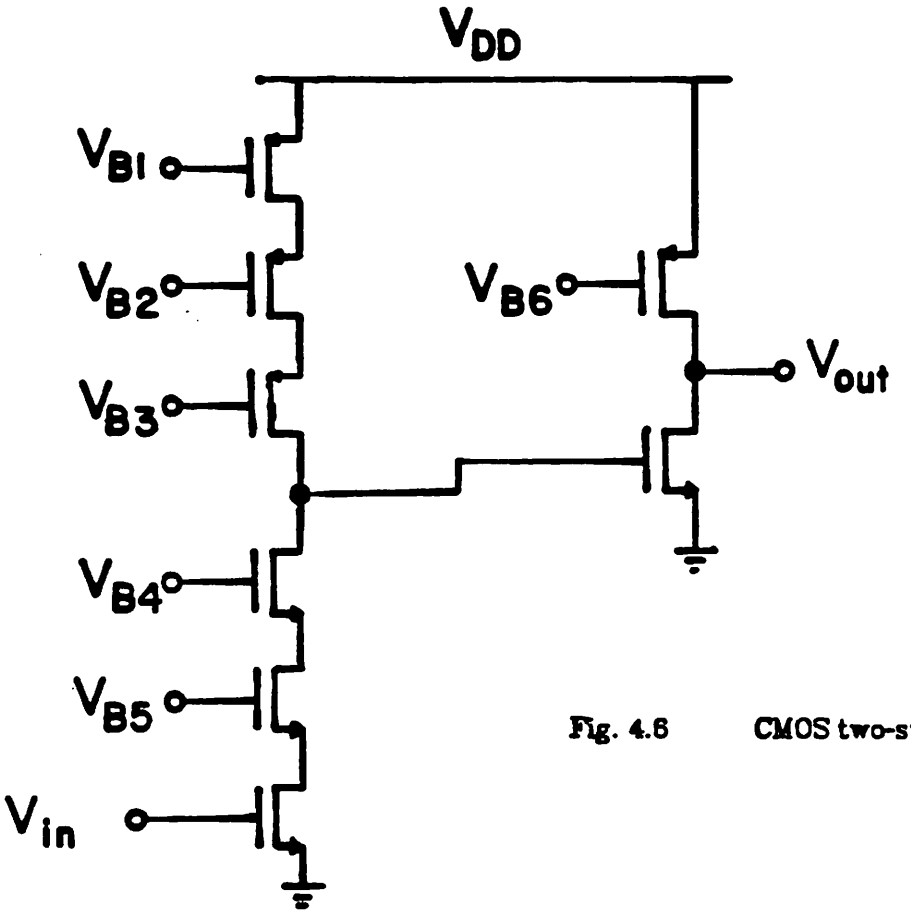


Fig. 4.6 CMOS two-stage amplifier.

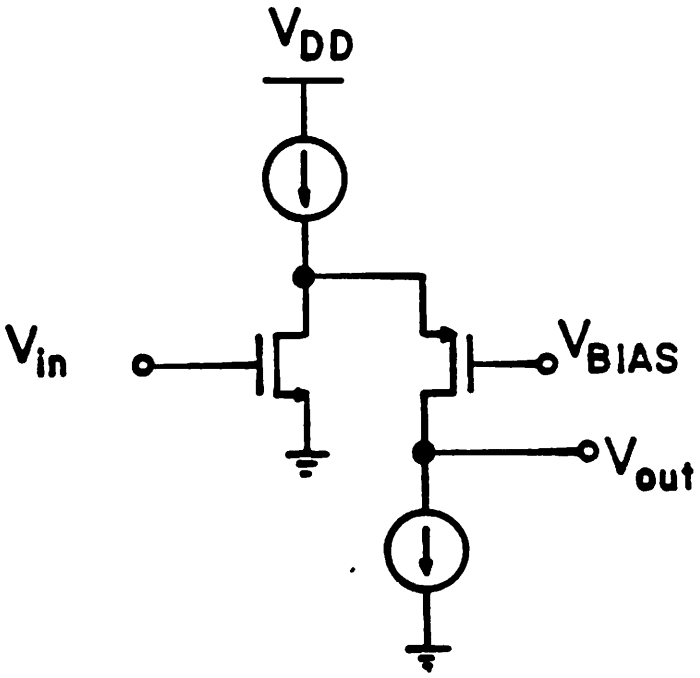


Fig. 4.7 CMOS folded-cascode.

There are other advantages in designing with CMOS. Because both p-type and n-type transistors are available, we can achieve voltage level shift with gain in both direction. This is shown in Fig. 4.7. This type of circuitry can be very convenient in single-stage designs because the output and the input D.C. voltage levels are the same. The detail design tradeoffs for such a folded cascode circuit is discussed in [26]. Another advantage with CMOS technology is that it is compatible with low power CMOS digital circuitry. As more and more digital IC's are designed in high speed CMOS, it will be advantageous to use the same technology for analog designs so that a higher level of integration is possible when we need the digital processor and the analog interface on the same chip.

The main drawback with CMOS technology is that the die area required is larger than NMOS designs. If we look at the CMOS inverter layout shown in Fig. 4.2b, we notice that a long separation is required between the n-diffusion in the substrate and the p-diffusion in the well. For present day CMOS technologies, this is the density limitation of CMOS IC's. Though we can shrink the transistor feature size down to the 1 μm level, this n-diffusion to p-diffusion distance cannot be substantially reduced due to latch-up considerations. Even with epitaxial CMOS technologies, the minimum distance between the two diffusions has to be roughly 8 μm to prevent latch-up. However, development in isolation technologies can drastically change this consideration. Presently, partial dielectric isolation with trench or groove techniques are gaining acceptance [37]. With oxide trench isolation, the n-diffusion to p-diffusion distance can be as small as the minimum feature size. In this case, the CMOS inverter still occupies a larger area because the NMOS inverter shares a common diffusion between the driver and the load. However, this small area advantage will no longer be a consideration in the choice between a CMOS or NMOS technology.

4.2 CMOS Technology Design Considerations

4.2.1 SHORT CHANNEL DEVICE CONSIDERATIONS

The most important consideration in a technology design is the device characteristics. Here we are interested in analog applications with 10 V power supply, so that we are particularly concerned with the breakdown phenomenon in the short channel transistors. Breakdown mechanisms in long channel MOS transistors are fairly well known [30]. The most dominant ones are gate-induced junction breakdown and channel avalanche breakdown. However, in short channel devices, subsurface as well as surface punchthrough can become important. We will first briefly discuss this two breakdown mechanisms, then we will analyze them in more detail with two-dimensional simulations and some experimental device characteristics.

In short channel devices, the drain depletion region can reach through to the source end and lower the potential barrier of the source junction (Fig. 4.8). With a sufficiently high voltage, the barrier can be lowered to a point where unwanted current will flow [31]. This subsurface punchthrough phenomenon is sometimes referred to as bulk drain-induced-barrier-lowering (bulk DIBL), and current flowing in this depletion region is diffusion limited. Subsurface punchthrough is only weakly affected by the gate voltage, although shallower junctions help.

Another breakdown mechanism in short channel devices is surface punchthrough, also known as the second gate effect (Fig. 4.9). In this case, field lines from the drain terminate on the gate. This field raises the surface potential towards $2\varphi_F$ against the influence of the grounded gate electrode. If the drain voltage is raised sufficiently high, the device enters weak inversion and unwanted current flows in this induced channel [32]. This phenomenon is sometimes referred to as surface drain-induced-barrier-lowering (surface DIBL), as

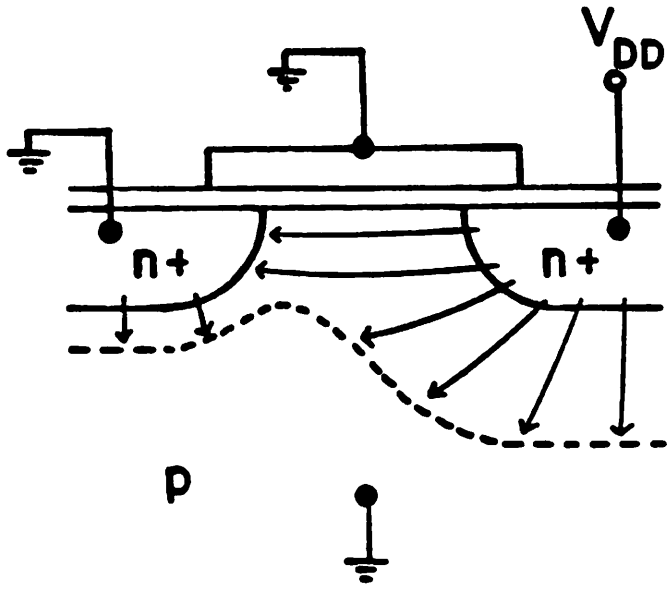


Fig. 4.8 Subsurface punchthrough mechanism.

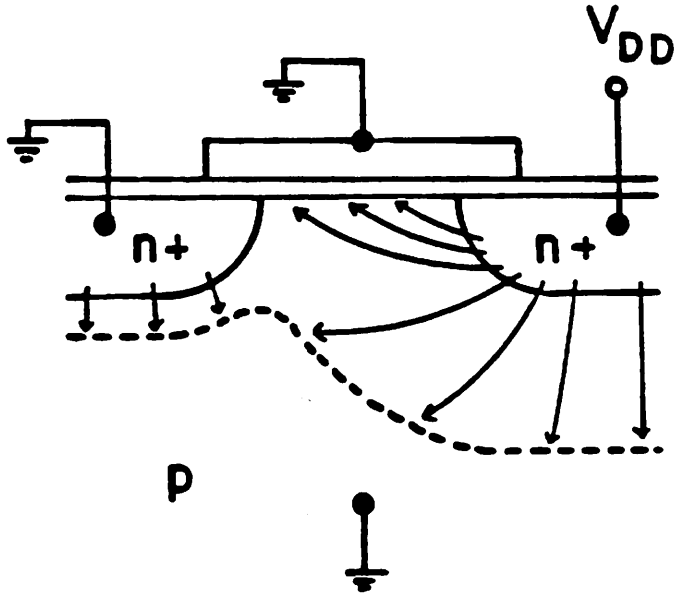


Fig. 4.9 Surface punchthrough mechanism.

opposed to the bulk DIBL in the previous case. As before, the current through the channel is diffusion limited, as in the case for subthreshold currents.

The mechanism behind these two breakdown phenomenon is the same, namely drain-induced-barrier-lowering. As a qualitative discussion of this mechanism, let us consider the structure shown in Fig. 4.10a, where the gate, source, and substrate are tied to zero volts. As the drain voltage is raised high enough (Fig. 4.10b), the potential barrier at the source junction is pulled low by the drain voltage. Even under this condition, when the two depletion regions already touch and the potential barrier at the source end is lower than the built-in voltage, the device may still not be affected since the current flow under this condition is so small that it can be totally neglected. Current flow under this condition is diffusion limited since the amount of electrons is so small in the depletion region. To calculate the amount of current flow, we need to know the potential profile and the carrier distribution of the device. This can be obtained from two-dimensional device simulation programs which calculate the solution to the two-dimensional Poisson's equation. As an example, we simulated the device in Fig. 4.11 using the two-dimensional device simulation program TWIST [38].

The device is an n-channel MOSFET with a substrate doping N_D of $5 \times 10^{14}\ cm^{-3}$, oxide thickness t_{ox} equal to 500 Å, and effective channel length L_{EFF} equal to 1 μm. The source and drain diffusions are formed by ion implantation, with a dose of $1 \times 10^{15}\ cm^{-2}$ and final junction depth x_j at 0.32 μm. There is also a shallow enhancement implant, with a dose of $5 \times 10^{11}\ cm^{-2}$ and penetration to 0.2 μm under the surface. The gate, source and bulk voltages, V_G, V_S, and V_B, respectively, are clamped to 0 V whereas the drain voltage V_D is at 3 V. The resulting potential contour map is shown in Fig. 4.12.

This is a typical potential profile from 2-D simulation programs in the punchthrough mode. The numbers on the contours correspond to the potential with

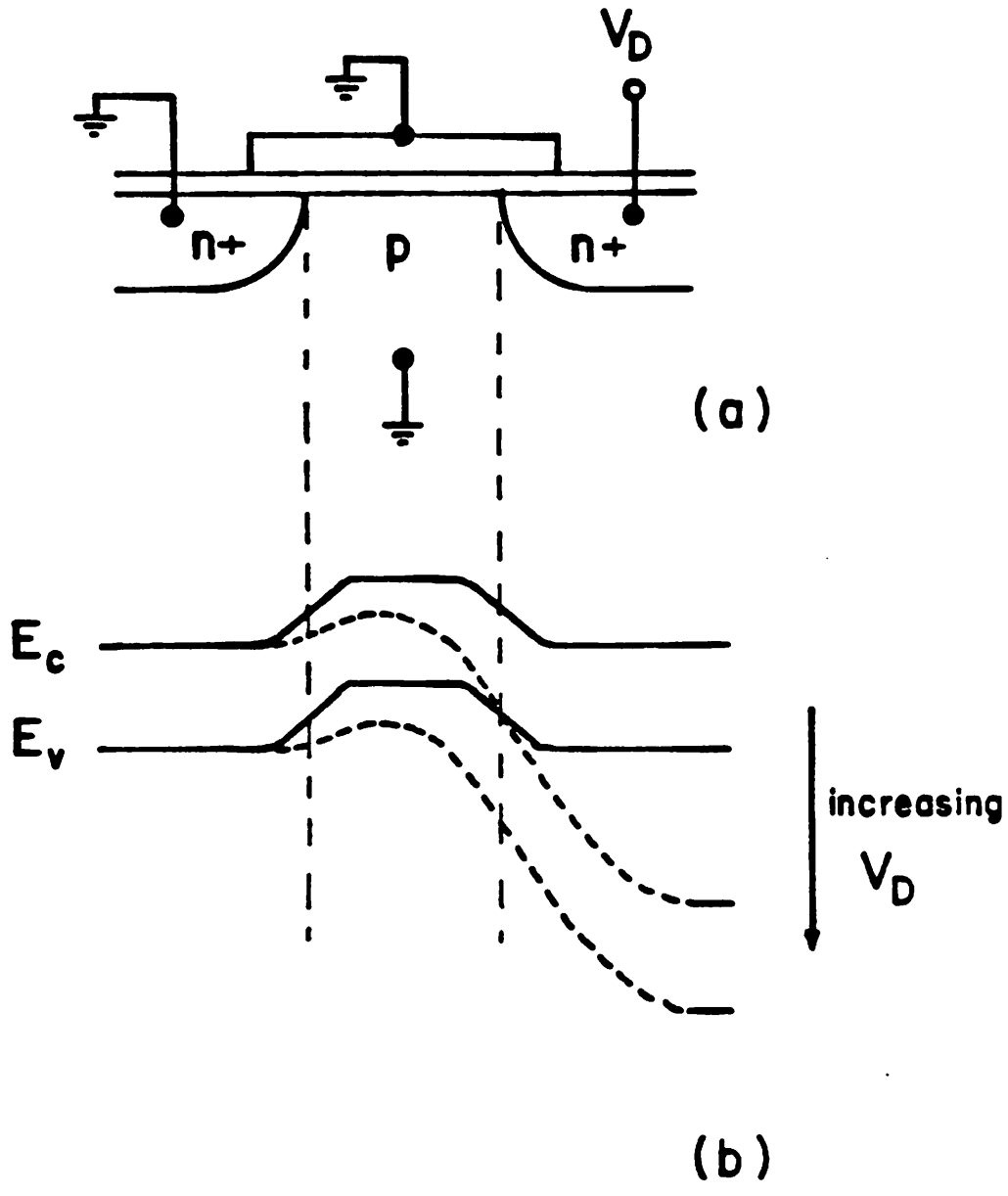


Fig. 4.10 (a) MOS device structure.
(b) Energy diagram showing drain-induced-barrier-lowering.

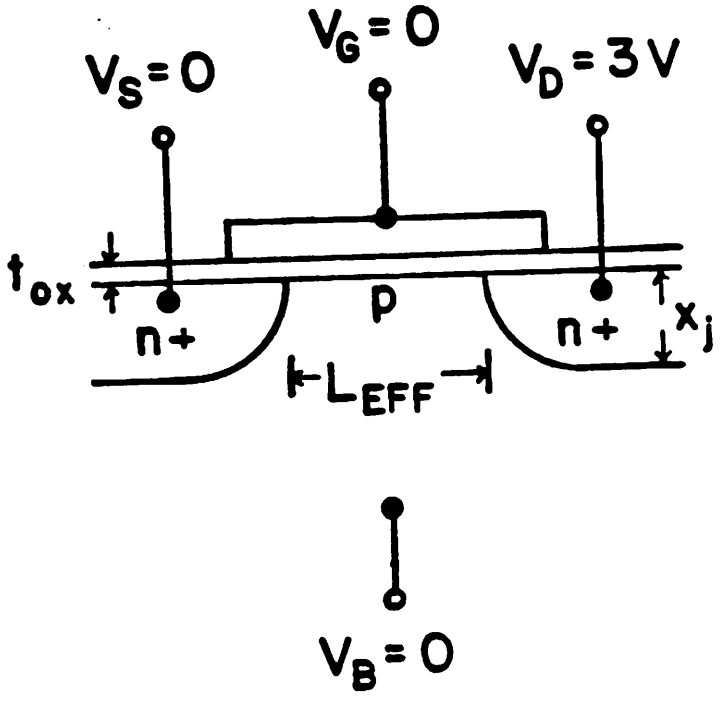


Fig. 4.11 n-channel MOSFET device structure for two-dimensional simulation.

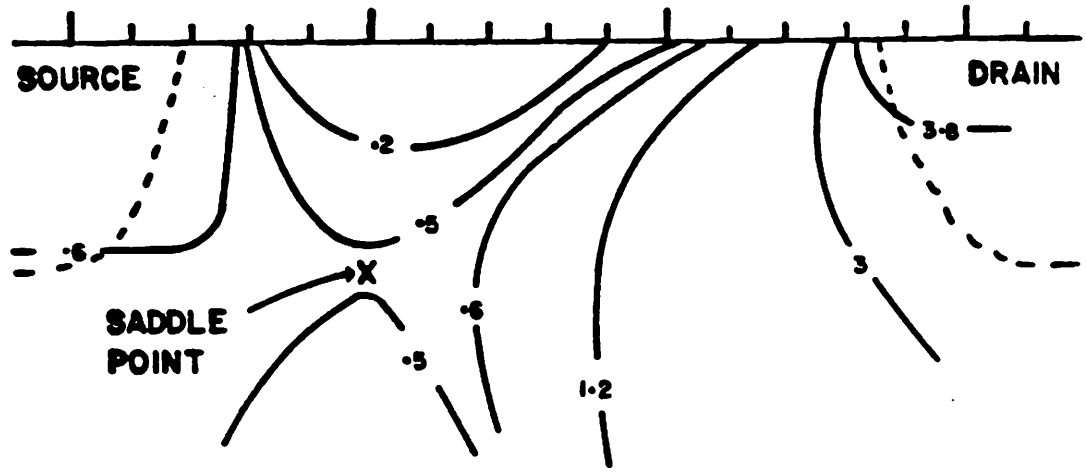


Fig. 4.12 Potential contours with a shallow implant and V_D=3V.

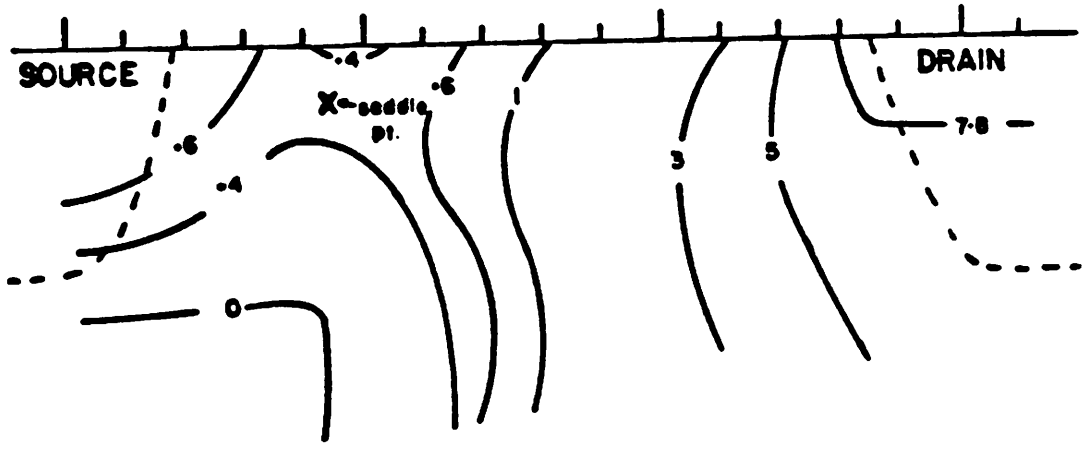


Fig. 4.13 Potential contours with a deep implant and V_D=7V.

respect to the substrate. In the vertical direction, the potential increases from a low surface potential (clamped low by the gate) to a maximum around x_j and decreases to zero at the bulk. In the horizontal direction, the potential decreases from the source end to a minimum, then goes up again towards the drain. The point that has the maximum vertical potential and the minimum horizontal potential is usually called the saddle point and the electric field at this point is zero in all directions. Here, current is purely diffusion. Since this is the maximum potential in the vertical direction, it also has the maximum electron density. We can assume the horizontal current across this vertical plane flows mainly around the saddle point. To evaluate this current, we first assume quasi-equilibrium and calculate the electron distribution from

$$n(\varphi) = N_S e^{\frac{q(\varphi-\varphi_s)}{kT}} \tag{4.9}$$

where N_S and φ_s are the electron concentration and potential at the source, φ is the potential at which we want to calculate the electron concentration n. This is typically done within the program. We next evaluate the current density J_n from

$$J_n = qD_n \frac{\Delta n}{\Delta x} \tag{4.10}$$

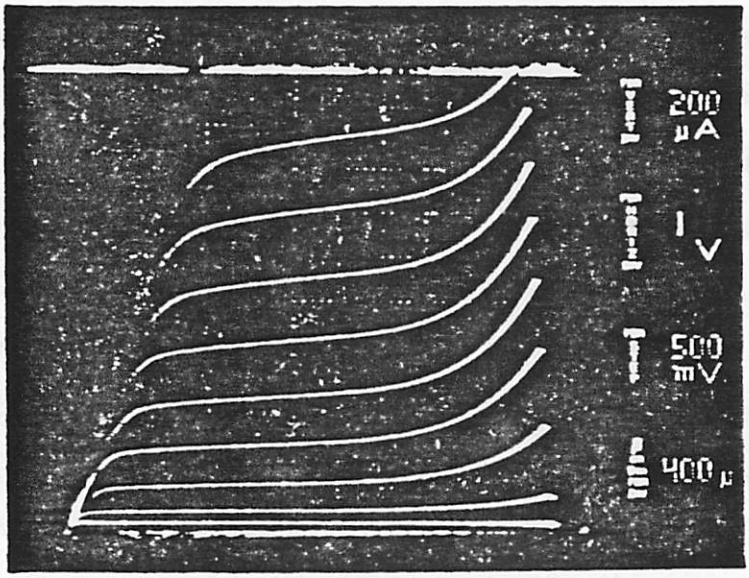
where D_n is the diffusion coefficient, Δn is the change in electron concentration, and Δx is the change in distance. Using this procedure, we approximate the drain current I_D equal to 6.7×10^{-8} A per μm channel width.

To prevent subsurface punchthrough, a common practice is to drive the enhancement channel implant fairly deep into the bulk. We re-simulate the above device with a deeper channel implant. This time we use the same dose (5×10^{-11} cm^{-2}) but the implant profile now peaks at 0.32 μm instead of at the surface. The device no longer exhibit subsurface punchthrough, but with a high enough drain voltage (V_D = 7 V), the saddle point is moved up near the surface

and we have a situation shown in Fig. 4.13. Again, using the above procedure, we approximate the drain current I_D equal to 3.9×10^{-8} A per μm channel width at $V_D = 7$ V. In contrast to the subsurface case, surface punchthrough is strongly affected by the gate and can be reduced with thinner gate oxides.

Fig. 4.14 shows a typical $i-v$ characteristics for an NMOS transistor with a deep threshold adjust implant. This device has an electrical channel length of about 4.3 μm, and the threshold adjust implant dose is $6 \times 10^{11}\ cm^{-2}$ at an energy of 100 KeV. The projected range with this energy is about 0.25 μm from the silicon surface. The problem with this device under 10 V operation is that it exhibits a "soft" breakdown at around 8 V. In fact devices with channel lengths in the 2 μm range show that the I_D characteristics begin to go up dramatically at around 6 V. The mechanism behind this "soft" breakdown characteristic is channel avalanche. This is illustrated in Fig. 4.15. When a MOS transistor is biased deeply into saturation, channel current flows and most of the drain-source voltage is sustained in the pinched-off region at the drain end. With a high enough drain voltage, impact ionization will occur at the drain end. This avalanche breakdown is "soft" as opposed to the "hard" breakdown characteristics of a pn diode. Electrons from impact ionization are swept towards the drain end whereas holes are injected into the substrate. If this hole current is large enough to forward bias the source junction, the parasitic npn bipolar transistor is turned on and the impact ionization hole current will serve as the base current of the transistor.

For the transistor shown in Fig. 4.14, we measure the substrate hole current I_{SUB} as a function of the drain-source voltage V_{DS} at a fixed gate voltage V_G equal to 4 V. The result is shown in Fig. 4.16. Current data below 0.001 μA results from noise in the measurement. Here, we indeed notice the substrate current to go up exponentially with V_{DS}, thus confirming our analysis of a channel avalanche breakdown mechanism. This exponential characteristic bends



$W = 40\ \mu m$

$L_{EFF} \approx 4.3\ \mu m$

$t_{ox} = 500\ A$

$N_D = 4 \times 10^{14}\ cm^{-3}$

V_T adjust implant : dose $6 \times 10^{11}\ cm^{-3}$; energy 100 KeV

Fig. 4.14 **Device characteristics with a deep threshold adjust implant.**

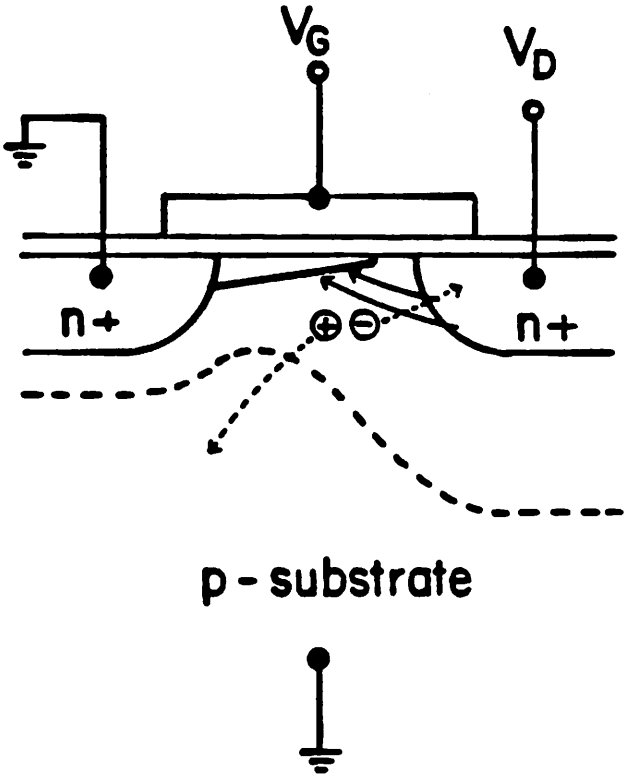


Fig. 4.15 Channel avalanche breakdown in an MOS transistor.

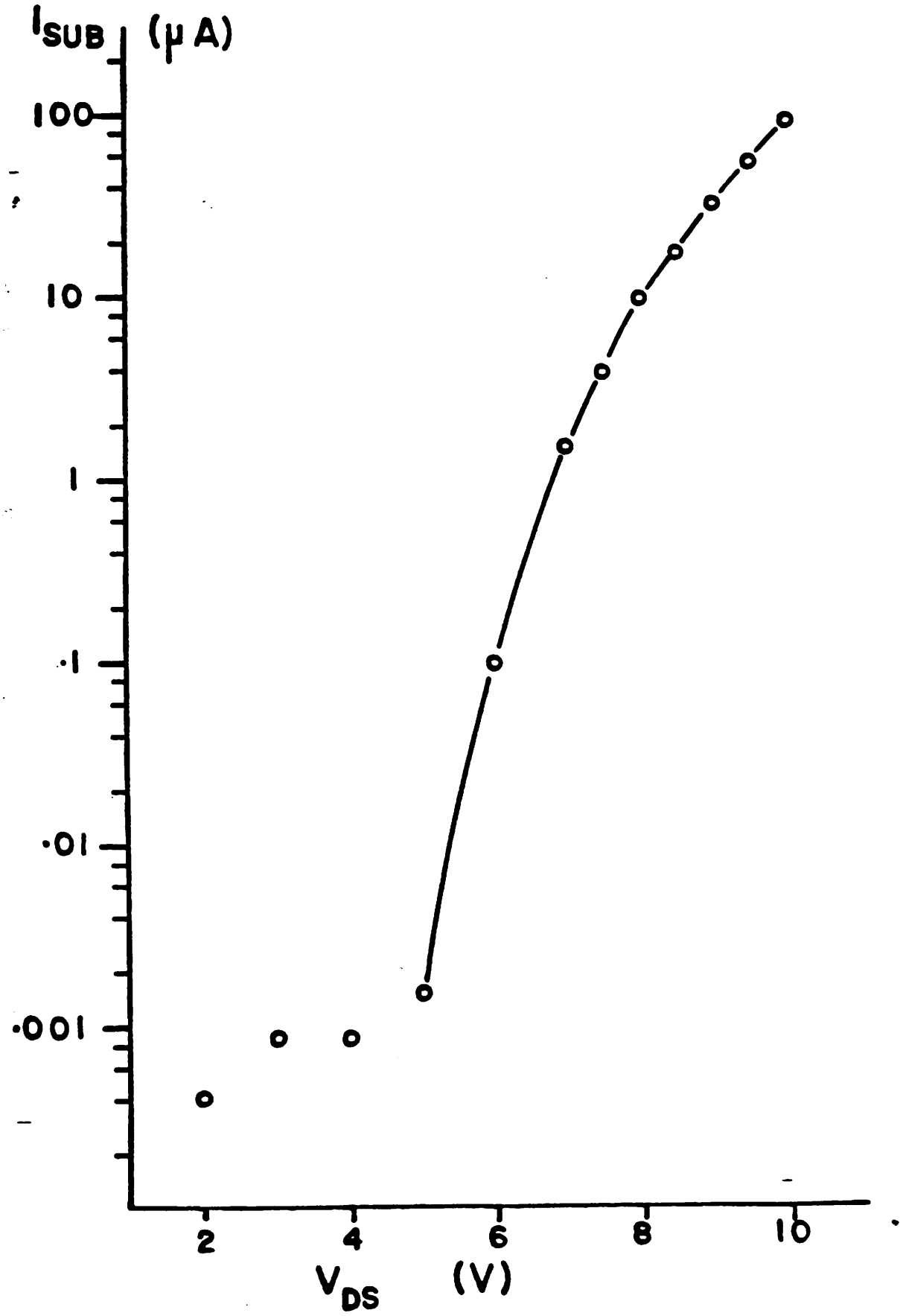


Fig. 4.16 Substrate current for the device shown in Fig. 4.14.

over at around 8 V, showing probable onset of source turn-on, when part of the substrate current recombines with the electrons injected from the source.

This early onset of channel avalanche is partly due to the deep implant. Consider the MOS transistor in Fig. 4.17 where it is biased in the saturation region. The electric field at the drain end is strongest at the corner of the junction due to curvature effects. With a deep threshold adjust implant, this electric field increases because of a higher concentration at the corner region. As the electrons are swept across the pinched-off region, they are pushed away from the surface due to a reversal of the vertical field direction as the drain voltage becomes greater than the gate voltage. This vertical field pushes the electrons towards the corner region of the drain junction where the electric field is strongest. Hence, impact ionization occurs at a smaller V_{DS}, resulting in the "soft" breakdown characteristics shown in Fig. 4.14. If the threshold adjust implant is shallow, the doping concentration at the corner region is smaller, thus lowering the electric field for the same drain voltage. In this case, channel avalanche will not occur until at a higher V_{DS}. This is indeed demonstrated to be true. Fig. 4.18 shows the current characteristics of an NMOS transistor with a shallow threshold adjust implant. Here, the dose is $5 \times 10^{11}\ cm^{-2}$ at an energy of 50 KeV. The projected range for this energy is about 0.1 μm below the silicon surface. The electrical channel length for this transistor is approximately 3.8 μm, which is 0.5 μm shorter than that of Fig. 4.14. However, the drain current does not go up until around 10 V, especially at lower gate voltages. This illustrates that a deep implant indeed results in the early "soft" breakdown characteristics of Fig. 4.14.

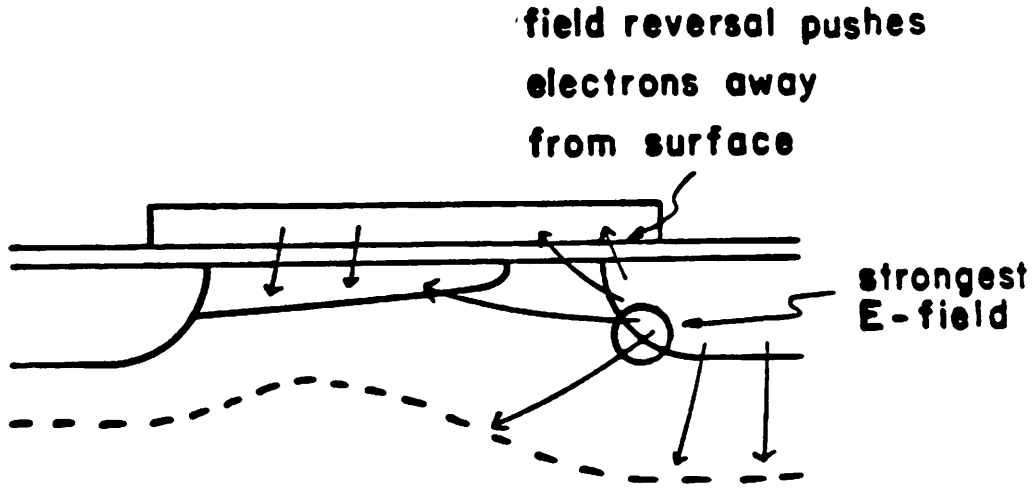
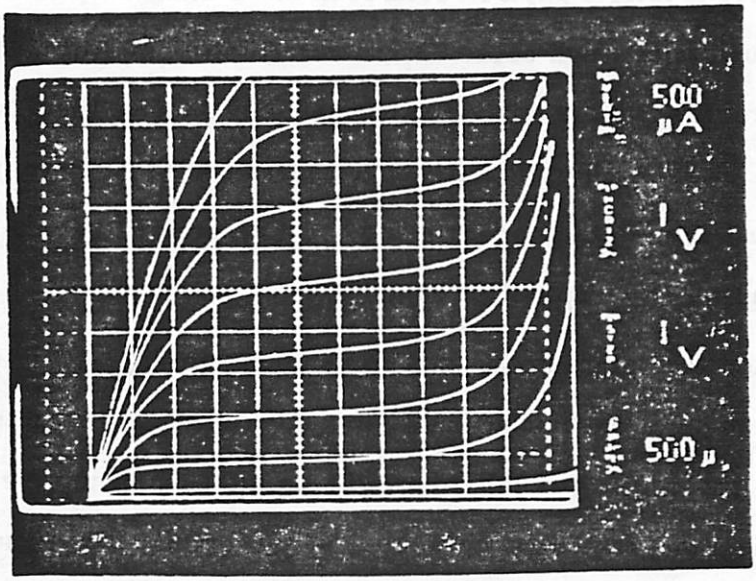


Fig. 4.17 **Electric field lines of a MOS transistor in saturation.**



$W = 40\ \mu m$

$L_{EFF} \approx 3.8\ \mu m$

$t_{ox} = 500$ A

$N_D = 4 \times 10^{14}\ cm^{-3}$

V_T adjust implant : dose $5 \times 10^{11}\ cm^{-3}$; energy 50 KeV

Fig. 4.18 **Device characteristics with a shallow threshold adjust implant.**

4.2.2 CMOS LATCH-UP CONSIDERATIONS

Presence of lateral pnpn structures in CMOS integrated circuits can lead to latch-up conditions which cause excess currents to flow. In many cases, these excess currents are large enough to permanently damage the IC's. The mechanism behind CMOS latch-up is fairly well understood [33]. Here, we will qualitatively discuss the latch-up process and the measures we have taken to prevent latch-up under normal operating conditions.

Consider the lateral pnpn structure shown in Fig. 4.19a, we notice that there are two parasitic bipolar transistors inherent in this structure. One is the lateral pnp shown in Fig. 4.19b, where the p+ diffusion inside the well forms the emitter, the n-well forms the base, and the p-substrate forms the collector. We choose the p-substrate as the collector because it is usually tied to the lowest potential. The other parasitic transistor is the lateral npn shown in Fig. 4.19c. Here, the n+ diffusion inside the substrate forms the emitter, the p-substrate forms the base, and the n-well forms the collector. Again, the n-well is chosen as the collector of the npn transistor since it is usually tied to a higher voltage. Notice that the n-well is the base of the pnp transistor as well as the collector of the npn transistor. Likewise, the p-substrate is the base of the npn transistor as well as the collector of the pnp transistor. Hence this pnpn structure can be modeled as two transistor tied back to back, as shown in Fig. 4.20a. This circuit model is redrawn in Fig. 4.20b for a clearer view.

It is very common to have the p+ diffusion tied to the positive power supply V_{DD} and the n+ diffusion tied to ground, as in the case of a simple inverter shown in Fig. 4.2a. In this case, the n-well will also be tied to V_{DD} and the substrate is grounded. We then have the situation shown in Fig. 4.21a, where we have a substrate resistance R_{SUB} between the substrate contact at the bottom and the place where the substrate forms the base of the npn transistor at the surface.

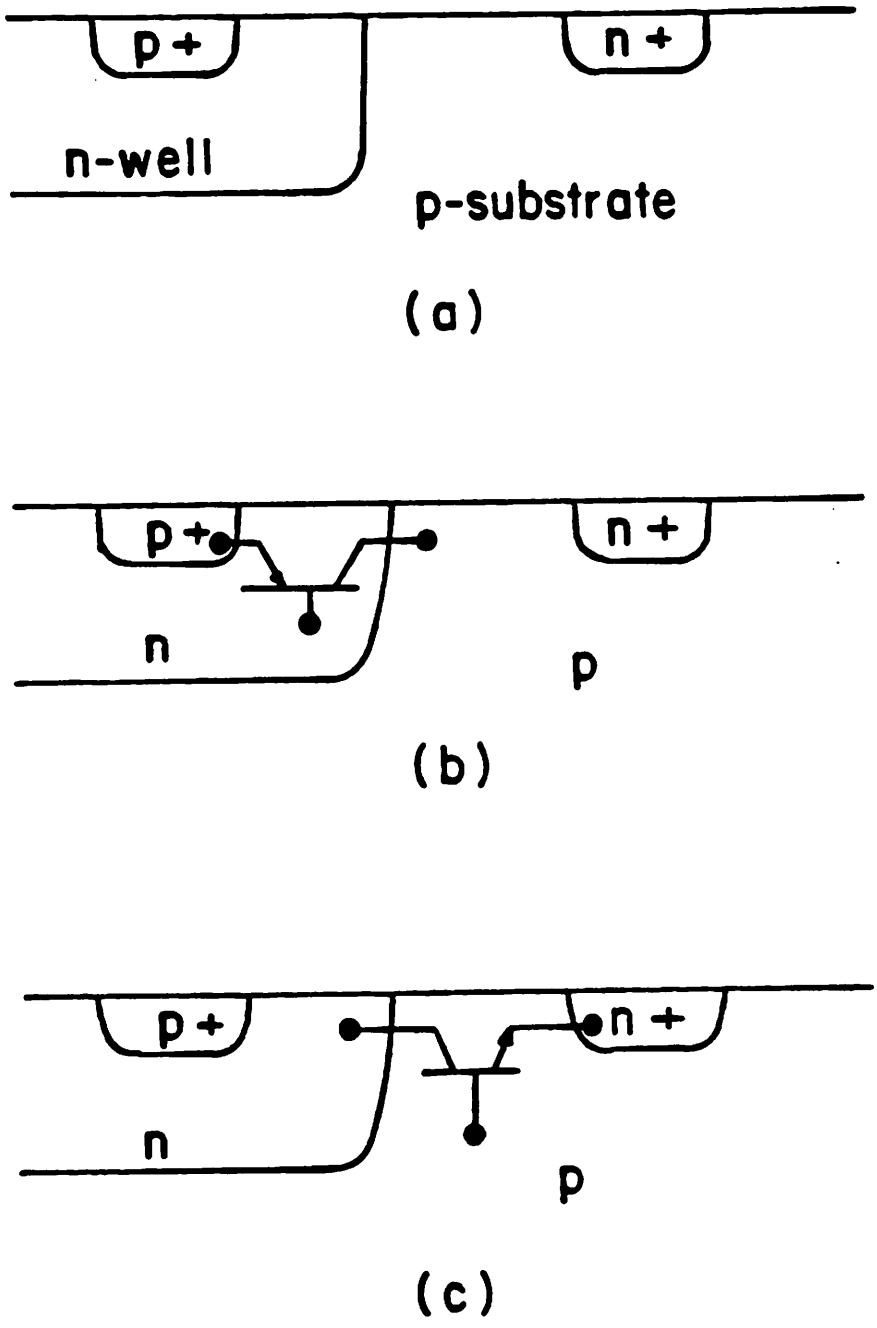
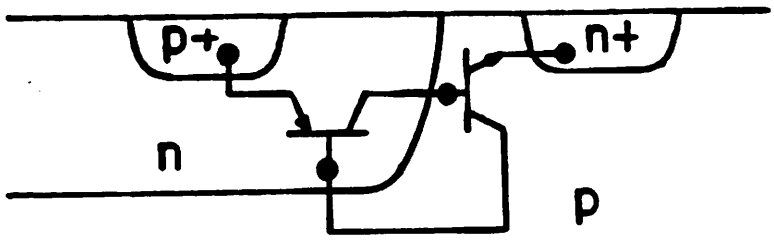
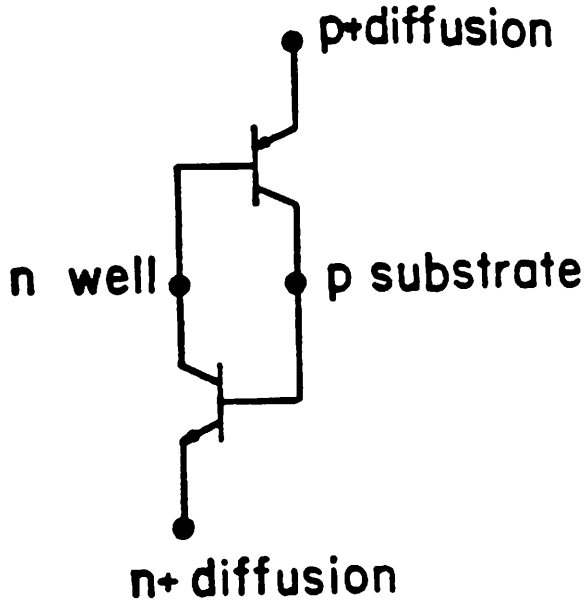


Fig. 4.19 (a) Lateral pnpn structure in a CMOS circuit.
(b) Parasitic pnp transistor in the lateral pnpn structure.
(c) Parasitic npn transistor in the lateral pnpn structure.

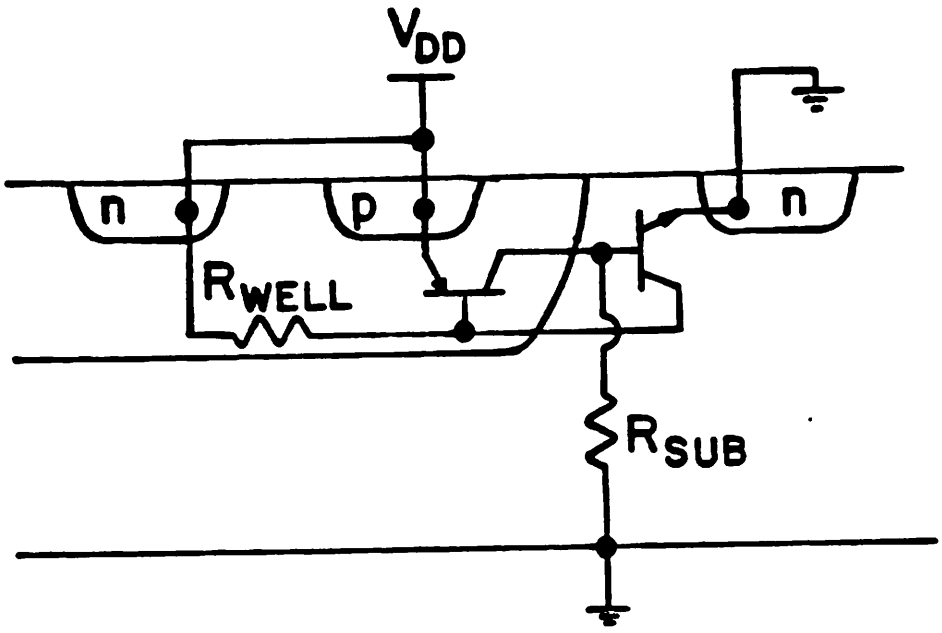


(a)

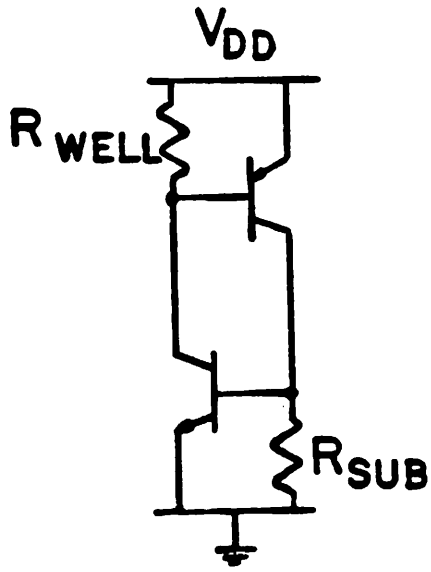


(b)

Fig. 4.20 (a) Parasitic bipolar transistors in the pnpn structure.
(b) Circuit model for Fig. 4.20a.



(a)



(b)

Fig. 4.21 (a) pnpn structure with well and substrate resistances.
(b) Circuit model for Fig. 4.21a.

Likewise, we have a well resistance R_{WELL} between the n+ well contact and the base of the pnp transistor. This situation is redrawn in Fig. 4.21b, where we notice that the two back to back transistors form a positive feedback structure. If somehow one transistor is turned on, the collector current will turn on the other transistor which causes more current to flow, thus further turning on the previous transistor. This regenerative procedure forms a positive feedback cycle which will eventually latch-up the circuit.

There are many ways which can trigger latch-up in the pnpn structure. One such possibility is shown in Fig. 4.22a and b. Suppose the n+ junction is the source junction of an NMOS transistor biased in the saturation region. Impact ionization in the pinched-off region can cause hole substrate current I_{SUB} to flow. When large enough, this substrate current can turn on the n+ source junction due to the potential drop $I_{SUB}R_{SUB}$, thus causing electron injection from the n+ diffusion to the substrate. If some of these electrons get close to the depletion region between the well and the substrate, they will be swept into the well by the electric field across the depletion region. These electrons will travel towards the n+ well contact and collected there. If this electron current is large enough, the potential drop across R_{WELL} can turn on the p+ diffusion in the well, causing hole injection to take place. These holes will also be swept across the depletion region between the well and the substrate by the aiding electric field. Eventually, these holes will be collected by the n+ diffusion since it is already on, or they will be collected by the substrate contact. In both cases, the n+ diffusion will be turned on harder, causing more electron injection to occur. This regenerative procedure thus forms a positive feedback cycle to latch-up the circuit.

Some other ways to trigger latch-up are sudden glitches in the power supply, causing displacement current to flow; or by a large reverse voltage between the well and substrate, causing impact ionization to take place in the well-

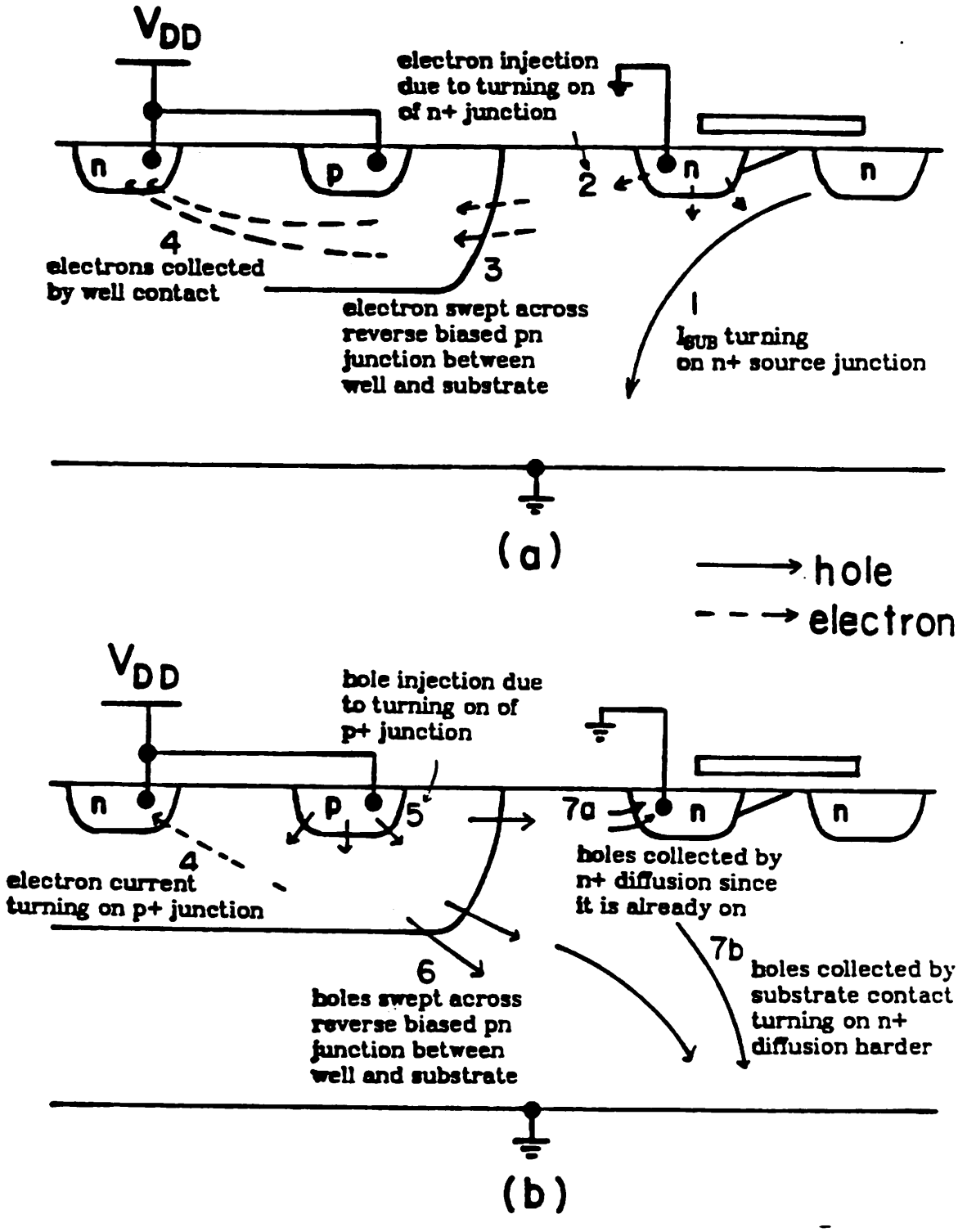


Fig. 4.22 (a,b) Mechanism that trigger CMOS latch-up.

substrate depletion layer. In all cases, once the procedure is initiated, it will regenerate itself until it latches up. There are ways to prevent regeneration of this feedback cycle, or to make it difficult to trigger the procedure under normal operating conditions. If the current gain β of the parasitic npn and pnp transistors are small, the injected electrons from the n+ diffusion will be recombined in the substrate before it can be swept into the well, and likewise for the holes. Thus, there is no positive feedback to regenerate the latch-up procedure. Or if the well and the substrate contacts are close to the p+ and n+ diffusions respectively, it will take a much bigger current to turn on these junctions, thus making it much more difficult to trigger latch-up.

This latter method forms the basis of putting n+ and p+ guard rings around the well to prevent latch-up. This is illustrated in Fig. 4.23. Electrons injected into the well are collected by the n+ guard ring to prevent them from traveling inside the well. Likewise, holes injected into the substrate will be collected by the p+ guard rings tied to ground. To reduce the current gain of the pnp and npn parasitic bipolar transistors, the p+ diffusions inside the well and the n+ diffusions inside the substrate are moved further away from the well-substrate interface, thus increasing the base width of the transistors, thereby decreasing their current gains. Both procedures will take up extra die area, but are necessary to prohibit latch-up. Other methods to control latch-up involve additional process complexity. One popular procedure is to grow an epitaxial layer on a heavily doped substrate and form a buried layer underneath the well. This will reduce the substrate and well resistance to much lower level. Since this method is too costly and time consuming, it is therefore not used in our process.

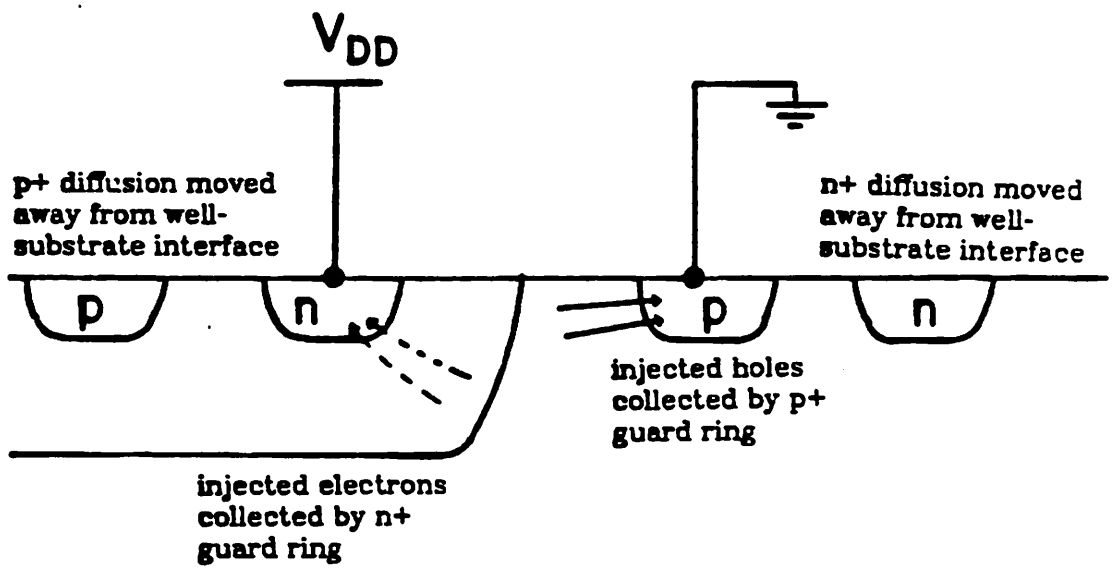


Fig. 4.23 Preventing latch-up in CMOS integrated circuits.

Chapter 5
Experimental Results

The design of a 260 KHz AM IF filter was discussed earlier in Section 3.5. Here, we will present the measured results of such a filter fabricated in CMOS technology and discuss how these results correlate with our designed performance.

The filter was fabricated by National Semiconductor using their 5 μm double-poly CMOS technology. The fully differential switched-capacitor realization shown in Fig. 3.28 was used. Also, since the bandwidth of the filter depends only on the coupling capacitor ratios, an additional "Q-switch" is incorporated which allows a choice of two sets of coupling capacitor ratios so that the bandwidth can be doubled without altering the center frequency. The entire circuit with the "Q-switch" feature is shown in Fig. 5.1, with corresponding capacitor values listed in Table 5.1.

The experimentally observed overall response of the filter is shown in Fig. 5.2a, and the detailed passband response is shown in Fig. 5.2b. The curve with a wider bandwidth corresponds to the case where we double the coupling capacitor ratio. Notice that the center frequency remains constant since the integrator capacitor ratios are not changed. These curves compare favorably with the designed response shown in Fig. 3.29a and b. The pertinent statistics are listed in Table 5.2 where we compare the specifications, the designed response and the experimental results. All the specifications are met and the small discrepancies between the designed and measured results can be explained by small errors in the capacitor ratios.

Table 5.2 also shows the dynamic range, average PSRR, total in-band noise and power dissipation of the experimental filter. The PSRR of the filter was smaller than expected since we would expect a fully differential filter

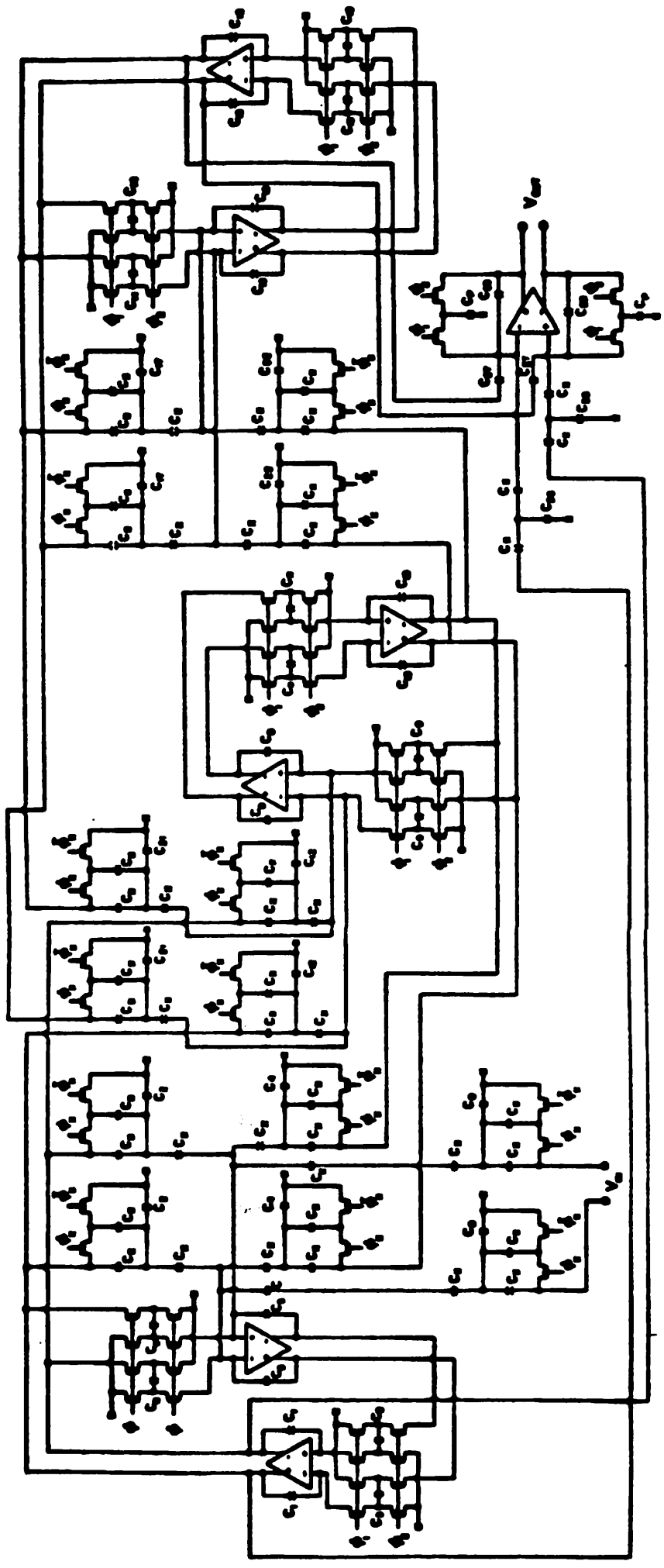


Fig. 5.1 Sixth order elliptic bandpass switched-capacitor filter with "Q-switch" feature for AM IF applications.

capacitor	value (pF)
C_z	0.2
C_1, C_5, C_8, C_{10}, C_{13}, C_{16} C_2, C_4, C_{22}, C_{17},	1.38202737
C_3, C_6, C_9, C_{11}, C_{14}, C_{18}	0.56006799
C_{12} , C_{21}	0.75372972
C_{25}, C_{26}, C_{27}	1.54197825
C_0	1.45049940
C_p	parasitic capacitance

Table 5.1 Component values

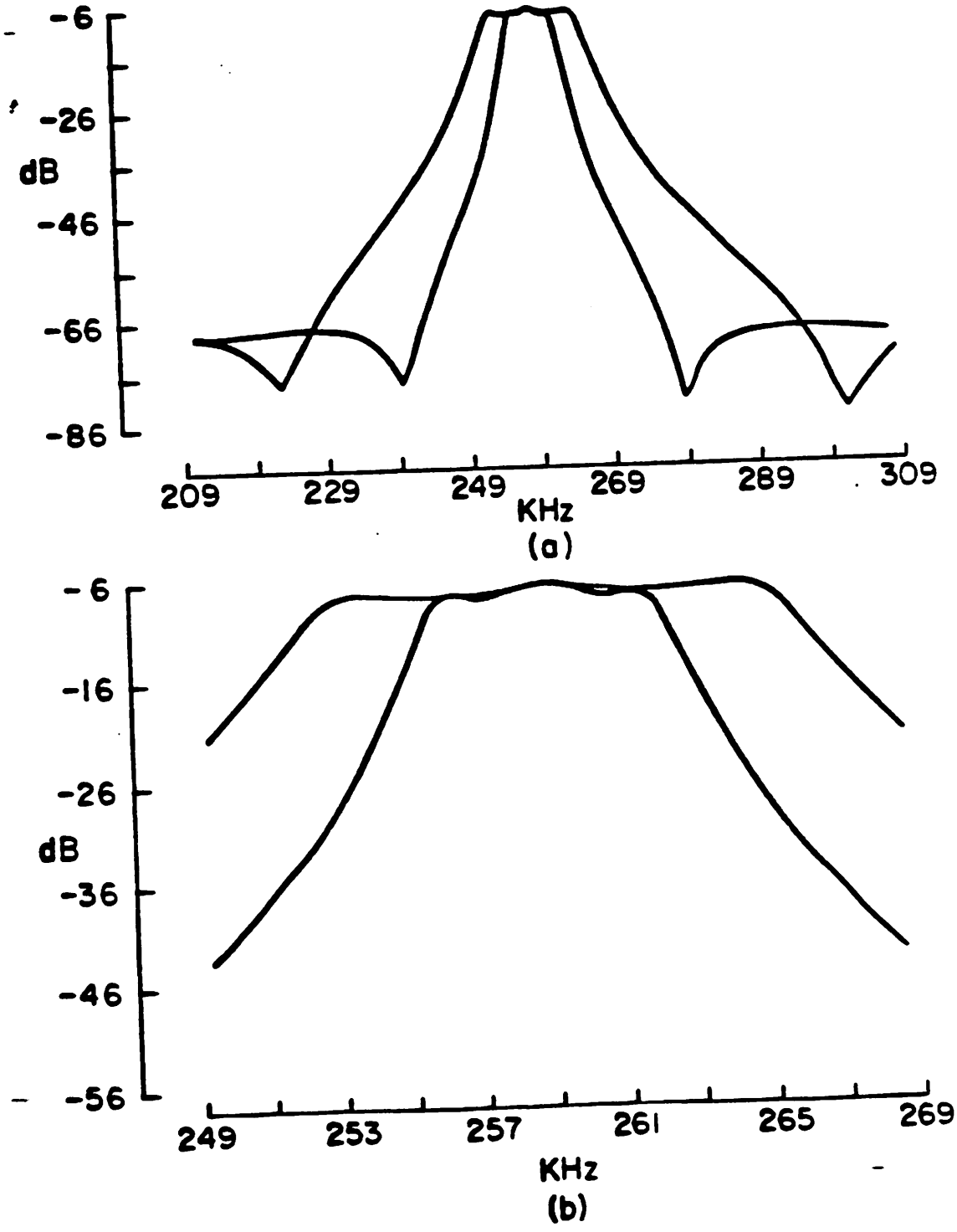


Fig. 5.2 (a) Experimentally observed overall response of AM IF filter.
(b) Experimentally observed detailed passband response of AM IF filter.

	specifications	designed value	experimental results
center frequency	260 KHz ± 1 %	260 KHz	259 KHz
ripple content	3 dB max	0.97 dB	1.3 dB
-3 dB BW	5 KHz min	7.2 KHz	6.5 KHz
rejection at ± 10 KHz	30 dB min	34.7 dB	38 dB
stopband rejection	55 dB min	60.4 dB	62 dB
clock frequency	4 MHz	4 MHz	4 MHz
gain	-6 dB	-6 dB	-6 dB
dynamic range			70 dB
PSRR			30 dB
total in band noise			300 μ V_{rms}
power dissipation			70 mV

Table 5.2 Measured results of AM IF filter at V_{DD}=10v

architecture to be insensitive to power supply variations. A plot of the PSRR as a function of frequency is shown in Fig. 5.3 where we notice that the minimum PSRR in the passband is about 17 dB. This suggested that the circuit was not well balanced, and a close examination of the chip showed non-symmetrical layouts in certain areas. To verify that this is the case, we measured the common-mode to differential-mode response of the filter by injecting an A.C. common-mode signal at the input. Indeed, we notice a large common-mode to differential-mode gain as shown in Fig. 5.4 where we plot the measured frequency response of both the differential-mode signal and the common-mode signal. From this figure, the common-mode rejection ratio in the passband is only about 20 dB, thus showing an unbalanced circuit which gives rise to the smaller than expected PSRR.

The in-band noise of 300 μV_{rms} is large compared with voice band lowpass filters, but this is expected of high Q bandpass ladders since we have large inherent gains inside the resonators. To see why this is true, let us examine the noise sources of switched-capacitor filters in more detail.

There are three kinds of noise sources in switched-capacitor filters [35]. The first kind is know as $\frac{kT}{C}$ noise which corresponds to the thermal noise in MOS switches. The second kind is flicker noise or $\frac{1}{f}$ noise in the MOS operational amplifiers. The third kind is broad band thermal noise of the operational amplifiers. The flicker noise is negligible in our high frequency AM filter since the passband is well above the $\frac{1}{f}$ noise corner frequency. Likewise, the noise contribution due to the amplifier thermal noise is small since transconductance amplifiers have very high output impedances which band-limit this otherwise broad band noise. Thus in our case, we are only concerned with the $\frac{kT}{C}$ noise in the switches. The contribution of this noise can be modeled by an equivalent

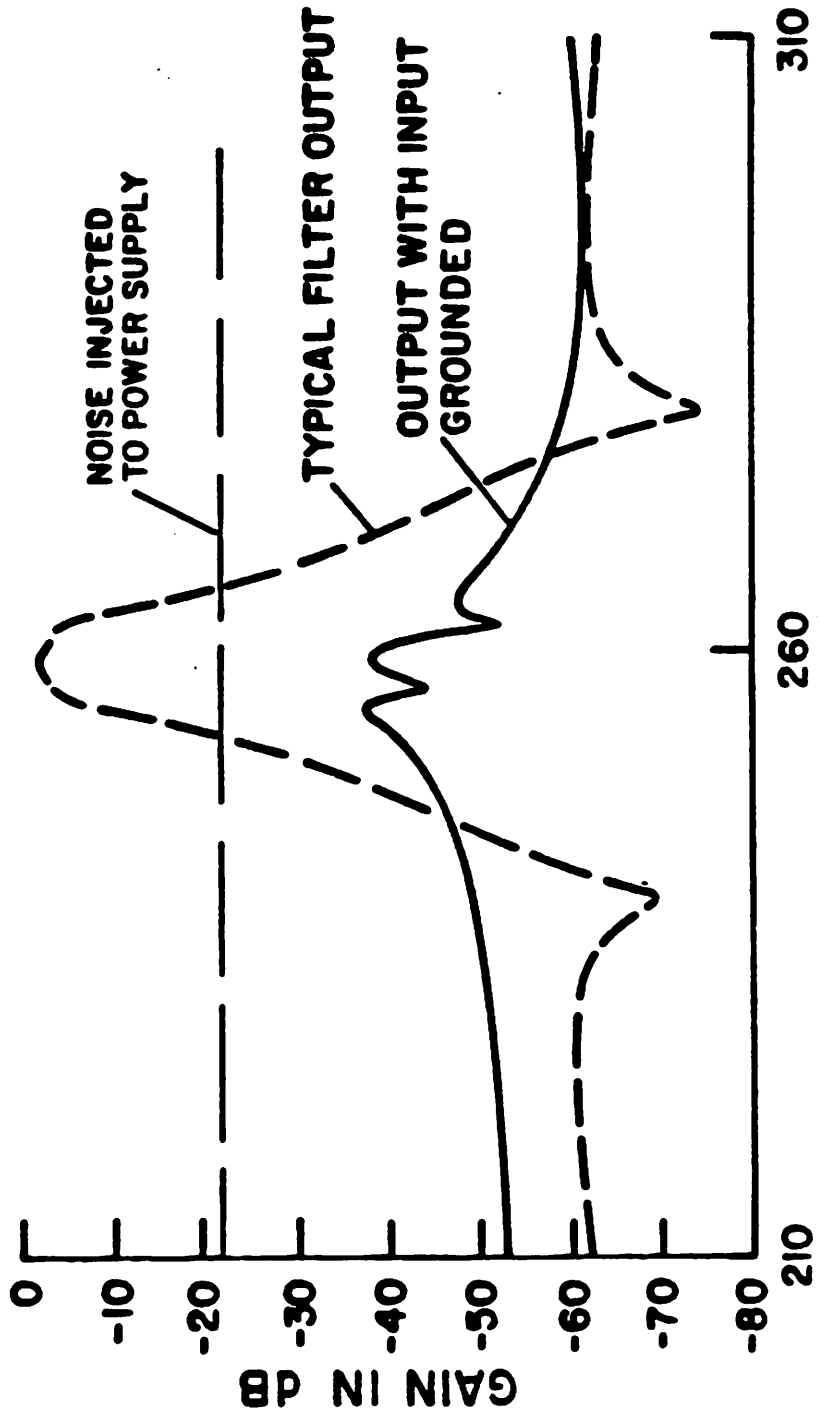


Fig. 5.3 Experimentally observed PSRR response of AM IF filter.

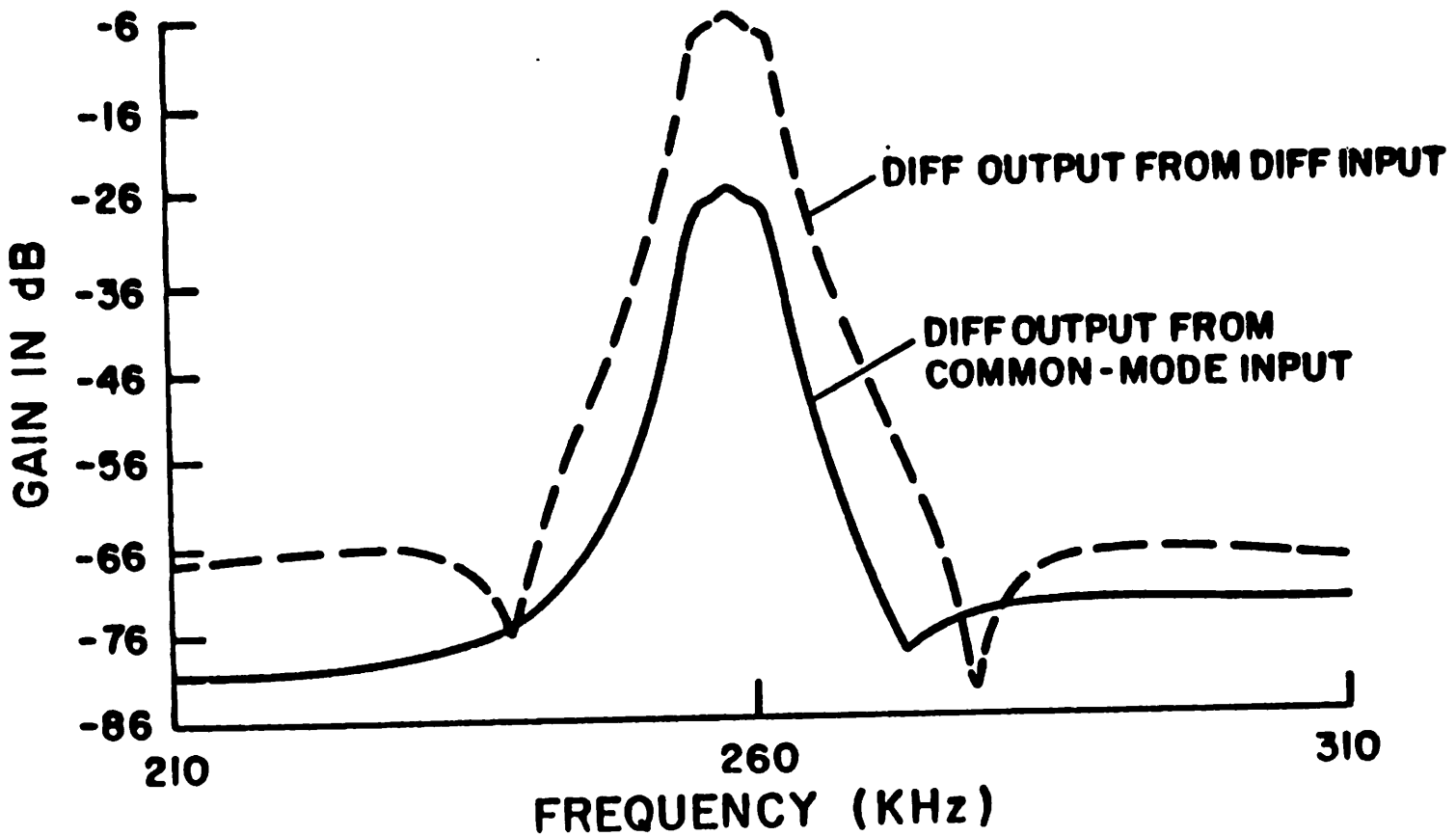


Fig. 5.4 Experimental common-mode and differential-mode frequency response.

input noise source $\overline{v_{in}^2}$ in front of each integrator, and the total output noise is given by

$$\int_{-\infty}^{\infty} \overline{v_o^2} d\omega = \sum_i \int_{-\infty}^{\infty} |H_i|^2 \overline{v_{in}^2} d\omega \tag{5.1}$$

where $\overline{v_o^2}$ is the power spectral density of the total output noise, $\overline{v_{in}^2}$ is the power spectral density of the $\frac{kT}{C}$ noise source in front of the i^{th} integrator, and H_i is the gain from the i^{th} integrator to the output. It can be shown that this gain, H_i, of a high Q bandpass ladder is on the order of Q in the passband, so that the higher the Q, the larger the gain, thus leading to the fact that the noise level of bandpass ladders increases with the Q of the filter.

To show that H_i is on the order of Q in the passband, let us consider a sixth order bandpass ladder with a signal flow diagram shown in Fig. 5.5. Here, we use identical resonator structures so that all the integrators have time constant τ equal to $\frac{1}{\omega_o}$ and all the coupling paths have gains k_i's on the order of $\frac{1}{Q}$. As an example, let us calculate the gain from node 4 to the output. By Mason's theorem, we have

$$H_4(s) = \frac{P_4 \Delta_4}{\Delta} \tag{5.2}$$

where Δ is the system determinant evaluated from the loop gains of the signal flow graph and, in this case, is given by

$$\begin{aligned}\Delta = 1 &+ \frac{1}{s\tau}(k_1 + k_6) + \frac{1}{s^2\tau^2}(3 + k_2k_3 + k_4k_5 + k_1k_6) \\ &+ \frac{1}{s^3\tau^3}(2k_1 + 2k_6 + k_2k_3k_6 + k_1k_4k_5) + \frac{1}{s^4\tau^4}(3 + k_2k_3 + k_4k_5 + k_1k_6) \\ &+ \frac{1}{s^5\tau^5}(k_1 + k_6) + \frac{1}{s^6\tau^6}.\end{aligned} \tag{5.3}$$

P_4 is the direct path from node 4 to the output and Δ_4 is the system determinant excluding the loops which touch P_4. Hence, for Fig. 5.5

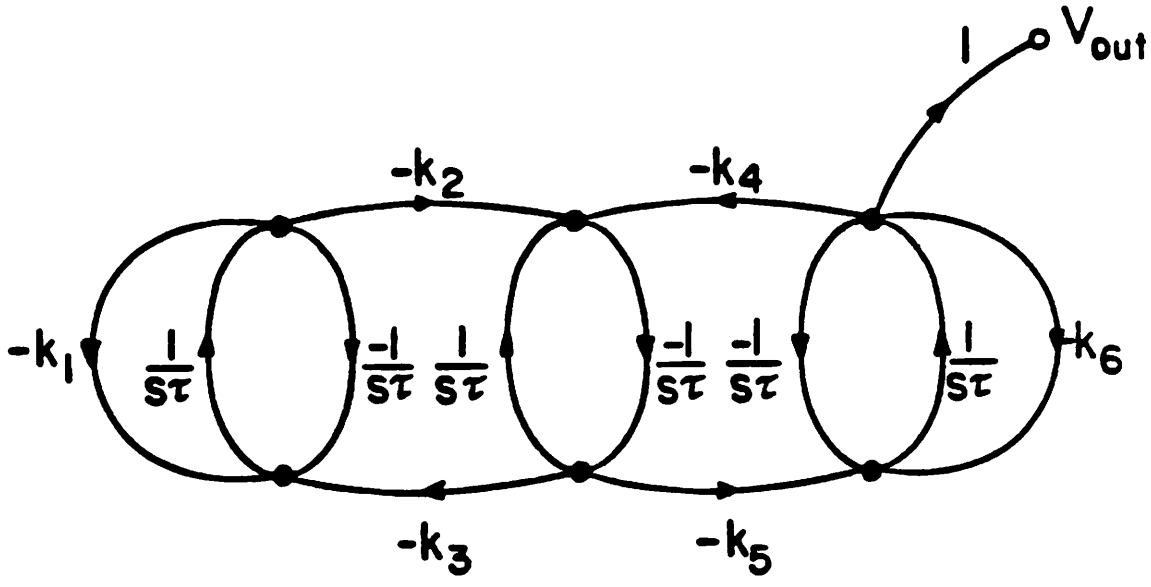


Fig. 5.5 Signal flow diagram of sixth order bandpass ladder for noise analysis.

$$P_4\Delta_4 = -\frac{k_5}{s\tau}\left[1 + \frac{1}{s^2\tau^2} + \frac{k_1}{s\tau}\right]. \tag{5.4}$$

At $s = j\omega_o$, $\omega_o\tau = 1$ so that

$$\Delta\Big|_{\omega_o} = k_2k_3k_6 + k_1k_4k_5 \tag{5.5a}$$

$$P_4\Delta_4\Big|_{\omega_o} = k_1k_5 \tag{5.5b}$$

and

$$H_4(j\omega_o) = \frac{k_1k_5}{k_2k_3k_6 + k_1k_4k_5}. \tag{5.6}$$

Since $k_i \sim \frac{1}{Q}$, we have

$$H_4(j\omega_o) \sim \frac{1/Q^2}{1/Q^3}$$

$$\sim Q. \tag{5.7}$$

In general, it can be shown that for an n^{th} order bandpass ladder, the system determinant Δ evaluated at the center frequency is on the order of $\frac{1}{Q^{\frac{n}{2}}}$ whereas $P_i\Delta_i$ is on the order of $\frac{1}{Q^{\frac{n}{2}-1}}$, so that $H_i(j\omega_o) \sim Q$ for all i's. Thus, the passband gain from the noise sources to the output is on the order of Q for bandpass ladders in general. For the AM filter we designed, we have a Q of about 40, so that we would expect $H_i(j\omega_o)$ to be on the order of 32 dB. Indeed, by simulations using DINAP, $H_i(j\omega_o)$ varies from 30 to 34 dB for the different nodes. If we assume all H_i's are equal to 40 throughout the passband, we can calculate a first order estimate of the total in-band noise using equation (5.1). For our sixth order bandpass filter

$$\int_{\text{passband}} \overline{v_o^2}\,d\omega \approx 6 \times 40^2 \times \overline{v_n^2}\Big|_{\omega_o} \times B \tag{5.8}$$

where B is the bandwidth and is equal to 6.5 KHz. The power spectral density of the equivalent input $\frac{kT}{C}$ noise in the passband is given by [35]

$$\overline{v_{in}^2} = \frac{2kT}{f_s C_S} \tag{5.9}$$

if the sampling capacitor C_S is much smaller than the integrating capacitor C_I. In our case, we have

$$\frac{C_I}{C_S} \approx 2 \tag{5.10}$$

so that we would expect our first order calculation would over-estimate the actual noise. In our case, C_S is roughly 0.5 pF and f_s is 4 MHz. Substituting into (5.8), we have

$$\int_{passband} \overline{v_o^2} d\omega \approx 6 \times 40^2 \times \frac{2kT}{4 \times 10^6 \times 0.5 \times 10^{-12}} \times 6.5 \times 10^3$$
$$\approx 2.6 \times 10^{-7} \tag{5.11}$$

so that the total in-band noise is $(2.6 \times 10^{-7})^{1/2}$ or 510 μV_{rms}. A more accurate SPICE [39] simulation of this noise contribution can be found in [36] which gives roughly the same result. Here, the switched-capacitor is replaced by an equivalent resistor which turns out to have the same noise contribution as the $\frac{kT}{C}$ noise. In both cases, we have assumed the sampling capacitor C_S to be much smaller than C_I so that we over-estimate the output noise. In fact, our hand calculated value of 510 μV_{rms} is on the same order of the measured value 300 μV_{rms}, which shows that our measured result is reasonable.

The fact that high Q filters have greater noise basically stems from the inherent large gains of the resonators which is only limited by the terminations, so that the noise contributions are enhanced by the large gains. In high frequency filters, this situation is made worse because the sampling capacitors cannot be too large. In our case C_S is roughly 0.5 pF which is an order of

magnitude smaller than the sampling capacitors found in most codec filters. A plot of the measured noise response of the AM IF filter is shown in Fig. 5.6.

In addition to the above measurements, the frequency response from different chips are measured to check the variations from die to die. This is shown in Fig. 5.7, which demonstrates very consistent frequency characteristics, with no variations in the center frequency at all. Thus, the identical resonator filter architecture indeed helps to reduce sensitivity problems. The photomicrograph of the fabricated AM filter is shown in Fig. 5.8. Additional characterization of this chip can be found in [36].

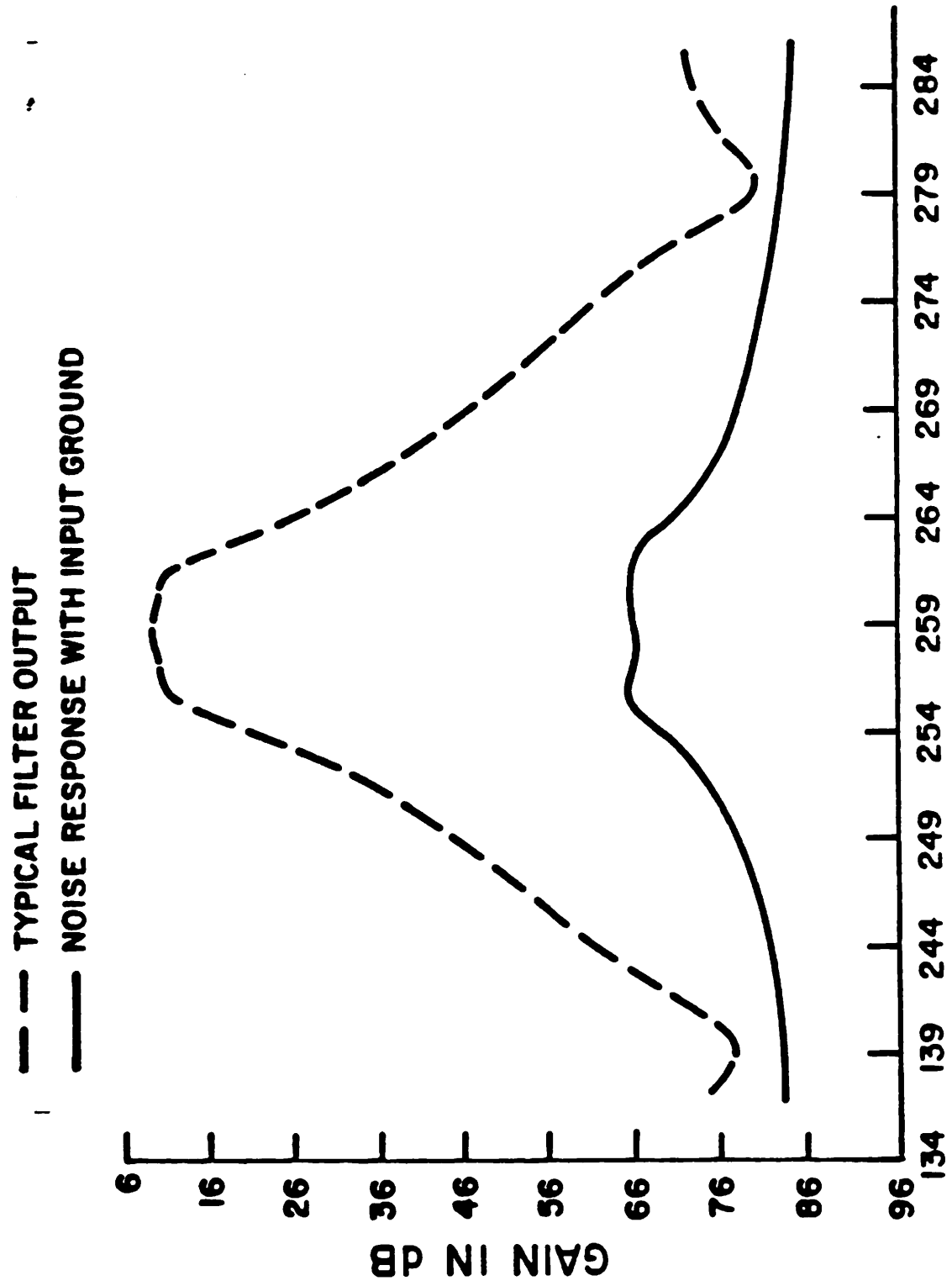


Fig. 5.6 Measured noise response of AM IF filter.

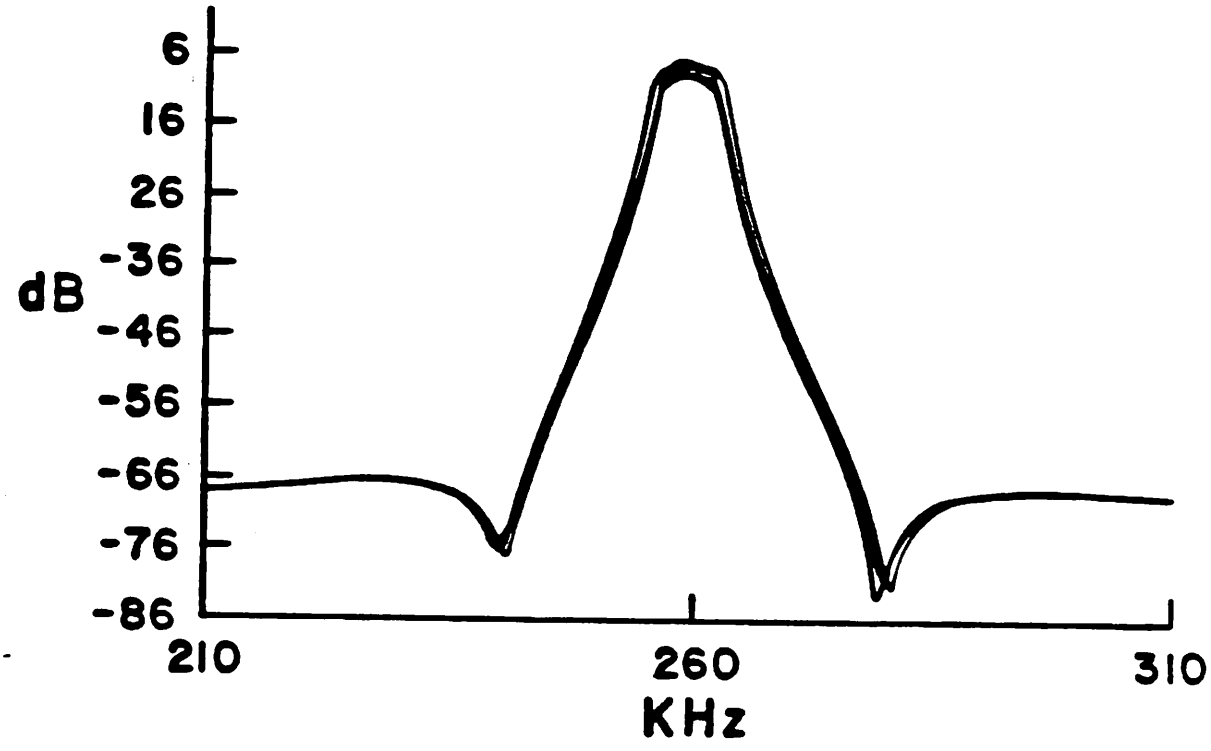


Fig. 5.7 Frequency response from five different dies.

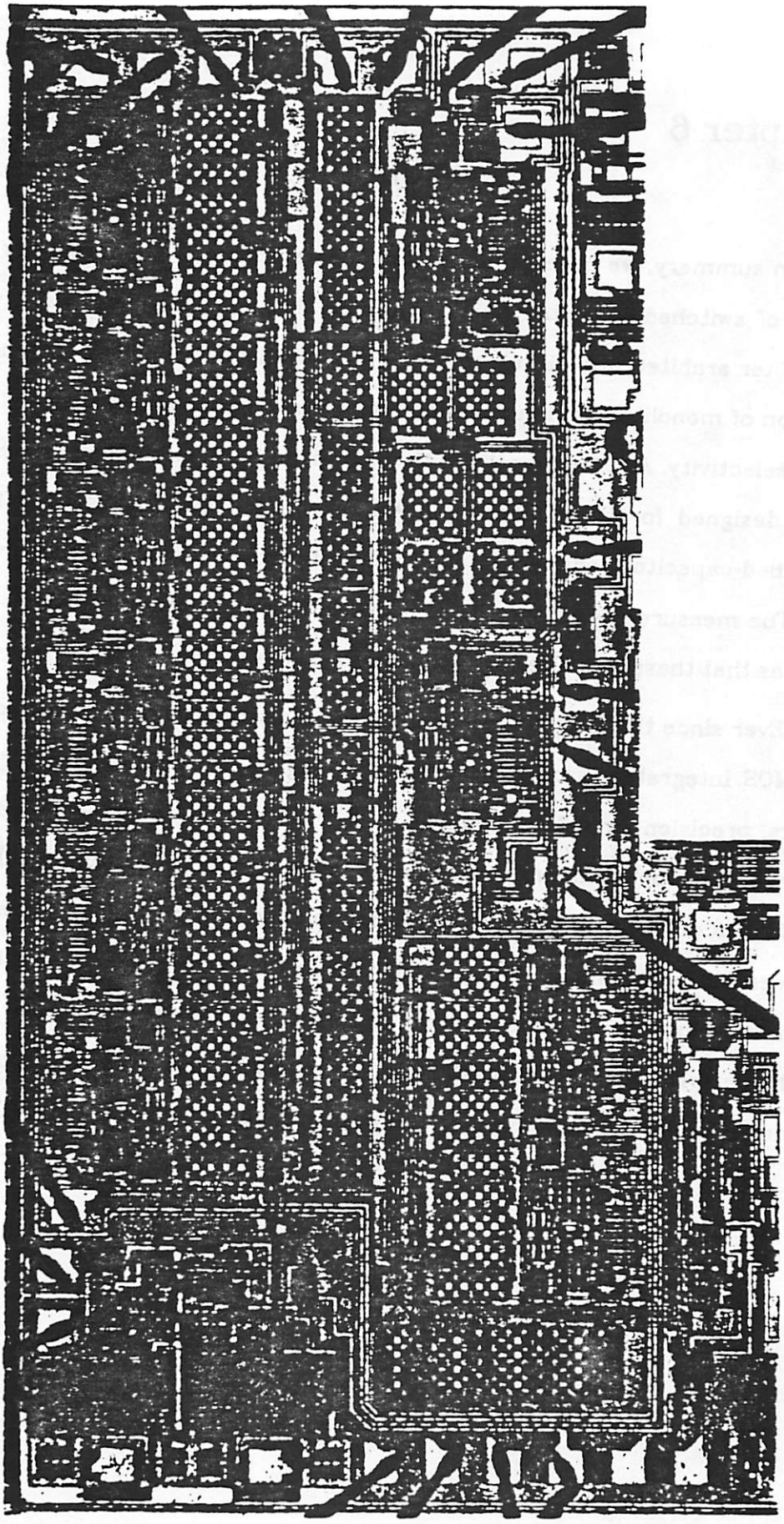


Fig. 5.8 Photomicrograph of AM IF filter.

Chapter 6

CONCLUSION

In summary, we have gained better understanding of the frequency limitations of switched-capacitor filters. Through this understanding, we developed new filter architecture, circuit concepts and advanced technology for a new generation of monolithic communication filters which require high speed as well as high selectivity. All these results in the successful fabrication of an experimental chip designed for IF application in AM receivers. This chip is a sixth order switched-capacitor bandpass filter with a center frequency at 260 KHz and a Q of 40. The measured results agree with the designed performance, which demonstrates that these new filtering techniques are feasible.

Ever since the advent of all-MOS charge-redistribution A/D converters, analog MOS integrated circuits have come a long way to present day high speed filters, precision amplifiers, and 14+ bit A/D converters. Looking ahead, we will be faced with new challenges as process technology advance towards the 1 μm level. The problem of channel length modulation and noise in the scaled MOS devices will be the main obstacles to overcome before we will be able to utilize the new technological improvements to further advance the application as well as the performance of analog MOS IC's. To conclude, there will always be a need for new concepts and new ideas in this ever-expanding field of analog MOS.

Appendix A:
Single-Level Polysilicon, Inverted CMOS Process

The following is a description of a single-level polysilicon, inverted CMOS process developed in the course of this work. The starting material is a <100> boron-doped wafer with a nominal resistivity of 30 Ω–cm. This appendix is aimed at understanding the process and is descriptive in nature. The detail instrument settings used in our Semiconductor Laboratory (like exposure settings on our Canon 4:1 stepper) can be found in [40].

1. Initial Oxidation

An initial oxide is grown in "WET" oxygen ambient at 1000° C for 35 minutes. This results in an oxide thickness of 3000 Å. The reason for such a thick oxide growth is to provide a step on the wafer for later alignments. This will be explained in Step 5.

2. MASK #1: Define N-Wells

Standard lithography procedure is used to define the n-well regions. Next, the oxide on top of the n-wells is etched away in buffered HF. For this step, line width control is not important so that we can easily tolerate a 50% over-etch. The photoresist is then stripped off after etching the oxide.

3. N-Well Implantation

The n-well regions are implanted with phosphorus at 100 KeV, with a dose of 1.5×10^{2} cm^{-2}.

4. N-Well Drive-In

First, the phosphorus ions are driven deep under the silicon surface at 1100° C in "DRY" oxygen ambient for 280 minutes. This is followed by a high

temperature (1150°C), long drive-in (720 minutes) in a 10% "DRY" oxygen, 90% nitrogen ambient. The purpose of the oxygen presence is to enhance the diffusion process by oxidation. This step results in a final well depth of 4 μm, with about 3500 Å oxide on the well regions and 4800 Å oxide on the field regions.

5. Gate Oxidation

The initial oxide on the wafer is stripped away in buffered HF for about 4 minutes. Since there is no photoresist on the wafer, non-buffered HF can also be used. A gate oxide is then grown at 1000° C for 48 minutes, resulting in an oxide thickness of 500 Å. The wafer profile after this step is shown in Fig. A1. Notice that there is a slight dent in the well region. This is because there is a 1700 Å difference in the additional oxide grown in Step 4 between the well and field regions. The reason for the slower growth on the field regions is due to the presence of a thick oxide to begin with. Since a thermally grown oxide consumes about 40% of its thickness in silicon, the 1700 Å oxide difference creates a 680 Å dip in the well regions. This dip is used for alignment purposes by later masks which use the n-well mask as reference.

6. Nitride Deposition

A layer of silicon nitride, about 1000 Å thick, is deposited over the wafer using a low-pressure chemical vapor deposition (LPCVD) process. The gas used is a mixture of ammonia and silane diluted in argon. This nitride layer is used to mask against oxygen diffusion in subsequent local oxidation steps.

7. MASK #2: Define NMOS Active Regions

This mask defines the NMOS active regions on the substrate. The nitride on the field regions is etched away using plasma etching techniques. The resulting profile is shown in Fig. A2. Notice that the nitride on the well regions is left

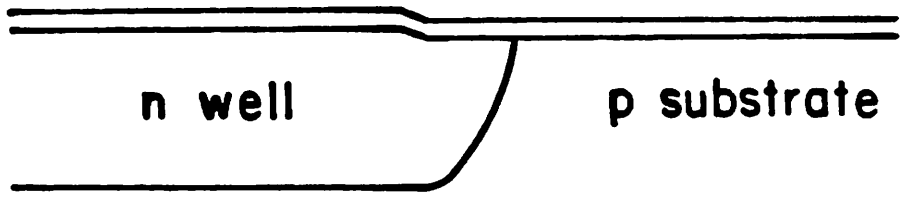


Fig. A1 Wafer cross-section after gate oxidation (Step 5).

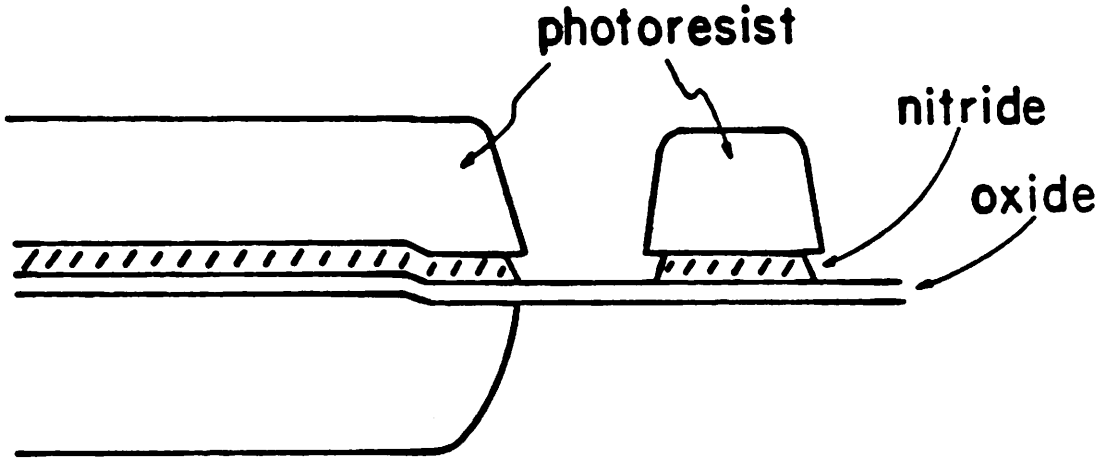


Fig. A2 Wafer cross-section after nitride etch (Step 7).

intact and will be etched again in subsequent masking steps for PMOS active regions. The photoresist remains after this step for protecting the active regions against the field implant.

8. P-Field Implantation

Boron is implanted into the p-field regions in the substrate to increase the field threshold to about 20 V. The energy used is 100 KeV, with a dose of 1 x 10^{13} cm^{-2}.

9. Backside Preparation and Implantation

Since we will be using aluminum instead of gold to contact the back of the wafer, a Schotty contact may result since the wafer has a low substrate doping (~ 4 x 10^{14} cm^{-3}). To ensure an ohmic contact for the substrate, boron is implanted into the backside to increase the doping level there. We first strip off the remaining photoresist from the previous step. Then a new coat of photoresist is put onto the front side to protect the structure already there. Next, the wafer is dipped into buffered HF to remove the native backside oxide. The wafer backside is then implanted with BF_2 at 200 KeV, with a dose of 2 x 10^{15} cm^{-2}. BF_2 is used instead of boron ions since it gives a higher beam current for faster through-put. Also, the florine ions will not be electrically active and thus will not affect the device electrical characteristics.

10. P-Field Drive-In

After the backside implant, the protecting photoresist on the front side is stripped away. Next, the boron field implant is driven-in at 1000° C in a "DRY" oxygen ambient for 20 minutes. We should be careful in tailoring this drive-in (and subsequent local oxidation steps) since if the boron is too deep, the diffusion edge capacitances will be too large. On the other hand, if it is too shal-

low, the device may leak along the channel edges.

Usually, this step is done right before local oxidation, but in this case, we do it before masking the second nitride etch. The reason for this is because the etch differential between nitride and oxide in SF_6+O_2 plasma is only about 4:1. Hence, part of the oxide is already etched away in the field regions by the first nitride etch. In the second nitride etch to follow, there will be regions where the oxide is etched again (with no nitride on top). To prevent pitting the silicon surface, this step is done before the second nitride etch since this will grow some oxide in the field regions.

11. MASK #3: Define PMOS Active Regions

This mask defines the PMOS active regions on the well. Plasma etch is used to remove the nitride in the field regions in the well. During this step, the nitride on the NMOS active regions is protected by photoresist.

12. Local Oxidation

The photoresist is first stripped off before local oxidation. The field oxide is then grown in 920°C in "WET" oxygen for 420 minutes, resulting in a final field oxide thickness of 7500 Å. The reason for the slow oxide growth is to prevent excessive boron segregation into the field oxide. The wafer cross-section after this step is shown in Fig. A3.

13. MASK #4: Define Polysilicon to Substrate Capacitor Regions

Most analog applications require a true capacitor which is insensitive to voltage variations across the capacitor plates. This process provides a polysilicon to substrate capacitor by a heavy n+ implant so that the MOS capacitor is always well into accumulation under 10 V operations. This mask defines the capacitor thin oxide regions and plasma etch is used to remove the nitride on

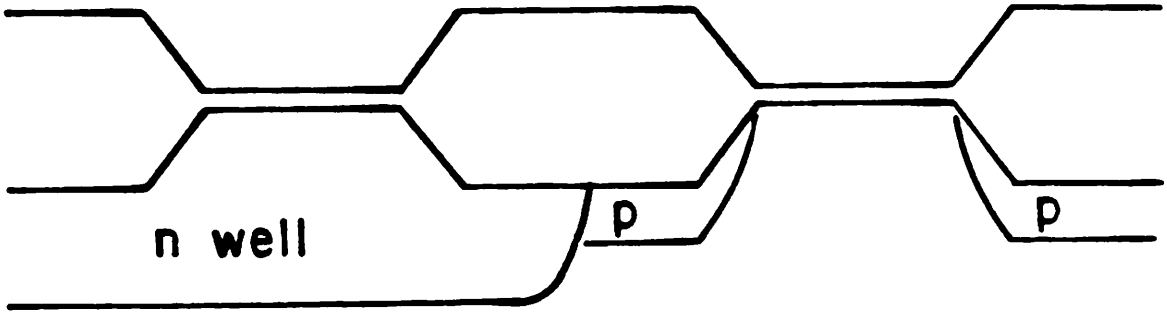


Fig. A3 Wafer cross-section after local oxidation (Step 11).

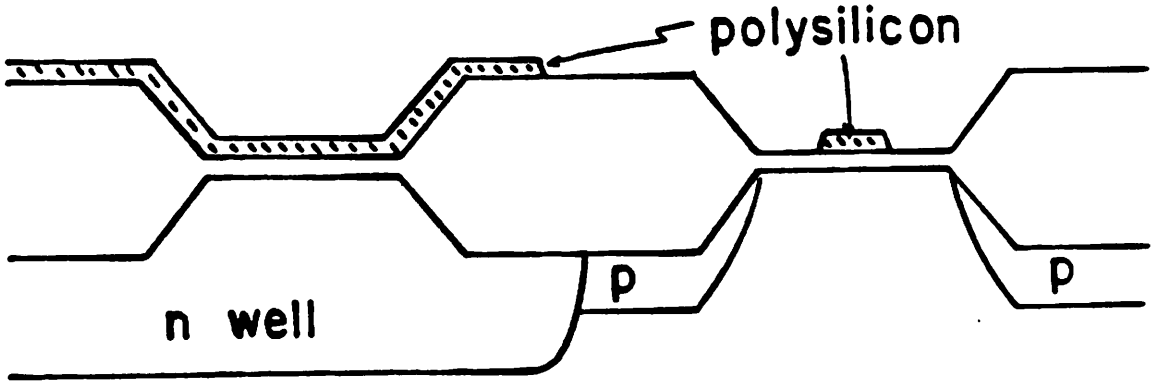


Fig. A4 Wafer cross-section after etching NMOS polysilicon (Step 19).

top of these regions. Since this will also partially remove the capacitor gate oxide, it has to be dipped off in buffered HF and regrown later.

14. Capacitor Implantation

A heavy phosphorus dose of 8 x 10^{14} cm^{-2} is implanted into the capacitor regions at 50 KeV. The photoresist is then removed after the implantation.

15. Regrow Capacitor Oxide

This is done at 1000° C in "DRY" oxygen ambient for 80 minutes, resulting in an oxide which is 750 Å thick. This step also drives-in the phosphorus ions.

16. Nitride Removal

A very thin layer of oxide is grown on the nitride surface during local oxidation. This has to be dipped off first in 10:1 HF. A thorough rinse is required before the nitride removal since remaining streaks of HF will attack the gate oxide once the nitride is removed. Next, nitride is etched in hot phosphoric acid at 155° C. To ensure constant temperature, a reflux system which returns the boiled off water vapor back to the phosphoric acid is used. This is done to ensure a constant boiling point at 155° C. If the temperature of the phosphoric acid is too high, the etch differential between the nitride and oxide will degrade, which will result in a partial removal of the gate oxide.

17. Threshold Adjust Implantation

A blank implantation of boron at 50 KeV with a dose of 5 x 10^{11} cm^{-2} is used to adjust the device threshold voltages. The nominal long channel device thresholds are 0.5 V for the NMOS transistors and -0.5 V for the PMOS transistors.

18. Polysilicon Deposition

Polysilicon is deposited using a LPCVD process. The gas used is silane diluted in argon. The thickness of the polysilicon layer deposited is 4000 Å.

19. Phosphorus Predeposition for Doping Polysilicon

This is done at 950° C in a $POCl_3$ and oxygen ambient for 30 minutes. The carrier gas is nitrogen. In addition to doping the polysilicon with phosphorus, this step also grows a thin layer of oxide on top of it, so that a short oxide dip is required after the predeposition. The wafer has to be thoroughly rinsed after the oxide dip to prevent remnants of HF from sipping through the grain boundaries of the polysilicon down to the gate oxide below.

20. MASK #5: Define NMOS Polysilicon

This mask defines the polysilicon runs on the substrate regions. A SF_6+O_2 plasma is used to etch the polysilicon. Line width control is important here since this defines the device geometry and thus the electrical parameters like threshold voltage, which varies greatly with channel length. As in MASK #2, the polysilicon on the well regions are not touched in this step. The device profile after stripping off the photoresist is shown in Fig. A4.

21. N+ Source/Drain Implantation

A heavy dose of arsenic is implanted to form the n+ source and drain diffusions. The energy used is 180 KeV, with a dose of 3×10^{15} cm^{-2}. The source/drain regions so formed are self-aligned to the polysilicon gate, which is also implanted along with the exposed silicon surfaces.

22. MASK #6: Define PMOS Polysilicon

This mask defines the polysilicon runs for the PMOS devices in the well regions. The NMOS polysilicon is protected by a layer of photoresist against

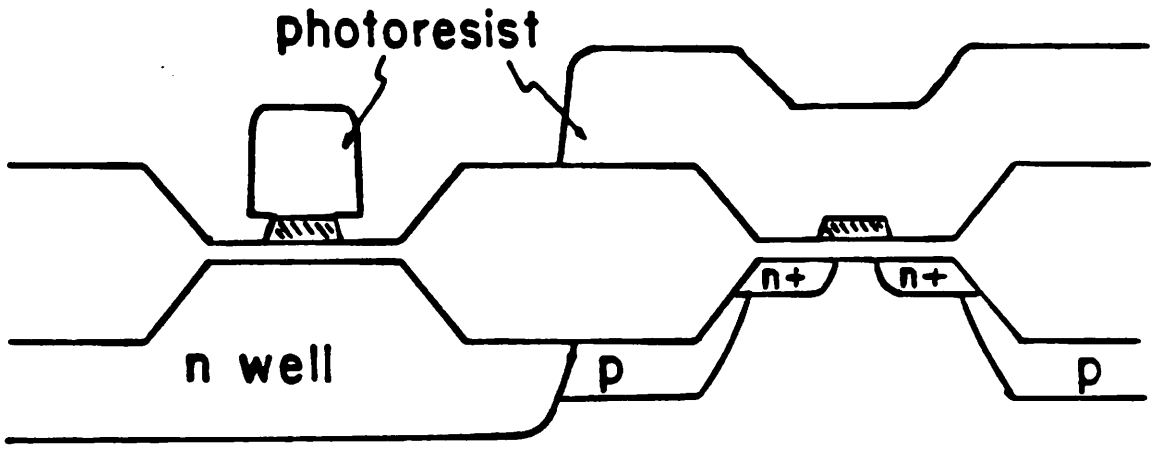


Fig. A5 Wafer cross-section after etching PMOS polysilicon (Step 21).

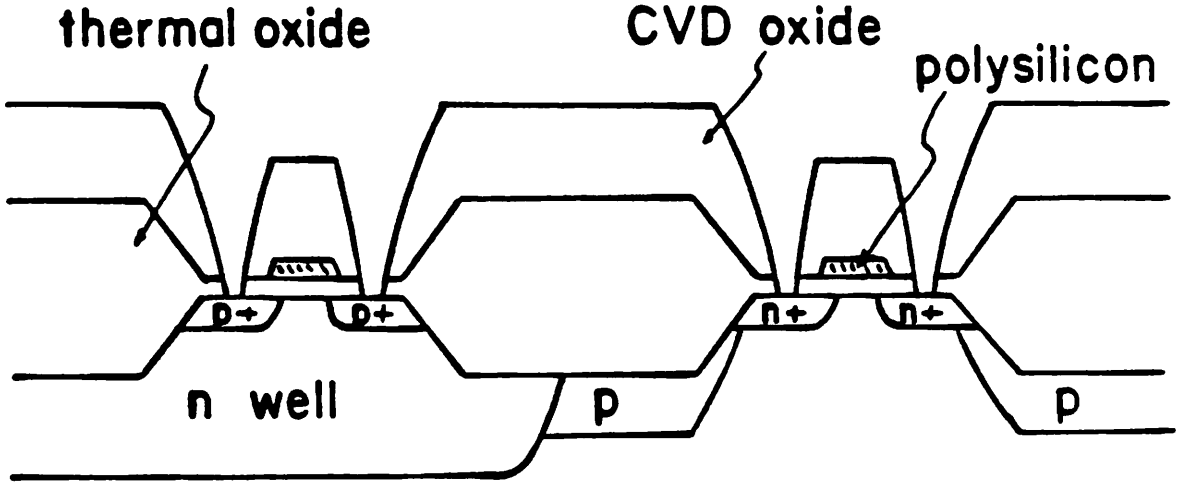


Fig. A6 Wafer cross-section after etching contact holes (Step 24).

plasma etching. The photoresist is left behind after this step to mask the NMOS active regions as well as the PMOS polysilicon gates against boron implantation. The wafer profile after this step is shown in Fig. A5.

23. P+ Source/Drain Implantation

The p+ source/drain implant is a heavy implant and if photoresist is used to mask against it, there is a possibility that it might crack under the heavy ion bombardment. To prevent it from cracking, the photoresist is hardbaked (in addition to the regular postbake at 110°C) in nitrogen plasma at 120°C. The implantation energy used is 60 KeV, with a dose of 2×10^{15} cm^{-2}. The photoresist is then removed afterwards.

24. Oxide Deposition

A thick CVD oxide is deposited for isolation between the polysilicon and metal runs. First, a 0.25 μm of undoped oxide is deposited, followed by a 0.75 μm of phosphosilicate glass (PSG) with 7% phosphorus content by weight. This oxide is densified and reflowed at 1050°C for 20 minutes in nitrogen ambient. The purpose of the undoped oxide underneath the PSG is to mask against phosphorus diffusion into the silicon during the reflow step.

25. MASK #7: Define Contact Holes

This is done in buffered HF with an etch-bake-etch technique. The oxide is first partially etched, then the photoresist is baked again to reflow the edge regions to prevent excessive undercut. This is required because the phosphorus doped CVD oxide etches 10 to 20 times faster than thermal oxide, and can cause the contact holes to bloom excessively. The remaining photoresist is removed after this step and the wafer cross-section is shown in Fig. A6.

26. Metalization

A palladium silicide technology is used for contacting the shallow junctions. First, palladium is evaporated onto the wafer and sintered in forming gas at 300° C for 5 minutes to form palladium silicide in the contact holes. This silicide acts as a barrier metal between aluminum and silicon to prevent aluminum from spiking through the shallow junctions. The excess palladium is etched away in gold etchant. The wafer is then cleaned and aluminum is evaporated on top of the silicide.

27. MASK #8: Define Metal Runs

This mask defines the metal layer and aluminum etchant is used to etch the metal away.

28. Backside Aluminum Evaporation

The remaining photoresist is removed and a new layer is spun on the top surface to protect the structure when removing the backside oxide. This backside oxide is etched in buffered HF. The protecting photoresist is then removed and aluminum is evaporated on the backside. The final wafer cross section is shown in Fig. A7.

29. Alloy

The metal is sintered in forming gas at 350° C for 20 minutes to form a good contact.

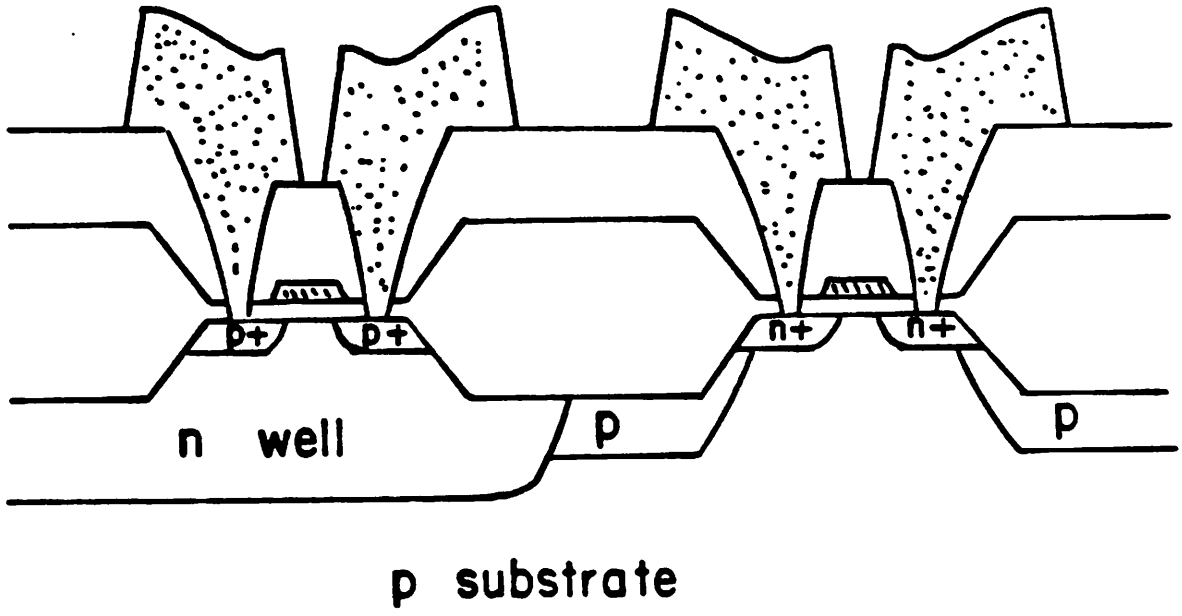


Fig. A7 Final wafer cross-section.

Appendix B:
CMOS Process Design Rules

In establishing the design rules, we assume a direct alignment tolerance of $\pm 1\ \mu m$, and an indirect alignment tolerance of $\pm 2\ \mu m$. Table B1 lists the alignment references for different mask levels. For example, the NMOS POLY mask (#5) is aligned to the NMOS ACTIVE mask (#2), so that the alignment tolerance between them is $\pm\ 1\mu m$. However, the alignment tolerance between the NMOS POLY mask and the N-WELL mask (#1) will be $\pm\ 2\mu m$.

To prevent CMOS latch up, n+ guard rings are required inside the n-wells. We further assume that the n-well lateral diffusion is 3 μm, metal and polysilicon undercuts are 0.5 μm, lateral diffusion for n+ and p+ junctions are 0.25 μm and 0.5 μm respectively, bird's beak from local oxidation is 1 μm, and undercuts in contact holes are 1 μm. From this, we come up with a set of design rules listed in Table B2 and illustrated in Fig. B1-3.

The design rules are established with certain safety margin in addition to the above assumptions. For example, the alignment error between the CONTACT mask and the NMOS POLY mask is 1 μm. The contact hole undercut is 1 μm, and allowing an extra 1 μm as safety margin, we come up with a minimum gate to contact distance of 3 μm. Actually, the polysilicon undercut is 0.5 μm so that we have an additional 0.5 μm safety margin. However, it will be fatal to short out the gate and the diffusion, so that we have allowed more room for error. On the other hand, there is no additional safety margin for non-critical cases. For example, the alignment error between the METAL and CONTACT masks is 1 μm, and contact hole undercut is 1 μm, we come up with a minimum 2 μm metal overlap of contact holes. Even if the metal does not entirely cover up the contact, the exposed regions will not be fatal so that no additional safety margin is required. Most of the rest of the design rules are self-explanatory in a

mask #	name	aligned to
1	N-WELL	
2	NMOS ACTIVE	1
3	PMOS ACTIVE	2
4	CAPACITOR	2
5	NMOS POLY	2
6	PMOS POLY	5
7	CONTACTS	5
8	METAL	7

Table B1 Alignment references for different mask levels

MASK #1	N-WELL (define N-WELL regions)	
1	well to well distance	12 μm
MASK #2	NMOS ACTIVE (define NMOS thin oxide regions, n+ diffusion, and P-ACTIVE shields)	
2	n+ diffusion width	5 μm
3	n+ diffusion to n+ diffusion spacing	5 μm
4	n+ diffusion to well distance	12 μm
5	P-ACTIVE shield overlap well	4 μm
MASK #3	PMOS ACTIVE (define PMOS thin oxide regions, p+ diffusion, n+ guard rings and P-ACTIVE shields)	
6	p+ diffusion width	5 μm
7	p+ diffusion to p+ diffusion spacing	6 μm
8	n+ guard ring width	5 μm
9	n+ guard ring to well edge	0 μm
10	p+ diffusion to n+ guard ring distance	6 μm
11	N-ACTIVE shield overlaps NMOS active regions	4 μm
12	N-ACTIVE shield to P-ACTIVE shield distance	4 μm
MASK #4	CAPACITOR (define capacitor implant regions)	
13	capacitor mask overlaps capacitor thin oxide regions (defined by MASK #2)	4 μm
MASK #5	NMOS POLY (define NMOS gate, poly interconnect and P-POLY shields)	
14	NMOS poly gate and interconnect width	4 μm
15	NMOS poly to NMOS poly spacing	4 μm
16	NMOS poly gate overlap field	3 μm
17	CAPACITOR mask (#4) overlap capacitor poly regions	4 μm
18	NMOS poly gate to edge of capacitor mask	4 μm
19	P-POLY shield overlap p+ active regions	2 μm
MASK #6	PMOS POLY (define PMOS gate, poly interconnect and N-POLY shields)	
20	PMOS poly gate width	6 μm
21	PMOS interconnect width	4 μm
22	PMOS poly to PMOS poly spacing	4 μm
23	PMOS poly gate overlap field	3 μm
24	N-POLY shield overlap n+ active regions and n+ guard rings	2 μm
25	P-POLY shield to N-POLY shield distance	2 μm
MASK #7	CONTACTS	
26	contact cut on diffusion	4 μm × 4 μm
27	contact cut on polysilicon	3 μm × 3 μm
28	diffusion overlap contact	3 μm
29	polysilicon overlap contact	4 μm
30	gate to contact distance	3 μm
MASK #8	METAL	
31	metal width	6 μm
32	metal to metal spacing	4 μm
33	metal overlap contact	2 μm

Table B2 Summary of design rules (min. draw dimensions)

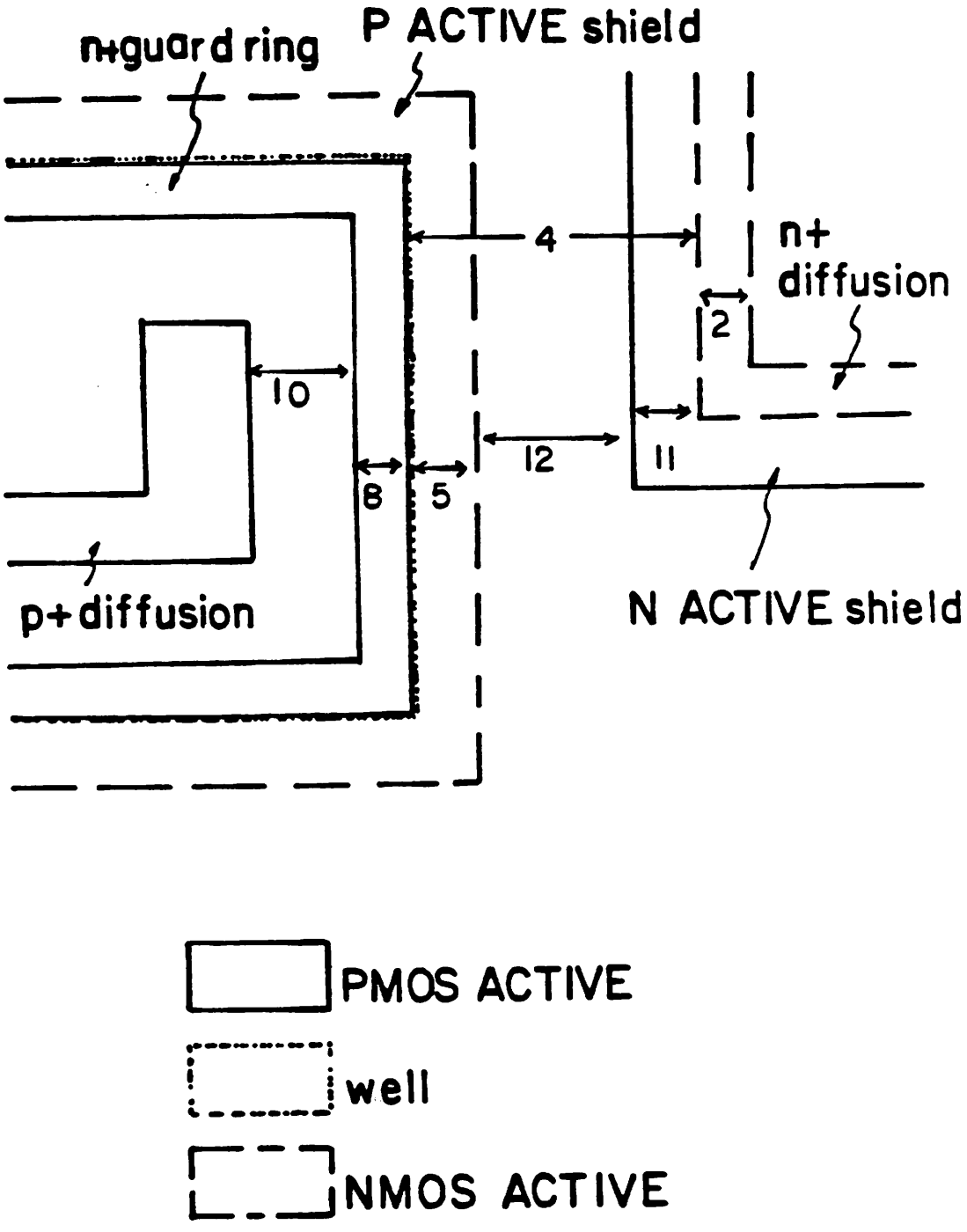


Fig. B1 Design rules for ACTIVE shields.

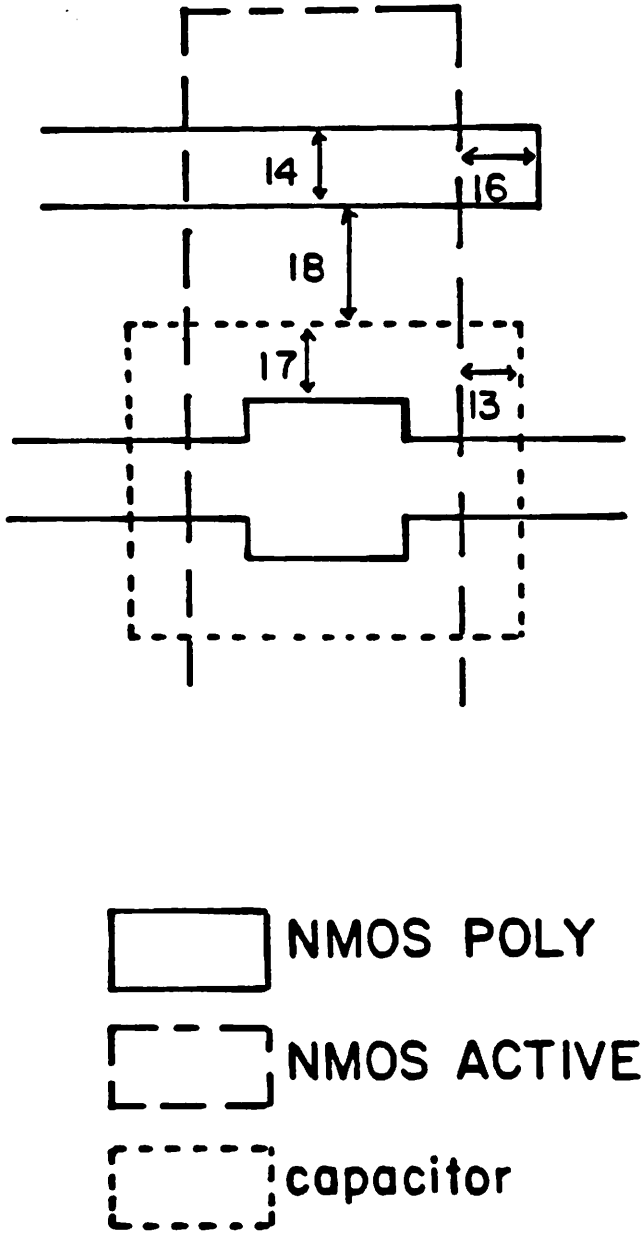


Fig. B2 **Design rules for polysilicon to substrate capacitors.**

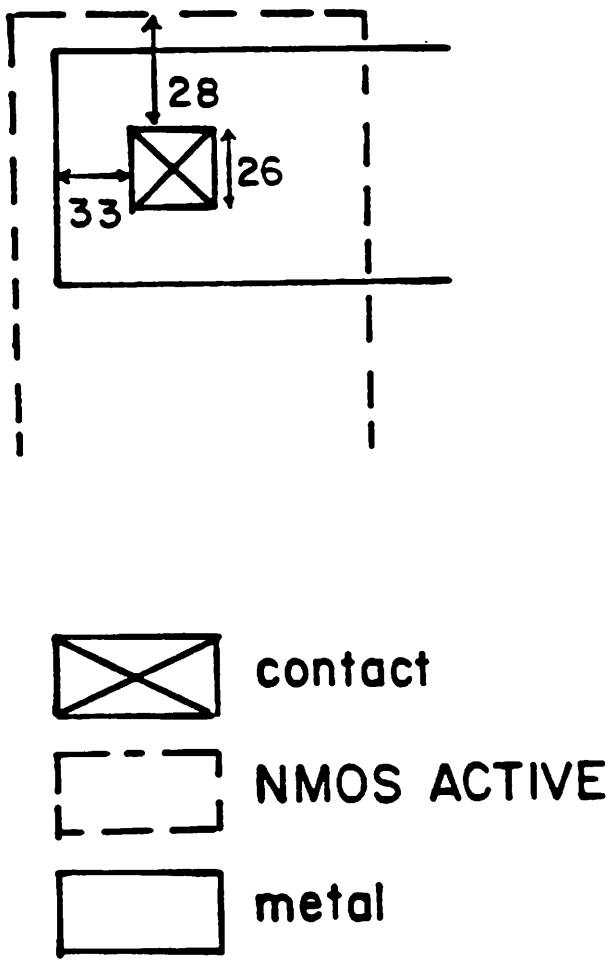


Fig. B3 Design rules for contact holes.

fashion similar to the above reasonings. However, there are certain rules which concern the shields that need further explanation.

There are two types of shields, the "ACTIVE" shields and the "POLY" shields. The P-ACTIVE shields, defined by the NMOS ACTIVE mask (#2), are used to shield the wells from the p-field implant. It has to overlap the well regions by 4 μm. The N-ACTIVE shields, defined by the PMOS ACTIVE mask (#3), are used to protect the nitride of the NMOS thin oxide (or active) regions and have to overlap them by 4 μm. In addition, we need a 4 μm space between the N-ACTIVE and P-ACTIVE shields to prevent leaving behind a strip of unused nitride. These design rules are illustrated in Fig. B1. These rules have plenty of room for error, but since we have to keep a 12 μm distance between the n+ diffusion and the well anyway, they do not take up extra space.

The P-POLY shields, defined by the NMOS POLY mask (#5), are used to shield the p+ diffusions from the heavy arsenic implant as well as to prevent the PMOS polysilicon runs from being etched away. Likewise, the N-POLY shields, defined by the PMOS POLY mask (#6), are used to shield the n+ diffusions and the n+ guard rings from the heavy boron implant as well as to prevent the NMOS polysilicon runs from being etched away. Therefore, both POLY shields have to overlap the regions that they protect by 2 μm. Also, as in the case for the ACTIVE shields, we need a 2 μm spacing between the POLY shields to prevent leaving behind a strip of floating polysilicon. The POLY shields are very similar to the ACTIVE shields and can be similarly illustrated by Fig. B1.

References

[1] R. W. Brodersen, P. R. Gray, and D. A. Hodges, "MOS switched-capacitor filters," **Proc. IEEE**, vol. 67, pp. 61-74, Jan. 1979.

[2] P. R. Gray, R. W. Brodersen, D. A. Hodges, T. C. Choi, R. Kaneshiro, and K. C. Hsieh, "Some practical aspects of switched-capacitor filter design," **Proc. IEEE International Symp. Circuits and Systems**, pp. 419-422, Apr. 1981.

[3] P. R. Gray, D. Senderowicz, H. Ohara, and B. M. Warren, "A single-chip NMOS dual channel filter for PCM telephony applications," **IEEE J. Solid State Circuits**, vol. SC-14, pp. 981-991, Dec. 1979.

[4] R. Gregorian, and W. E. Nicholson, Jr., "CMOS switched-capacitor filters for a PCM voice CODEC," **IEEE J. Solid State Circuits**, vol. SC-14, pp. 970-980, Dec. 1979.

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

[6] I. A. Young, "A low-power NMOS transmit/receive IC filter for PCM telephony," **IEEE J. Solid State Circuits**, vol. SC-15, pp. 997-1005, Dec. 1980.

[7] L. T. Lin, H. F. Tseng, D. B. Cox, R. G. Runge, and D. P. Conrad, "A monolithic audio spectrum analyzer for speech recognition systems," **Digest of Technical Papers, International Solid State Circuits Conference**, pp. 272-273, Feb. 1982.

[8] G. M. Jacobs, D. J. Allstot, R. W. Brodersen, and P. R. Gray, "Design techniques for MOS switched-capacitor ladder filters," **IEEE Trans. Circuits and Systems**, vol. CAS-25, pp. 1014-1021, Dec. 1979.

[9] T. C. Choi, and R. W. Brodersen, "Considerations for high frequency switched-capacitor ladder filters," **IEEE Trans. Circuits and Systems**, vol. CAS-27, pp. 545-552, June 1980.

[10] K. C. Hsieh, P. R. Gray, D. Senderowicz, and D. G. Messerschmidt, "A low-noise chopper-stabilized differential switched-capacitor filtering technique," **IEEE J. Solid State Circuits**, vol. SC-16, pp. 708-715, Dec. 1981.

[11] T. C. Choi, R. T. Kaneshiro, R. W. Brodersen, and P. R. Gray, "High frequency CMOS switched-capacitor filters for communications applications," to be published in **Digest of Technical Papers, International Solid State Circuits Conference**, Feb. 1983.

[12] D. A. Hodges, P. R. Gray, and R. W. Brodersen, "Potential of MOS technologies for analog integrated circuits," **IEEE J. Solid State Circuits**, vol. SC-13, pp. 285-294, June 1978.

[13] L. T. Lin, H. F. Tseng, and L. Querry, "Monolithic filters for 1200 baud modems," **Digest of Technical Papers, International Solid State Circuits Conference**, pp. 148-149, Feb. 1982.

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

[15] J. J. Friend, C. A. Harris, and D. Hilberman, "STAR: an active biquadratic filter section," **IEEE Trans. Circuits and Systems**, vol. CAS-22, pp. 115-121, Feb. 1975.

[16] 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, Dec. 1975.

[17] L. T. Bruton, "Low sensitivity digital ladder filters," **IEEE Trans. Circuits and Systems**, vol. CAS-22, pp. 168-176, Mar. 1975.

[18] D. J. Allstot, R. W. Brodersen, and P. R. Gray, "MOS switched-capacitor ladder filters," **IEEE J. Solid State Circuits**, vol. SC-13, pp. 806-814, Dec. 1978.

[19] L. E. Franks, and I. W. Sandberg, "An alternative approach to the realization of network transfer functions: the N-path filter," **Bell System Technical Journal**, pp. 1321-1350, Sept. 1960.

[20] A. Fettweis, and H. Wupper, "A solution to the balancing problem in N-path filters," **IEEE Trans. Circuit Theory**, vol. CT-18, pp. 403-405, May 1971.

[21] M. S. Lee, and C. Chang, "Exact synthesis of N-path switched-capacitor filters," **Proc. IEEE International Symp. Circuits and Systems**, pp. 166-169, Apr. 1981.

[22] M. B. Ghaderi, G. C. Temes, and J. A. Nossek, "Switched-capacitor pseudo N-path filters," **Proc. IEEE International Symp. Circuits and Systems**, pp. 519-522, Apr. 1981.

[23] A. V. Oppenheim, and R. W. Schafer, **Digital Signal Processing**, New York: Prentice-Hall, 1975.

[24] H. J. Orchard, "Inductorless filters," **Electronic Letters**, vol. 2, pp. 224-225, June 1966.

[25] G. M. Jacobs, "Practical design considerations for MOS switched-capacitor ladder filters," **U. C. Berkeley Electronics Research Lab. Memorandum**, no. UCB/ERL M77/69, Nov. 1977.

[26] R. T. Kaneshiro, "Circuit and technology considerations for high frequency switched-capacitor filters," **Ph. D. Dissertation, Univ. of California, Berkeley**, June 1983.

[27] R. W. Daniels, **Approximation Methods for Electronic Filter Design**, New York: McGraw-Hill, 1974.

[28] G. Szentirmai, "FILSYN - a general purpose filter synthesis program," **Proc. IEEE**, vol. 65, pp. 1443-1458, Oct. 1977.

[29] S. E. Belter, and S. C. Bass, "DINAP - a digital network analysis program," School of Electrical Engineering, Purdue University.

[30] A. S. Grove, **Physics and Technology of Semiconductor Devices**, Wiley, 1967.

[31] J. J. Barnes, K. Shimohigashi, and R. W. Dutton, "Short channel MOSFET's in the punchthrough current mode," **IEEE Trans. Electron Devices**, vol. ED-26, pp. 446-453, Apr. 1979.

[32] R. R. Troutman, "VLSI limitations from drain-induced-barrier-lowering," **IEEE Trans. Electron Devices**, vol. ED-26, pp. 461-469, Apr. 1979.

[33] D. B. Estreich, "The physics and modeling of latch-up and CMOS integrated circuits," **Stanford Integrated Circuit Laboratory Technical Report**, no. G-201-9, Nov. 1980.

[34] P. R. Gray, "Basic MOS operational amplifier design - an overview," **Analog MOS Integrated Circuits**, New York: IEEE Press, pp. 25-49, 1980.

[35] K. C. Hsieh, "Noise limitations in switched-capacitor filters," **Ph. D. Dissertation, Univ. of California, Berkeley**, May 1982.

[36] C. L. Hoang, "Evaluation of a fully integrated high frequency switched-capacitor filter," **U. C. Berkeley Electronics Research Lab. Memorandum**, no. UCB/ERL M82/67, Sept. 1982.

[37] R. D. Rung, H. Momose, and Y. Nagakubo, "Deep trench isolated CMOS devices," **Technical Digest, International Electron Devices Meeting**, pp. 237-240, Dec. 1982.

[38] S. Liu, B. Hoefflinger, and D. O. Pederson, "Interactive two-dimensional design of barrier-controlled MOS transistors," IEEE Trans. Electron Devices, vol. ED-27, pp. 1550-1558, Aug. 1980.

[39] L. W. Nagel, "SPICE2, a computer program to simulate semiconductor circuits," U. C. Berkeley Electronics Research Lab. Memorandum, no. ERL-M520, May 1975.

[40] R. T. Kaneshiro, "The Berkeley CMOS process," unpublished memorandum, Sept. 1982.