INTEGRATED SELECTIVE AMPLIFIERS USING FREQUENCY TRANSLATION

by

G. A. Rigby

Memorandum ERL-M168

27 May 1966

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

This memorandum is based on a thesis submitted for the Ph.D. degree at the University of California, Berkeley.

The research reported herein was supported wholly by the Joint Services Electronics Program (U. S. Army, U. S. Navy, U. S. Air Force) under Grant AF-AFOSR-139-65.

ABSTRACT

Frequency selective amplifiers with high selectivity and good desensitivity may be realized in silicon integrated circuit form by making use of the frequency translation principle. The amplifiers thus realized have a number of desirable properties which distinguish them from selective amplifiers based on other active-RC techniques. In particular, the gain, bandwidth and center frequency are virtually independent and consequently, the passband of the amplifier may be made very narrow without impairing its sensitivity.

The system consists of two identical channels, each of which contain two analog multipliers, to perform the frequency translation, and a lowpass filter whose passband is translated up to the center frequency of the system. The outputs of the two channels are summed in an amplifier to form the required output, and a local oscillator, whose frequency becomes the center frequency of the system, provides the signal needed for the frequency translation. Previously reported realizations of this system used periodic switches, instead of analog multipliers, to perform frequency translation, but such realizations suffer from a number of disadvantages with respect to frequency range and noise performance. The realizations reported here have been made possible by the development of an integrated bipolar transistor analog multiplier with a wide frequency range of operation and well-defined transfer

characteristics. The characterization of this multiplier is described in detail.

The design of and performance of the selective amplifier in integrated form is developed on the basis of the characteristics of the individual functional blocks and particular attention is paid to the sensitivity aspects of its performance. It is shown that the use of an overall negative-feedback loop results in the reduction of the temperature sensitivity of most performance parameters, and a method is described for compensating balance-drift of the analog multipliers. A very low drift sensitivity is thereby obtained. In cases where the long time-constants of the lowpass stages are beyond the range of simple monolithic resistance-capacitance combinations, alternative realizations using hybrid construction or monolithic multiplied-capacitance techniques are described.

Three prototype amplifiers, using integrated functional blocks have been constructed and their measured performances support the theoretical predictions. One of the realizations, using a negative feedback loop, achieves a stable Q of 600 at a 1.6 Mc/s center frequency.

ACKNOWLEDGMENTS

The writer wishes to thank his colleagues for many useful discussions during the course of this work. In particular, he acknowledges the supervision, advice, and constant encouragement from Professor D. O. Pederson, and the technical assistance in the laboratory given by Mrs. D. McDaniel and Mrs. M. L. Conragan.

CONTENTS.

CHAPTER 1. INTRODUCTION

1.1 The integrated selective amplifier problem.

The problem of realizing stable, highly selective bandpass characteristics in electronic circuits, without using inductors or electromechanical elements is a familiar one and many possible solutions have been suggested. Although there are many instances in discrete-component circuitry where RC selective amplifiers are useful, the challenge, in recent years, of realizing a wide variety of electronic functions in semiconductor integrated circuits has motivated increased activity in this area. Numerous solutions have been proposed, which are both compatible with the limitations of integrated circuits and also exploit their virtues; and although none is attractive in all respects, these proposals represent significant progress.

There are three categories into which selective amplifier schemes may be divided; namely,

(a) linear, active RC circuits,

(b) active RC circuits incorporating one or more nonlinear elements*, and

(c) electromechanical systems.

---

* For the purposes of this classification, the analog multiplier is regarded as a nonlinear element, although it may perform "linear multiplication".

Class (b) contains the frequency translation systems which are the main subject of this thesis and which will be introduced in Chap. 2. Classes (a) and (c) are considered briefly here and the reader is also referred to the broad survey presented by Newell[1], with the caution that this field is changing quite rapidly.

1.2 Linear active RC circuits.

The largest class of integrated selective amplifiers is, in terms of research activity, the linear active RC type. In treating these circuits it is convenient to make a distinction between those which contain one or more explicit feedback paths, and those which contain elements such as the negative-immittance-convertor[2](NIC). It is possible, though not always convenient, to model the implicit feedback mechanisms of the second type in such a way as to conform with those of the first. But regardless of the circuit model used, these circuits may be treated together in terms of a complex frequency-plane display of their Laplace-transformed transfer characteristics, i.e. the s-plane. In this way, most of the essential properties of such circuits are made obvious.

The main goal in a selective amplifier design can be stated, in terms of the s-plane, as the realization of a system transfer function containing pairs of complex conjugate poles, which lie close to the imaginary axis in the left half plane and whose positions in the plane are stable with respect

to changes in all sensitive parameters of the circuit. Passive RC networks have transfer functions containing real negative poles and may have real or complex conjugate zeros. Combinations of these networks with linear active elements and feedback loops are then used to generate transfer functions with the required complex-conjugate poles. Of the two aspects of this problem - that of obtaining poles in the desired positions and that of stabilizing their positions - the latter is generally the more difficult.

The block diagrams of Fig. 1.1 (a) and (b) show two typical linear active RC systems: a single-loop feedback type and a negative-immittance-converter type. In the first, the forward path consists of two active, frequency-independent gain blocks A_1 and A_2 with a passive RC network N_1. The feedback path contains a second RC network N_2. In the second circuit, an NIC is placed between two passive RC two-ports Y_1 and Y_2. The interaction between these networks through the NIC may be arranged to provide complex conjugate poles.

We now consider the four typical root-locus plots presented in Fig. 1.2. These represent the variation of transfer function pole positions for a given circuit as a suitable gain parameter is varied. In general, they may be produced by either of the two schemes given above.

The first example, (a), is typical of a potentially unstable circuit, for example, an RC oscillator adjusted below the threshold of oscillation. Since the complex pole pair may be set arbitrarily close to the imaginary axis, very

high selectivity is possible. However, the pole positions in such a case are usually very sensitive to active and passive device parameter changes. Case (b) is a special case representing a transition between potentially unstable and unconditionally stable systems. The presence of any excess phase-shift in such a circuit will distort the locus from the idealized straight line shown and, depending on its sign, will raise or lower the selectivity. This, again is a very sensitive arrangement.

Cases (c) and (d) provide similar control over pole positions, but also enable the sensitivity to be reduced. The zeros shown are complex "phantom" zeros which are introduced in the feedback path of a circuit, e.g. N_2 of Fig. 1.1(a), but do not appear in the closed-loop transfer function.[3] These zeros stabilize the position of the complex pole pairs for large loop gains and reduce their dependence on active device parameters. The (c) case suffers from two disadvantages, however. For high selectivity applications, the realization of complex zeros with small real parts requires careful adjustment of the network which produces them and their positions are, of course, sensitive to passive parameter values.[4,5] In addition, the large loop-gains required for high selectivity place heavy demands on the forward-path amplifier. Finally, case (d) provides a method for the simultaneous optimization of selectivity and sensitivity. This property is obtained by incorporating into a system sufficiently many degrees of freedom, in the form of extra feedback and

feedforward loops.[6] A portion of the root-locus lies closer to the imaginary axis than the complex conjugate zeros and the closed-loop poles are placed at points in this region where the first-order sensitivity of their positions is zero. Sensitivity results reported for this technique [6,7] are, at present, considerable better than those obtained with other RC schemes.

To summarize these comments, the characteristics of the linear active RC circuits are that they are relatively simple in terms of component requirements and may, in most cases, be realized in monolithic integrated form. It is sometimes advantageous, however, to resort to hybrid construction so that precision thin-film resistors and capacitors may be used for critical components. The tuning of these circuits is difficult because their gain, bandwidth and center frequency, as well as their respective sensitivities, are usually inter-related. The realization of Q-values above 50 with good desensitivity has been shown by many workers to be very difficult, but the limitations existing at very high center frequencies have not been thoroughly investigated.

1.3 Electromechanical systems.

The use of electromechanical resonant elements in selective amplifiers precludes conventional monolithic construction, but raises a number of attractive possibilities. For example, studies of the "sandwich" structure, in which a resonant quartz plate is mounted on a stack of quartz wafers, each one quarter

wavelength thick at the resonant frequency[1], have shown that this structure enables filters to be rigidly attached to substrates without introducing significant damping. Another approach, which is based on the principle of energy trapping, makes use of the surface-loading produced by evaporated metal electrodes on a quartz wafer. These confine resonant modes to the regions of the electrodes, thereby enabling a number of filter elements to be formed in a single wafer.[1,8] Piezoelectric elements such as these are not useful at very low frequencies, however, because the elastic wavelengths require physical dimensions to be much larger than those normally encountered in microcircuits.

Another device which is currently being studied is the resonant gate field-effect transistor.[9,10] This is basically a metal-oxide-silicon transistor (MOST) without the conventional gate electrode. The gate in this case is a beam of tungsten or electro-formed gold which is cantilever-mounted at one or both ends so that it stands a small distance above the oxide surface. A suitable signal applied between the gate and channel excites a resonance in the beam by electrostatic force. The motion of the beam, in turn, modulates the channel conductivity by charge induction, thus providing a feedback path. At present these devices require large polarizing voltages and exhibit rather poor selectivity, but an improvement in performance may be expected.

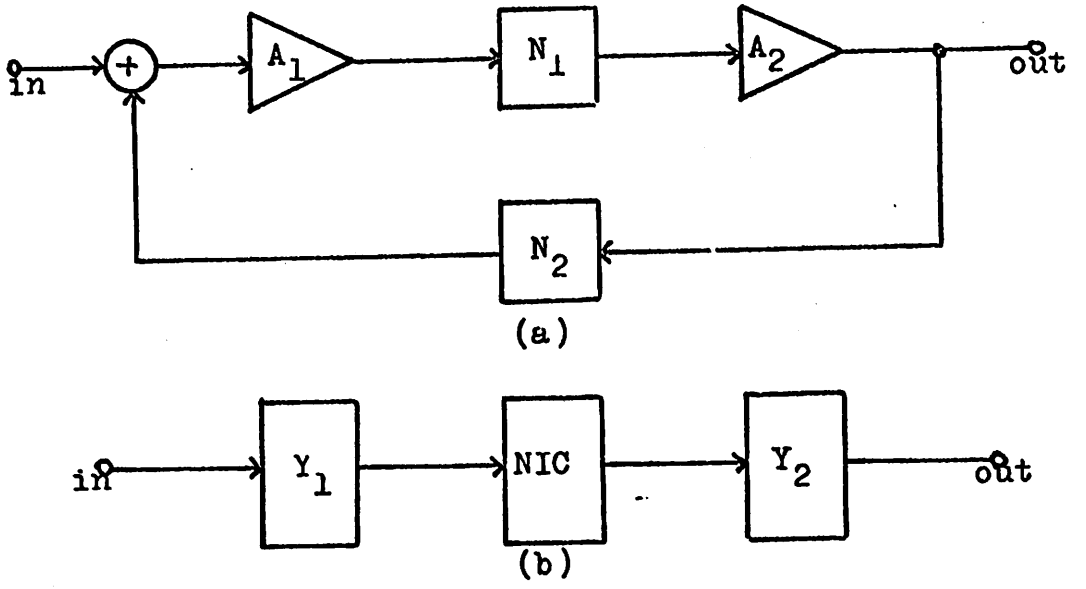


Figure 1.1 Linear active RC system examples:
(a) single-loop feedback, (b) negative immitance converter

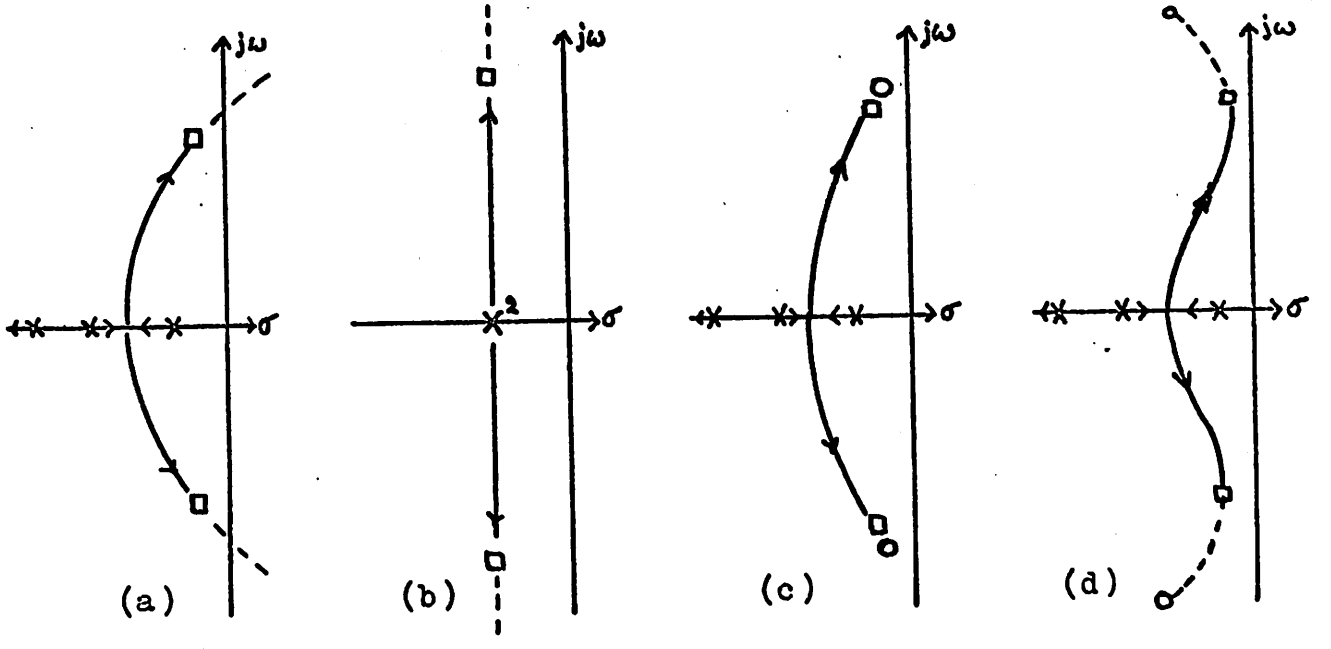


Figure 1.2 Typical root-locus plots for linear active RC systems.

CHAPTER 2. THE FREQUENCY TRANSLATION FILTER

2.1 Principle of operation.

Frequency translation is a very fundamental and long-established technique in communication systems, and the historical development of ideas which led to the superheterodyne receiver and the synchronous detector [11] is also responsible for the technique to be described here. This technique differs from the superheterodyne and synchronous detector in some respects, however. The "image-frequency" phenomenon is absent, in comparison with the former, and in the second case the necessity for maintaining a phase relation between the input frequency and local oscillator signal - a characteristic of the synchronous detector - is absent in the frequency translation filter.

A two-channel frequency translation filter was described by Barber [12] in 1947 and system considered here is the same in principle. Early proposals such as this one, however, suggested no efficient way in which the frequency translations could be performed, and it is therefore a primary purpose of this thesis to consider this question. During recent years, the theory of time-varying networks has seen considerable development and a particular member of this class, the periodic

switch [13], is of interest here because it provides one of the possible ways of performing the required type of frequency translation.[14] In 1960, Franks and Sandberg[15] presented a theory of the general frequency translation filter shown schematically in Fig. 2.1. This consists of N identical channels, each of which contains two ideal analog multipliers M, to perform the frequency translation, and a real, time-invariant, linear network with a time-domain impulse response h(t). The multipliers are driven at their second inputs by periodic signals p(t) and q(t) as shown. The period of these signals is $T = 2\pi/\omega_o$, where ω_o becomes the center frequency of the filter passband. The p and q inputs are shifted in phase between adjacent channels by an amount $\tau = T/N$.

In general, p and q are not sinusoidal and it has been shown [15] that if they are decomposed into complex Fourier components,

$$p(t) = \sum_{m=-\infty}^{+\infty} P_m \exp(j\omega_o mt)$$
$$q(t) = \sum_{n=-\infty}^{+\infty} Q_n \exp(j\omega_o nt) \qquad (2.1)$$

the system transfer function F(s) is given by

$$F(s) = N \sum_{n=-\infty}^{\infty} P_{-n} Q_n H(s - jr\omega_o) \qquad (2.2)$$

where $H(s) = \mathcal{L}[h(t)]$ and s is the complex frequency variable.

From (2.2) we see that the system transfer function consists of a spectral array of images of the basic transfer function H(s), centered on those harmonics of ω_o for which

$P_{-n}Q_n$ is non-zero. This is illustrated by Fig. 2.2, in which $|F(j\omega)|$ is plotted up to the fourth harmonic of ω_o, for a given $H(j\omega)$. The particular case of this general system which concerns us here is the case where H(s) is a lowpass function and the input and output of the system are band-limited so that a simple bandpass characteristic $F(s) = H(s-j\omega_o) + H(s+j\omega_o)$ is obtained. In this form, the system has been studied by a number of workers who have used periodically switched gates to approximate the multiplying process.[15-18].

These periodic switching realizations suffer from a number of practical disadvantages, although they satisfy the theoretical requirements as outlined above. In the first place, the finite switching speed of the gates has been found to limit the center frequency of the system to values of 100 kc/s or less. Furthermore, the transients produced by these gates contribute significantly to noise in the system. A minimum of three channels is necessary to realize the required bandpass function and the circuitry needed for the generation of the appropriate gate-drive waveforms is extensive. Since periodic gating is equivalent to analog multiplication with a pulse train, we see that this technique produces images of the passband at harmonics of the desired center frequency. (See Fig. 2.2.) It is therefore necessary to incorporate extensive filtering circuitry.

The frequency translation filter could be synthesized in a more direct way if the driving functions p and q were pure sinusoids and the analog multipliers could be realized in practice. In order to do this, an integrated analog multiplier

has been developed and is described in Chap. 3. The system which has thus been realized is the priciple subject of this thesis. It will be seen that the behavior of this system conforms more closely with the theoretical model described above and greatly reduces the problems associated with the periodic switching realizations. At the same time, the characteristic advantages of the frequency translation filter are retained; namely, the virtual independence of gain, bandwidth and center frequency. This independence enables stable, highly selective bandpass systems to be realized in integrated circuit form.

2.2 A prototype selective amplifier.

The system to be described here is shown schematically in Fig. 2.3 and is a special case of the general filter of Fig. 2.1. Note that only two channels are used, in contrast to the minimum of three for the switching version, and the reason for using two will become apparent from (2.5) below. Each channel consists of two analog multipliers, assumed linear in the initial analysis, and a lowpass filter H with a bandwidth ω_b. A local oscillator provides a sinusoidal signal at a frequency ω_o. which is separated into two quadrature components by an RC phase-shift network and applied to the multipliers as shown. The two channels have identical transfer functions in the ideal case. The outputs of the two channels are summed at the input of an amplifier A which incorporates a simple highpass filter to attenuate undesired low-frequency components and a simple

notch filter to attenuate the small amount of second harmonic distortion produced, in practice, by the output multipliers.

Each of the analog multipliers is characterized by two transfer coefficients P and Q which are defined as follows. If the two inputs to a multiplier are x and y, the output is defined as $2Pxy + Qx$ so that P is the transfer coefficient for the product term and Q gives the magnitude of the direct transfer of input x. In an ideal multiplier Q is zero, but in the type of multiplier to be used it is non-zero. Expressions for P and Q in terms of circuit parameters are given in (3.15) and (3.26). In the following, these coefficients are given subscripts to identify them with a given multiplier, e.g. P_{a1}, Q_{a1} are the coefficients of multiplier M_{a1}.

To illustrate the operation of the system, we consider a simple input signal $v_i = V_{im} \sin\omega_i t$ and use the signal and phasing notation of Fig. 2.3. The output of M_{a1} is thus found to be:

$$v_{a1} = P_{a1}V_{im}\left[\cos(\omega_o-\omega_i)t - \cos(\omega_o+\omega_i)t\right] + Q_{a1}V_{im}\sin\omega_i t \quad (2.3)$$

If ω_i is close to ω_o and the bandwidth of $H(j\omega)$, ω_b, is much less than ω_o, the high-frequency terms in (2.3) are attenuated to a negligible level, leaving:

$$v_{a2} \cong H_a\left[j(\omega_o-\omega_i)\right] P_{a1}V_{im}\cos(\omega_o-\omega_i)t \quad (2.4)$$

The corresponding signal in channel (b) is of the same form but 90^o out of phase. When these two signals are applied to their respective output multipliers and the outputs are summed, the

resultant signal is:

$$v_3 \simeq V_{im}\left\{P_{a1}P_{a2}H_a\left[j(\omega_o-\omega_i)\right] + P_{b1}P_{b2}H_b\left[j(\omega_o-\omega_i)\right]\right\}\sin\omega_i t$$
$$+V_{im}\left\{P_{a1}P_{a2}H_a\left[j(\omega_o-\omega_i)\right] - P_{b1}P_{b2}H_b\left[j(\omega_o-\omega_i)\right]\right\}\sin(2\omega_o-\omega_i)t$$
$$+ F\left[(\omega_o-\omega_i)t\right] \qquad (2.5)$$

The low-frequency sinusoidal term $F\left[(\omega_o - \omega_i)t\right]$ is removed by the highpass network in the summing amplifier. The coefficient of the second term is zero if the transfer functions of the two channels are exactly matched and under these conditions the system response is given by the first term only.

If (2.5) is now recalculated without assuming that the high-frequency terms in (2.3) are negligible, the following general system transfer function K(s) is obtained:

$$K(s) = 2AP_1P_2\left[H(s-j\omega_o) + H(s+j\omega_o)\right] \qquad (2.6)$$

where K(s) is defined to include the summing amplifier gain A and has been simplified by the assumptions: $P_{a1} = P_{b1} = P_1$, $P_{a2} = P_{b2} = P_2$ and $H_a = H_b = H$. This expression is seen to be analogous to (2.2) for the case $P_{-n}Q_n=0$ for all $n \neq +1,-1$.

If we consider a simple example in which $H(s) = H_o/(sp_1+1)$, K(s) becomes a transfer function with a complex conjugate pole pair at $s = -p_1 \pm j\omega_o$ and a real zero at $s = -p_1$. The response of this system is therefore a conventional second-order bandpass function, with a bandwidth of $2p_1$ and a center frequency ω_o.

In the following chapters the analysis and design of this system are examined in detail, with particular reference to its

realization in integrated circuit form. The individual functional blocks, i.e. analog multipliers, lowpass stages, oscillator phase-shift network and summing amplifier are treated separately. The results obtained are then used to characterize the performance of the system when realized with real, rather than ideal components.

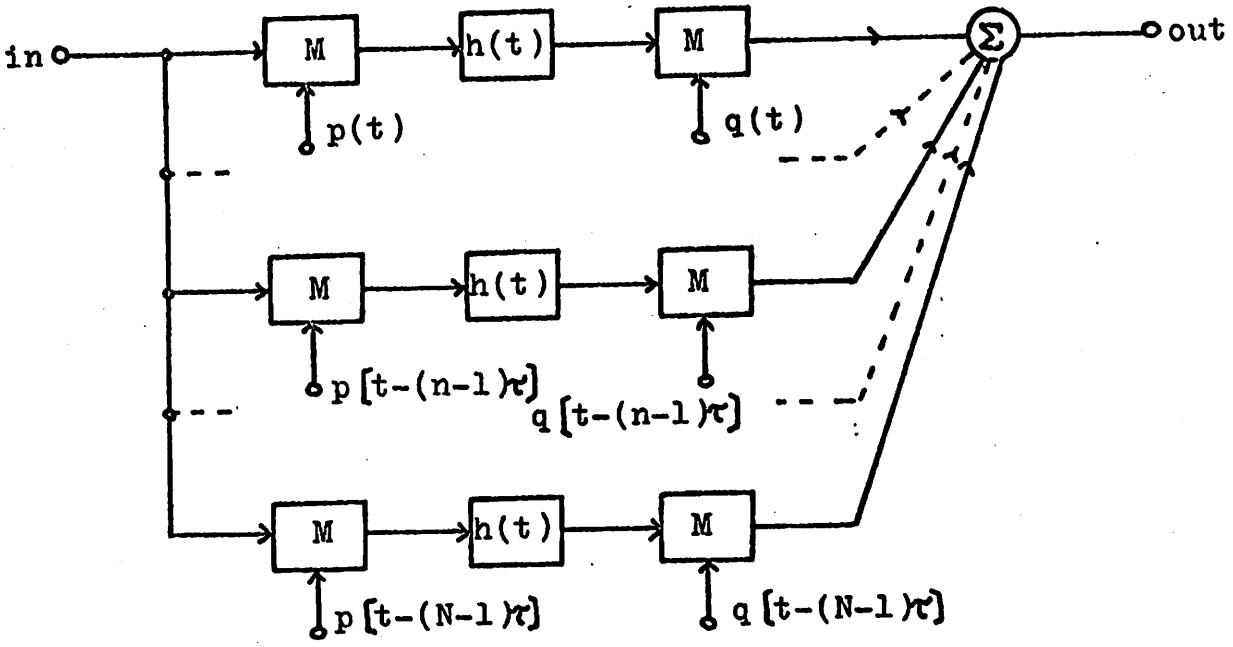


Figure 2.1 General frequency translation filter. (after Franks & Sandberg.)

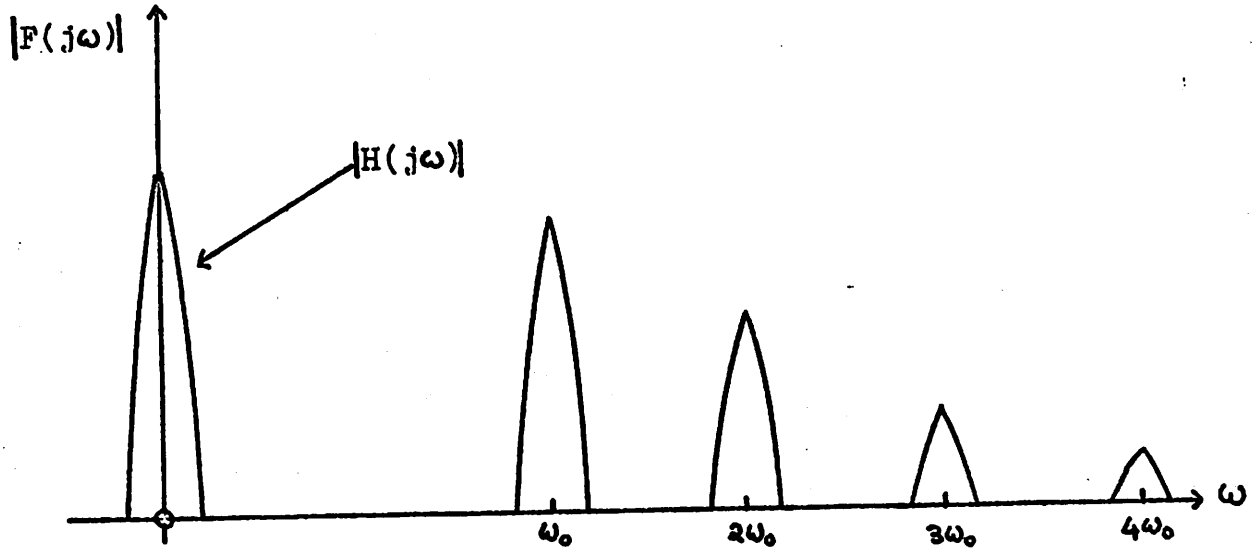


Figure 2.2 Frequency response of the system of Fig. 2.1.

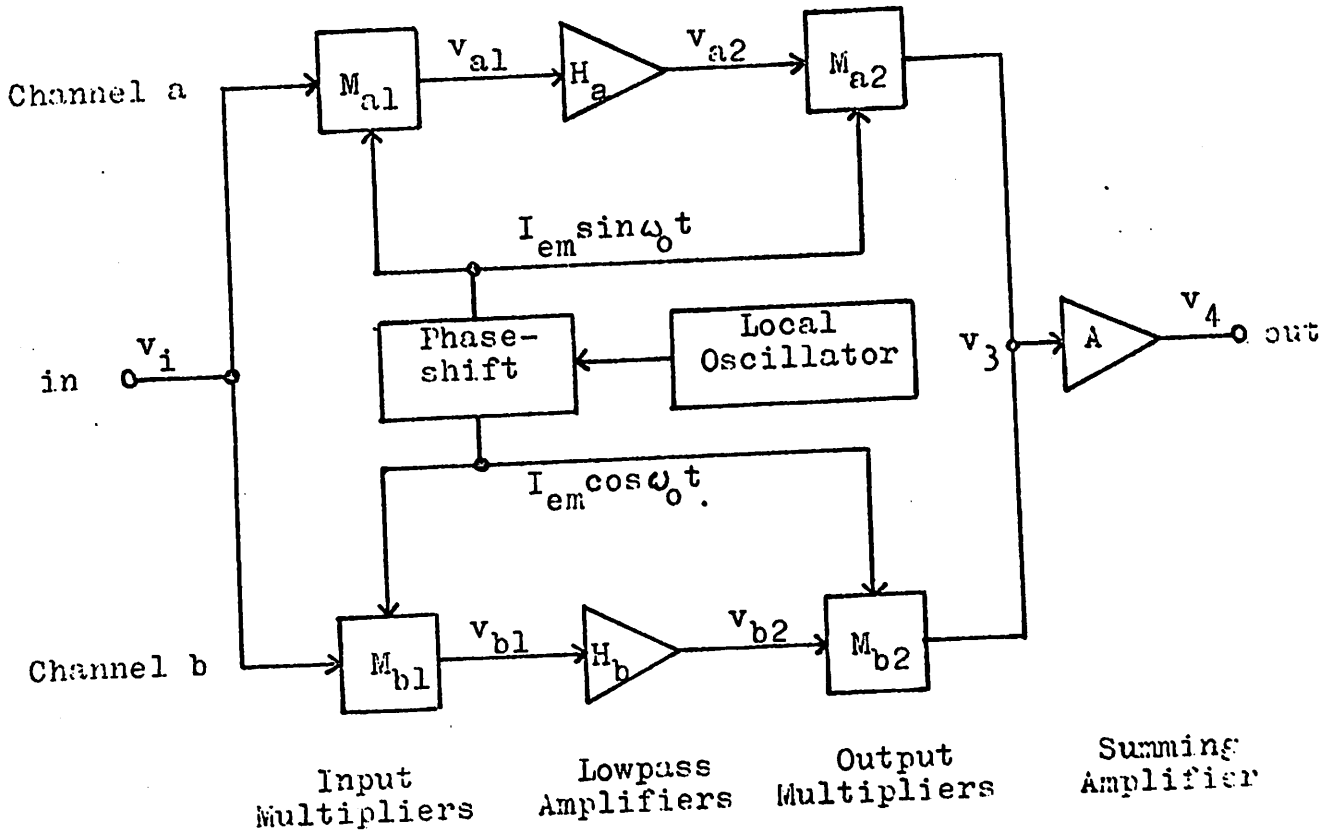


Figure 2.3 Two-channel frequency translation filter.

CHAPTER 3. THE ANALOG MULTIPLIER

3.1 Choice of circuit.

Four analog multipliers are used in the frequency-translation filter and it is desirable that these should be simple in terms of component and fabrication requirements. The choice of a suitable circuit, therefore, depends partly on this factor and also on how closely its operation conforms to the transfer characteristic defined above in Sec. 2.2. That is, if x and y are the two inputs, the output should contain a term k_1xy, where k_1 is constant over a wide range of frequencies and signal levels. This defines linear multiplication. For this application, the presence of additional terms of the form k_2x and k_3y is permissible, since these terms may be separated or cancelled in the system; but terms proportional to x^2 or y^2 are undesirable. Two familiar semiconductor devices have suitable properties for analog multiplication: the field-effect transistor (FET), in which the channel conductance is a function of both gate and drain voltage below pinch-off, and the bipolar transistor (BJT) in which the transconductance is a function of emitter current.* Circuits using

---

* There are other devices such as the four-terminal transistor and the Hall-effect plate which possess the multiplying property, but these are not suitable for this application since they cannot be realized simply in diffused monolithic circuits.

these devices are now examined in the light of the above criteria.

Applications of the FET multiplier have been described in the literature[19,20] in terms of the circuit shown in Fig. 3.1. One input is applied as a differential signal $(V_{G1} - V_{G2})$ between the two gate electrodes and the other as a differential drain voltage $(V_{D1} - V_{D2})$.* The common source current is sampled by the small resistor R_s to provide the output signal V_o. Below pinch-off, the drain current of an FET is approximated by[21]

$$I_D \simeq G(V_P + V_G - V_D/2)V_D \qquad (3.1)$$

where G is a constant parameter, V_P is the pinch-off voltage and the gate and drain voltages are measured with respect to the source. Using this expression and denoting by the subscripts 1 and 2 the parameters of Q_1 and Q_2, we find;

$$V_o/R_s \simeq G_1(V_{G1}-V_{G2})(V_{D1}-V_{D2}) + (V_{D1}-V_{D2})(G_1V_{P1}-G_2V_{P2}) + (V_{D1}-V_{D2})(G_1-G_2)V_{G2} - (V_{D1}-V_{D2})^2(G_1-G_2)/2 \qquad (3.2)$$

In terms of the notation introduced at the beginning of this section, we identify the first term as k_1xy, the sum of the second and third as k_2x, (note that k_2 is dependent on V_{G2}) and the fourth term as the undesired distortion term x^2. A more accurate characterization of the device reveals small

---

* An alternative mode of operation in which the second input is applied as a differential drain current has also been considered, but leads to a similar result to the one derived here.

higher-order distortion terms as well. To minimize the distortion, therefore, the characteristics of the two devices should be closely matched, i.e. $G_1 \simeq G_2$, $V_{P1} \simeq V_{P2}$. It has also been shown [19] that second-order distortion terms are reduced by minimizing the ratio $(V_{G1}+V_{G2})/(V_{P1}+V_{P2})$.

A consideration of these factors, together with the performance data reported in Ref. 19 and the present difficulty of fabricating high-frequency integrated FET pairs with closely matched parameters, has led to the development of a bipolar transistor multiplier for this application. The characteristics of this circuit are described in the following sections and the reasons for preferring it (e.g., low distortion, a well-defined transfer function and operation at high frequencies) will become apparent. It is possible, nevertheless, that the objections to the FET scheme may be overcome in the future.

3.2 The bipolar transistor multiplier: low-frequency characterization.

The bipolar transistor analog multiplier, shown schematically in Fig. 3.2, is seen to resemble closely the conventional differential pair configuration. Both (npn) transistors and their associated resistors are assumed to be fabricated by adjacent diffusions in a single silicon chip. The two input signals are the differential base voltage V_1 and the emitter signal current I_e, while the output is taken as the differential collector voltage V_o. To bias the two transistors in

their linear operating regions, a dc emitter current I_E is also supplied. In this representation, the resistors R_{B1} and R_{B2} have values which include both the source resistance and the internal base resistance of the transistors.

For these conditions of source and load, two gain parameters are defined: the emitter-to-collector current gain a_I^b and the base-to-collector current gain a_I^e for each of the transistors Q_1 and Q_2. It is assumed that these are constant over the operating current range of the circuit. In terms of more familiar parameters, a_I^e and a_I^b are equal, respectively, to the β and α of the transistors for the special case of short-circuit collector load. The frequency dependence of these parameters limits the frequency range of operation of the multiplier and, as will be seen below, the dominant effect is due to a_I^e. Since this section is concerned with low-frequency characterization, it will be assumed that all signal frequencies are such that these parameters take their dc values.

At this point, a subscript convention is defined which is used throughout the remainder of this chapter: all parameters of Q_1 are labelled with a subscript 1 and those of Q_2 with 2. The absence of a subscript implies either a parameter which is not identified with a particular device, or one which is averaged between the two sides.

We now postulate that the I-V characteristics of the emitter-base junctions are given by the simple diode law:

$$I_E = I_{Eo}\exp(qV_{BE}/kT) \qquad (3.3)$$

in which I_{Eo} is a parameter related to the emitter saturation current. From Fig. 3.2 we note that:

$$V_{BE1} - V_{BE2} = V_1 + I_{B2}R_{B2} - I_{B1}R_{B1} \tag{3.4a}$$

$$= V_1 + \frac{I_{C2}R_{B2}}{a^e_{I2}} - \frac{I_{B1}R_{B1}}{a^e_{I1}} \tag{3.4b}$$

and thus, from (3.3) amd (3.4a):

$$I_{E1} - I_{E2} = I_{Eo1}\exp(qV_{BE1}/kT)\left[1 - \exp\{-(V_1+I_{B2}R_{B2}-I_{b1}R_{B1}-\theta)q/kT\}\right] \tag{3.5}$$

where θ is defined by $\exp(q\theta/kT) = I_{Eo2}/I_{Eo1}$. At the collector node, the output voltage V_o is related to the differential emitter current by:

$$V_o = R_{L1}I_{C1} - R_{L2}I_{C2} = a^b_{I1}R_{L1}I_{E1} - a_{I2}R_{L2}I^b_{E2} \tag{3.6}$$

Therefore, from (3.5), (3.6) and the continuity relation: $I_{E1}+I_{E2}=I_E+I_e$, the following is obtained:

$$V_o = a^b_{I1}R_{L1}(I_E+I_e)\frac{1-\exp\left[-(V_1+I_{B2}R_{B2}- I_{B1}R_{B1} - \theta - \phi)q/kT\right]}{1+\exp\left[-(V_1+I_{B2}R_{B2}- I_{B1}R_{B1} - \theta)q/kT\right]} \tag{3.7}$$

where ϕ is defined by $\exp(q\phi/kT) = a^b_{I2}R_{L2}/a^b_{I1}R_{L1}$.

The exponents in the numerator and denominator of (3.7) differ only by the parameter ϕ which is a measure of the load and gain imbalance between the two transistors. If these quantities are closely matched, ϕ, and therefore the difference between the two exponents, becomes small so that (3.7) may be

rewritten as:

$$V_o \simeq a_{I1}^b R_{L1}(I_E+I_e)\tanh(qV_1'/2kT) \qquad (3.8)$$

in which V_1' is the effective input voltage to the multiplier. It is defined as:

$$V_1' = V_1 + I_{B2}R_{B2} - I_{B1}R_{B1} - \Theta - \emptyset \qquad (3.9)$$

so that it takes into account the voltage drop across the R_B's and also represents as an equivalent input offset voltage $(\Theta+\emptyset)$ the total effect of imbalance in the circuit.

If the argument of the tanh function in (3.8) is small compared with 1, the function may be approximated by the first term of its power series expansion. Thus, (3.8) becomes:

$$V_o \simeq a_I^b R_L(I_E+I_e)(q/2kT)V_1' \qquad (3.10)$$

in which a_I^b and R_L are mean values for Q_1 and Q_2. This last approximation is correct to within 5% for $|V_1'| \leq 20$ mV and to 10% for $|V_1'| \leq 28$ mV. From (3.10) we see that the circuit forms the analog product between the effective input voltage V_1' and the emitter signal current I_e. In addition, the output contains a term proportional to $I_E V_1'$, in which I_E is constant. The presence of the latter term, however, does not constitute a problem in this application.

To bring out more clearly the relation between V_1 and V_1', (3.4b), (3.6) and (3.10) are used to obtain:

$$V_o\left[1 + \frac{a_I^b}{a_I^e}\frac{q}{2kT}(I_E+I_e)R_B\right] = a_I^b R_L(I_E+I_e)(q/2kT)(V_1-\Theta-\emptyset) \qquad (3.11)$$

Ideally, the coefficient of V_o on the left side should be close to unity and the following design constraint is therefore imposed:

$$\frac{a_I^b}{a_I^e}\,\frac{q}{2kT}\,(I_E+I_e)R_B \ll 1 \qquad (3.12)$$

In practice this coefficient is greater than unity and contains a time-varying component arising out of the sinusoidal nature of I_e. (I_e is the local oscillator input when the multiplier is used in the selective amplifier). Typical numerical values of the circuit parameters, taken from the realization of Chap. 6, are:

$a_I^b = 0.98$, $a_I^e = 50$, $q/2kT = 19.2\ V^{-1}$ (at $T = 300^oK$),

$R_B = 100$ ohm, $R_L = 1{,}400$ ohm, $I_E = 4$ mA, $I_e = 2$ mA pk-pk.

The mean value of the coefficient of V_o in (3.11) is found to be 1.16 and may be made closer to unity by the reduction of I_E and R_B. The parameters Θ and ϕ measure the mismatch between the two sides of the multiplier which results from random variations in integrated circuit processing. For an extreme case of 20% mismatch between all respective circuit parameters we find, from the definitions following (3.5) and (3.7): $\Theta, \phi = \pm 4.5$ mV.

In the following calculations it will be assumed that (3.12) is satisfied by design and the low-frequency transfer function given by (3.11) is now used to characterize the operation of the multiplier in the selective amplifier.

When used as an input multiplier, the circuit is assumed to have a differential base input signal $V_1 = V_{im}\sin(\omega_i t+\psi) + V_r$, where V_r is an externally applied dc offset used for balancing.

The emitter current signal is $I_e = I_{em}\sin\omega_o t$. From (3.9) and condition (3.12) the effective input voltage is:

$$V_1' = V_{im}\sin(\omega_i t+\psi) + V_r - (\theta+\phi) \qquad (3.13)$$

Substitution into (3.11) gives:

$$V_o = P_o\left\{V_{im}\cos\left[(\omega_o-\omega_i)t-\psi\right]-V_{im}\cos\left[(\omega_o+\omega_i)t+\psi\right]+2\left[V_r-(\theta+\phi)\right]\sin\omega_o t\right\}$$
$$+Q_o\left\{V_{im}\sin(\omega_i t+\psi) + V_r -(\theta+\phi)\right\} \qquad (3.14)$$

in which P_o and Q_o are low-frequency values of the transfer coefficients defined previously in Sec. 2.2. They are evaluated as:

$$P_o = a_I^b R_L(qI_{em}/4kT) \quad \text{and} \quad Q_o = a_I^b R_L(qI_E/2kT) \qquad (3.15)$$

Continuing with the numerical example cited above, we obtain $P_o \simeq 14$ and $Q_o \simeq 56$. More accurate values, which take (3.12) into account are $P_o = 12.1$ and $Q_o = 48$.

When used as an output multiplier, the circuit has the same emitter current input but the base signal is at the differency frequency $(\omega_o-\omega_i)$. However, the values of P_o and Q_o are the same as in (3.15).

The terms in (3.14) which are proportional to $(V_r-\theta-\phi)$ are unwanted output components arising out of imbalance in the multiplier. If, at a given temperature, the multiplier is balanced by setting $V_r = \theta+\phi$, it is then important to know how this balance point changes with temperature. To calculate

this we assume that V_r may vary with temperature but that the ratios I_{Eo2}/I_{Eo1} and $a_{I2}^b R_{L2}/a_{I1}^b R_{L1}$, which define Θ and ϕ, are constant with temperature. This is reasonable in view of the earlier statement that the transistors and their loads are fabricated together on the same silicon chip. The following partial derivative, therefore, gives the balance-point sensitivity S_b^T, expressed in terms of an equivalent input drift voltage;

$$S_b^T = \frac{1}{Q_o}\frac{\partial}{\partial T}\left[Q_o(V_r-\Theta-\phi)\right] = (V_r-\Theta-\phi)\frac{1}{a_I^b I_E R_L}\frac{\partial}{\partial T}(a_I^b I_E R_L) + \frac{\partial V_r}{\partial T} - \frac{V_r}{T} \tag{3.16}$$

An analogous expression is found for the output components proportional to P_o.

A case of particular interest here is where I_E is produced by a constant voltage source in series with a diffused resistor. It is apparent that, since I_E then has an equal and opposite coefficient to the diffused resistor, the product $I_E R_L$ is virtually constant with temperature. Also, the coefficient of a_I^b is very small, so that the the first term in (3.16) is negligible in this case. Furthermore, if we assume that V_r is constant with temperature and that it takes a typical value [22] of 3 mV, the drift sensitivity is approximately $S_b^T \simeq 10\ \mu V/^oC$. Eqn. (3.16) is used in Chap. 5 to determine the drift charactersitics of the total system.

In conclusion, we note a second-order effect which arises out of imbalance in the multipliers. If (3.8) is represented as $w \simeq k \tanh(v+u)$, where v is the signal input

and u is the dc imbalance term, this may be expanded as

$$w = k\left[(v+u) - \frac{1}{3!}(v+u)^3 + \ldots\ldots\right]$$
$$= k\left[v(1-\tfrac{1}{2}u^2+\ldots) + u(1-\tfrac{1}{2}v^2+\ldots)+\ldots\right] \qquad (3.17)$$

Because even powers of sinusoidal terms contain dc components, it may be seen from the second term in this expression that the output of the multiplier contains a dc component which varies with even powers of the input amplitude. Thus, the balance point of the multipliers is sensitive to signal level. Fortunately, it has been found experimentally that this effect becomes important only near the overload point of the circuit, i.e. V_1 of the order 20 mV. The reduction of this effect is largely a function of the high degree of parameter matching attainable in monolithic structures.

3.3 High-frequency characterization.

The frequency dependence of the multiplier transfer function, which is due to charge-storage effects in the transistors, results in a reduction of the coefficients P and Q at high frequencies, as well as a phase-shift which is not predicted by the characterization of the previous section. As a qualitative indication of the effect of frequency, we note that a_I^e decreases at high frequencies and that this increases the coefficient of V_o in (3.11), leading to a reduction in output. This argument is not sufficiently accurate, however, and the following characterization is developed.

In both the input and output multipliers the local oscillator signal is applied at the emitter, while at the bases the input siganl is at ω_i for the input multipliers and at the differency frequency $(\omega_o - \omega_i)$ for the output multipliers. The input multiplier case is treated first, and the calculations are simplified somewhat by assuming initially that $\omega_i = \omega_o$, i.e. the input signal is in the center of the passband.

The base-to-collector current gain is represented by a single dominant-pole approximation:

$$a_I^e(s) \simeq \frac{a_I^e(0)}{1 + sT_o} \tag{3.18}$$

in which s is the complex frequency variable and T_o may be measured or calculated from the hybrid-pi model, for Q_1 (or Q_2) loaded by R_L. With the multiplier exactly balanced, i.e. $V_r - \theta - \phi = 0$, (3.11) is rewritten in the form:

$$V_o = a_I^b R_L (I_E + I_e)(q/2kT)\left[V_1 - V_o R_B/R_L a_I^e(s)\right] \tag{3.19}$$

in which the relations given by (3.4b) and (3.6) have again been used. From (3.18), $1/a_I^e(s)$ is now interpreted as a differential operator in the time domain and the last factor in (3.19) becomes: $V_1 - V_o R_B/R_L a_I^e(0) - T_o R_B \frac{d}{dt}\left[V_o/R_L a_I^e(0)\right]$. However, if condition (3.12) is satisfied, the second term in this expression is negligible, leaving only the input voltage and the time-dependent term.

To simplify the resulting form of (3.19), the following

parameters are defined:

$$\begin{aligned} \omega_o t &= \tau \\ a_I^b R_L (q/2kT) I_E &= M \\ I_{em}/I_E &= \rho \\ \omega_o T_o R_B / R_L a_I^e(0) &= x \end{aligned} \tag{3.20}$$

where it is seen that x is the normalized input and local oscillator frequency variable. With these substitutions, (3.19) reduces to:

$$V_o = M(1 + \rho \sin\tau)(V_{im} \sin\tau - x\frac{dV_o}{dt}) \tag{3.21}$$

which can be rearranged into the standard form of a first-order differential equation with variable coefficients:

$$y' + P(t)y = Q(t) \tag{3.22}$$

This equation has an exact solution[22] given by:

$$y = \exp(-\int_o^t P(t)dt) \left[C + \int_o^t Q(t)\exp(-\int_o^t P(t)dt)dt\right] \tag{3.23}$$

However, in this problem, the form of the integrands is such that the solution cannot be obtained in closed form. Therefore an approximate method of solution is sought.

The output voltage is first approximated by the series

$$V_o = a_o + a_1 \sin\tau + a_2 \sin 2\tau + b_1 \cos\tau + b_2 \cos 2\tau \tag{3.24}$$

This is substituted into (3.21) and coefficients are compared.

The following system of equations results:

$$\begin{bmatrix} 1 & 0 & 0 & -\frac{1}{2}M\rho x & 0 \\ 0 & 1 & -M\rho x & -Mx & 0 \\ 0 & \frac{1}{2}M\rho x & 1 & 0 & -2Mx \\ 0 & Mx & 0 & 1 & -M\rho x \\ 0 & 0 & 2Mx & \frac{1}{2}M\rho x & 1 \end{bmatrix} \begin{bmatrix} a_o \\ a_1 \\ a_2 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2}M\rho V_{im} \\ MV_{im} \\ 0 \\ 0 \\ -\frac{1}{2}M\rho V_{im} \end{bmatrix} , \qquad (3.25)$$

We are interested here in the output at the difference frequency $(\omega_o-\omega_i)$, which is the dc term a_o since we have assumed that $\omega_o=\omega_i$. Its value is obtained by a Cramer's rule solution of (3.25) and the result is expressed as the value of the frequency-dependent transfer coefficient $P(\omega_o)$ for a down-conversion from ω_o to dc:

$$P(\omega_o) = P_o \frac{1 + (Mx)^2(4+\rho^2/2)}{1 + (5+\frac{3}{2}\rho^2)(Mx)^2 + (Mx)^4(4-3\rho^2+\frac{1}{2}\rho^4)} \qquad (3.26)$$

Since this function varies only slowly with frequency over the range of interest, it also represents the down-conversion transfer coefficient for input frequencies slightly different from ω_o.

For the up-conversion mode, where the multiplier is used at the output of a channel, the calculation is repeated. The rather unexpected result is that the corresponding transfer coefficient is identical with (3.26).

As a check on the accuracy of this approximate solution, an analog computer solution of the multiplier equation has been performed. To set this up, it is necessary to make the following minor approximation: $(1+\rho\sin\tau)^{-1} \simeq (1-\rho\sin\tau)$ for

$\rho \ll 1$. If is made small, it is seen from (3.26) that the frequency dependence of the transfer coefficient becomes insensitive to the value of ρ, so that the making of this approximation is justified. Accordingly, (3.21) is rearranged and integrated to produce:

$$V_o \simeq \frac{1}{Mx} \int_0^t \left[MV_{im} \sin\tau - V_o(1 - \sin\tau) \right] dt \qquad (3.27)$$

The computer schematic for solving this is shown in Fig. 3.3. The driving function $\rho \sin\tau$ is formed in the two integrators and inverting amplifier at the top of the diagram, the computation is done in the second and third rows, while the last row forms the time-average of V_o. Since this average is analogous to a_o in the previous calculation, a plot of its variation with Mx gives the required dependence of $P(\omega_o)$.

In Fig. 3.4 the solutions for $P(\omega_o)$ obtained by these two methods are plotted as a function of the normalized frequency Mx, for $\rho = 0.25$. In comparing the two curves it is seen that the approximate analytical solution gives a slightly optimistic prediction of the frequency range for constant $P(\omega_o)$, but is sufficently accurate to provide a first-order description.

If $P(\omega_o)$ is recalculated for a base input of $V_{im} \cos\omega_o t$ and denoted by $P(\omega_o, \pi/2)$, it is found that $P(\omega_o, \pi/2) = Mx P(\omega_o)$. Therefore, for a general phase angle α, between the base signal at ω_o and the emitter signal at ω_o, it follows that the transfer coefficient is:

$$P(\omega_o, \alpha) = P(\omega_o)(\cos\alpha + Mx \sin\alpha) \qquad (3.28)$$

and the condition for this to be a maximum for a given Mx is that α = arctan(Mx). This result may be expressed as follows. At high frequencies there is a phase-shift in the multiplier owing to the base current flowing in R_B. In the case where $\omega_o = \omega_i$, the signal at the "internal" base nodes of Q_1 and Q_2 lags the external base signal V_1 by the phase angle α, as given. If this phase-shift is large, the multiplier ceases to operate correctly in the system. By comparing this phase data with the amplitude data of Fig. 3.4, we see that both effects become important at similar frequencies. From these considerations, therefore, an upper frequency limit for correct multiplier operation is set at Mx = 0.2.

This condition is now related to the circuit parameters. From (3.20), Mx = 0.2 corresponds to

$$\omega_o = 0.4 \frac{a_I^e(0)}{T_o g_m R_B} \qquad (3.29)$$

where $g_m = a_I^b q I_E / kT$. As a numerical example we take the following case:

$1/2\pi T_o$ = 3.0 Mc/s, $a_I^e(0)$ = 50, I_E = 4 mA, R_B = 50 ohm. Substitution in (3.25) gives, for the maximum frequency, 7.5 Mc/s.*

3.4 Design considerations.

For correct multiplier operation, the transistors should be operated in an emitter current range where their current gains are not current-dependent. To ensure this, the amplitude

* The upper frequency limit for the multipliers of Chap.6 is approximately 1.6 Mc/s.

of the emitter signal current I_{em} should be somewhat less than the bias current I_E, although, if it is too small, transfer gain is sacrificed unnecessarily. A good compromise is obtained for $I_{em} = 0.25I_E$. To minimize the contribution of the I_BR_B terms in the transfer function, the emitter current should be high enough to optimize the dc current gain of the transistors but not so high as to violate inequality (3.12). For similar reasons, R_B should be minimized. Since T_o increases with collector load, because of the collector-base feedback capacitance, the ac impedance of the load should be small. For the case of the input multiplier, this may be attained simply by incorporating the lowpass network into the load so that a large shunt capacitance dominates.(See Fig. 4.1).

The emitter current sources may be realized by common-emitter transistor stages or, as is the case here, by resistors. It is sufficient that these resistors have a conductance which is a tenth (or less) of the g_m of the transistors. The input multiplier bases may be driven by a single-ended signal source without appreciably affecting their operation, provided that a capacitor is used for dc isolation. This capacitor also attenuates and low-frequency components which may be passed by the lowpass stage, As noted in Sec. 3.2, it is normally necessary to provide a dc offset adjustment either at the input or output of the input multipliers for balancing each channel. Practical circuits are given in Chap.6.

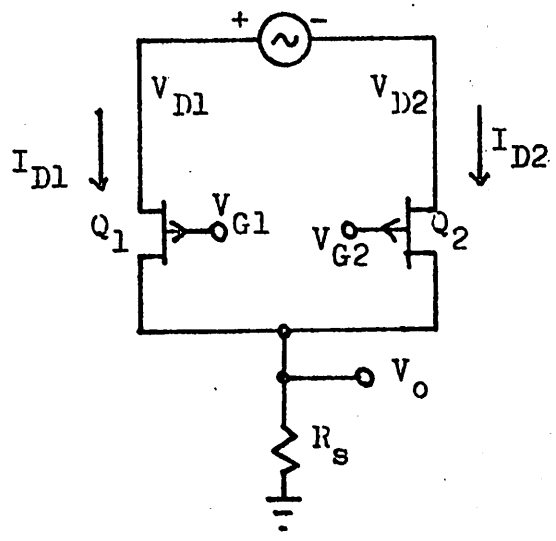


Figure 3.1 Balanced analog multiplier using field-effect transistors.

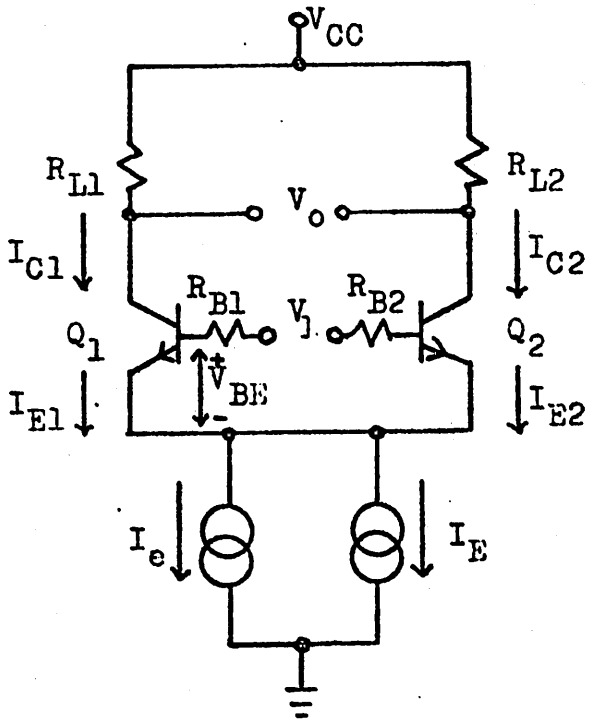
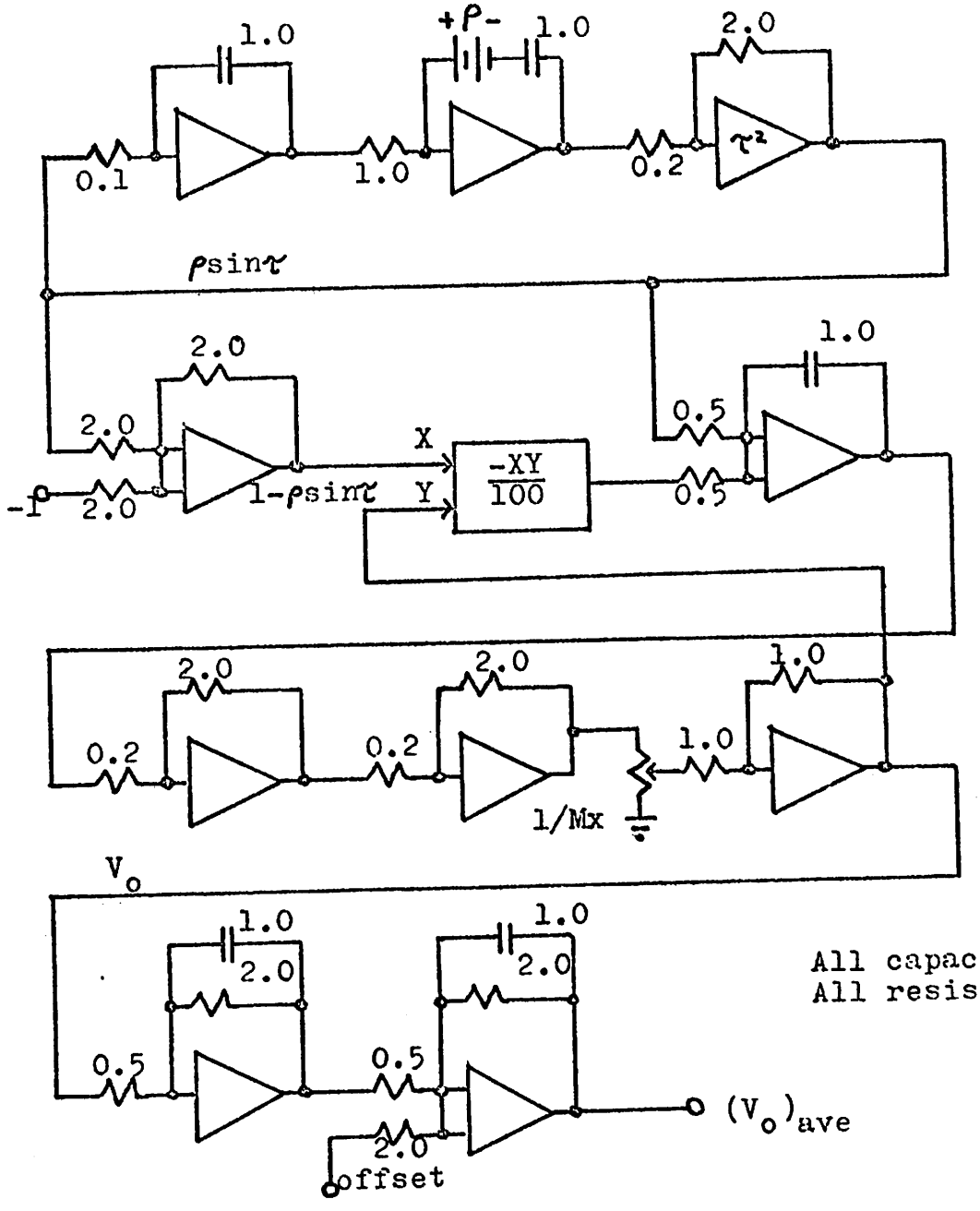


Figure 3.2 Balanced analog multiplier using bipolar transistors.



All capacitors in μF
All resistors in MΩ

Figure 3.3 Analog computer schematic for solving:

$$V_o = (1/Mx)\int_o^t [MV_{im}\sin\tau - V_o(1-\rho\sin\tau)]\,dt.$$

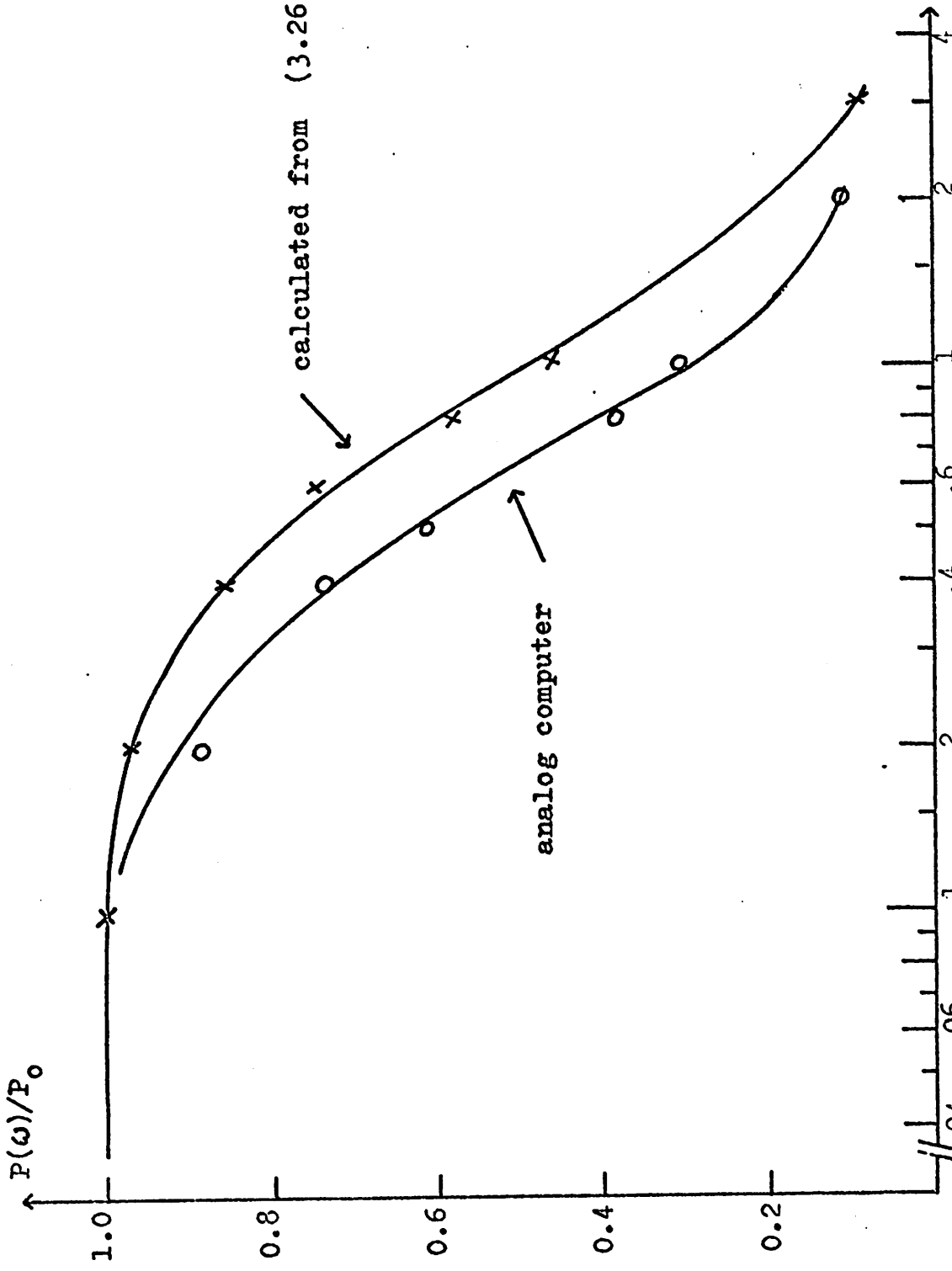


Figure 3.4 Frequency dependence of multiplier transfer coefficient

CHAPTER 4. LINEAR CIRCUITRY.

4.1. The lowpass stage.

Both channels of the system contain lowpass amplifying stages which, ideally, have equal characteristics. As we have seen, the passband of the system is an image of these lowpass characteristics. In principle, therefore, any desired bandpass characteristic can be obtained by designing these stages for the corresponding lowpass function. The realization of very narrow-band filters is consequently only dependent on the availability of suitably long time constants. A restriction is made in this analysis to lowpass transfer functions given by two real poles, since these provide sufficient flexibility for most design goals of interest here to be met.

With the exception of the means by which its long RC time constants are realized, the design of the lowpass stage presents no special problems. The value of the midband gain can influence the dynamic range of the system and this factor is treated in Sec. 5.1 below. At this point it is sufficient to know that the midband gain should lie in range 0.5 to 5 for typical circuits, and that this can be obtained with one, or

perhaps two, gain stages. The range for linear amplification should be larger than the dynamic range of the multipliers, so that it does not limit the response of the system. Furthermore, the value of the gain should not be strongly dependent on transistor parameters such as β_o which have a wide production tolerance. A configuration which may be simply designed to meet these requirements is shown in Fig. 4.1 and its midband gain is:

$$H(0) \simeq \frac{R_1+r_x+\beta_o(R_e+1/g_m)}{R_C+R_1+r_x+\beta_o(R_e+1/g_m)} \cdot \frac{R_2}{R_e+1/g_m+r_x/\beta_o} \qquad (4.1)$$

where r_x, g_m and β_o are averaged hybrid-pi parameters of Q_1 and Q_2. Note the total collector load on the input multiplier is now $R_L = R_C \| [R_1+r_x+\beta_o(R_e+1/g_m]$ and that R_2 is approximately the source resistance for the output multipliers. R_2 should therefore be small. The R_1C_1 networks attenuate the common-mode signal from the input multipliers as well as high-frequency difference components, so that this stage need not have a large common-mode rejection ratio.

A practical upper limit on monolithic junction or MOS capacitors is 400 pF. For conventional diffused resistors the limit is about 10 kΩ.[24] The product of these maximum values corresponds to a lowpass bandwidth of the order 50 kc/s. Therefore, a simple monolithic realization of the lowpass filter is convenient for bandwidths greater than this figure, but for smaller bandwidths it is necessary to resort to other realizations.

One possible method is a hybrid construction using subminiature ceramic capacitors which are now available, or even external capacitors of conventional dimensions if size is not a consideration. Besides simplicity, the use of ceramic capacitors is also advantageous from the viewpoint of sensitivity. Base-diffused resistors have temperature coefficients of the order +2000 ppm/^{o}C[24] and if these are used with MOS or junction capacitors which are almost constant with temperature, the bandwidth of the system will have the same temperature coefficient as the resistors. Ceramic capacitors, however, have controllable coefficients. They can be fabricated with a coefficient which is equal and opposite to that of the resistors, thus resulting in a net desensitivity of the system bandwidth.

An alternative approach to the long time-constant problem is provided by the Miller-effect multiplication of a capacitance when it is placed in the negative feedback path of a linear amplifier. A simple form of a multiplied-capacitance circuit, which may be realized in monolithic form, is shown in Fig. 4.2. The capacitance C_1 may be MOS or junction-type although the former is normally preferable from a linearity viewpoint. Using a first-order model of the transistor, the low-frequency input capacitance C_B presented at port B in this circuit is given approximately by:

$$C_B \simeq C_1(1 + g_m R_L) \qquad (4.2)$$

where $g_m = qI_C/kT$. At high frequencies, where the reactance of C_1 approaches R_L, this value decreases towards C_1.

A multiplied capacitance also appears in shunt at port C and in this case its low-frequency value is approximately:

$$C_C \simeq C_1\left(1 + \beta_o \cdot \frac{R_{B1}}{R_{B1}+r_\pi+r_x}\right) \qquad (4.3)$$

where $r_\pi = \beta_o/g_m$. Again, this value decreases at high frequencies with the decrease in base-to-collector current gain.

Both ports therefore present multiplied capacitance values which enable RC time constants to be realized which are a factor of ten or more larger than their simple monolithic counterparts. The maximum voltage swing which may be applied at port B is of the order 10 to 50 mV, whereas port C may accomodate signals of several volts. The noise level in the second case is correspondingly higher, however, and the dynamic ranges at the two ports are approximately the same. The choice of which port to use is therfore made on the basis of signal level. This circuit has been used in a realization of the frequency translation filter using a port C entry. (See Chap.6.)

There are a number of limitations with this simple multiplication circuit. The frequency at which the capacitance value starts to decrease is often too low to obtain a bandpass characteristic with good skirt selectivity. Also, the capacitance value is either β_o dependent (at port C) or g_m dependent (at port B) and is therefore not accurately defined. For these reasons, the more involved circuit of Fig. 4.3 has been developed. It is seen to be basically a balanced shunt-series feedback pair. The current gain from the base

of Q_1 to the collector of Q_2 and the corresponding gain from Q_4 to Q_3 are both stabilized by the feedback applied through the resistors R_F. Because of the balanced configuration, the signal collector current in Q_3 is equal but oppostite in phase to that in Q_2 so that the current gain i_3/i_1, as marked, has the 180° phase-shift needed for capacitance multiplication.

The low-frequency input capacitance at port B for this circuit is given by:

$$C_B = C_1(1 - i_3/i_1) \simeq C_1(1 + R_F/R_E) \qquad (4.4)$$

The approximation which defines C_B in terms of a stable resistance ratio is valid only if the loop-gain A_L for the feedback paths through R_F is high, i.e.

$$A_L \simeq \beta_{o1} \frac{R_1}{R_1+r_{\pi 1}} \cdot \frac{\beta_{o2}R_2(R_E+1/g_{m2})}{R_2+\beta_{o2}(R_E+1/g_{m2})} \cdot \frac{1}{R_F} >> 1 \qquad (4.5)$$

with the additional constraints: $R_E >> 1/g_{m2}$, $R_1 > \beta_{o1}/g_{m1}$. In these expression, subscripts 1 and 2 have been used to identify the hybrid-pi parameters of Q_1 and Q_2, respectively.

An experimental discrete-component prototype of the circuit of Fig. 4.3 has been constructed with the following parameter values:

R_1 = 560 ohm, R_2 = 4,700 ohm, R_3 = 3,300 ohm, R_4 = 900 ohm, R_F = 2,700 ohm, C_1 = 150 pF, V_{CC} = +15 V, I_E = 2 mA (per transistor)

The measured performance is: i_3/i_1 = 22 (cf. R_F/R_E = 27) with a -3 dB cut-off frequency of 8 Mc/s. The port B capacitance C_B is 3,500 pF up to at least 3 Mc/s. The maximum signal-to-

noise ratio at port B is 60 dB.

One of the multiplied-capacitance blocks may be used to replace each of the capacitors in the lowpass filter. For a simpler realization, the block may be modified by adding a second capacitor between the collector of Q_2 and the base of Q_4. Each block may then replace a balanced pair of capacitors by using the collector of Q_2 as the second entry.

4.2 The summing amplifier.

A convenient means of summing the outputs of the two channels is to use small common collector loads for the two output multipliers as is illustrated in Fig. 6.5. The summed output is applied to an amplifier which performs four functions in the system, namely, to provide a balanced-to-single-ended transition in the signal, to raise the system gain to the desired level, to attenuate low-frequency signals which are transmitted directly by the output multipliers and, finally, to attenuate any second harmonic component of the local oscillator signal. The balanced-to-single-ended transition may be realized in several ways and a convenient method here is to take the output signal from one collector of a balanced stage with a high common-mode rejection. This transition is not ideal at high frequencies, because of parasitic capacitances which couple a portion of the common-mode signal into the output. This signal may be cancelled very simply, in practice, by a small adjustment of the balance controls in each channel so that an equal and opposite

differential signal is produced by the output multipliers.

For a typical narrow-band selective amplifier, the bandwidth of the lowpass stages is more than a decade below the center frequency. Therefore, by ac coupling two or more of the summing amplifier stages, the low-frequency components transmitted from the lowpass stages are easily removed.

The amount of second and higher-order harmonic distortion in the output depends partly on the purity of the local oscillator signal and can not, therefore, be predicted exactly. In practical realizations, this distortion has been found to be less than 10% of full output without filtering, and a notch filter which attenuates the second harmonic by 20 dB relative to the first reduces this distortion to a negligible level. Fig. 4.4 shows a simple active RC notch filter which meets these requirements. Its transfer function is given approximately by:

$$\frac{v_2}{v_1}(j\omega) \simeq \frac{R_3}{R_2} - \frac{j\omega R_1 C_1}{1-\omega^2 R_1^2 C_1^2 + 3j\omega R_1 C_1} \cdot \frac{R_3}{R_e + 1/g_m} \qquad (4.6)$$

and for correct operation, the following design relations should be satisfied: $R_1 C_1 = 1/2\omega_o$, $R_2 = 3(R_e + 1/g_m)$. For center frequencies of 1 Mc/s or more, the capacitance values in the summing amplifier are such that the whole circuit may be realized simply in monolithic form.

4.3 The phase-shift network

In principle, it is better to obtain the 90° phase-shift

between the oscillator inputs to the two channels by means of a lag-type network, since a phase lead is normally accompanied by an increase in the harmonic content of the local oscillator signal - because of the magnitude response of a phase-lead network. A network which is simple and provides a 90^{o} phase-shift is shown in Fig.4.5. Its transfer function is:

$$\frac{v_2}{v_1}(j\omega) = \frac{1}{1+(R_1+R_2)/R_3+j\,\omega\left[R_1C_1(1+R_2/R_3)+C_2(R_1+R_2)\right]-\omega^2R_1R_2C_1C_2} \tag{4.7}$$

The condition for 90^{o} phase-shift is: $\omega_o^2R_1R_2C_1C_2 = 1+(R_1+R_2)/R_3$. This may be set exactly in practice by making any one of the five elements adjustable over a small range, e.g. C_2.

If the phase-shift is set correctly at a given temperature and frequency, it is then sensitive to changes in these quantities and the sensitivity is computed as follows. Assuming that all resistors have the same temperature coefficient, and that the coefficients of the two capacitors are also equal to each other, we define a variable $\nu = \omega_o R_1C_1$. The transfer function (4.6) is now written as $1/(A+jB)$ in which $A = 0$ for a 90^{o} phase-shift. If a small change $\Delta\nu$ in ν causes a change ε in the phase-shift, these changes may be shown to be related by:

$$\frac{\varepsilon}{\Delta\nu} = \frac{1 + B(\omega)}{B^2(\omega)} \tag{4.8}$$

where ε is measured in radians. This is the general sensitivity relation for this network and may be interpreted for a

particular case as follows.

Let $R_1 = R_2 = R_3$ and $C_1 = C_2$ in (4.7). The condition for A = 0 is $\nu = \sqrt{3}$ and B = $4/\sqrt{3}$, and from (4.8), $\epsilon/\Delta\nu$ = .057. A resistance change of 2000 ppm/^{o}C therefore produces a phase change of 0.057 x 2 x 3 x 10^{-3} rad/^{o}C, i.e. 0.0114 deg/^{o}C. From the definition of ν, it then follows that the phase sensitivity to a given fractional change in frequency or capacitance is the same. This result is used in Sec 5.2 where its effect on system performance is evaluated.

Finally, we note that the magnitude of the transfer function (4.7) is a strong function of frequency, as for other similar phase-shift networks, so that the main consequence of a shift in the local oscillator frequency is not a phase effect, but a change in the amplitude of the drive to one of the channels, causing a gain imbalance. To correct for this effect and to compensate for any other cause of gain imbalance, a potentiometer-type adjustment may be included. Fig. 6.5 shows a practical realization of the phase-shift network.

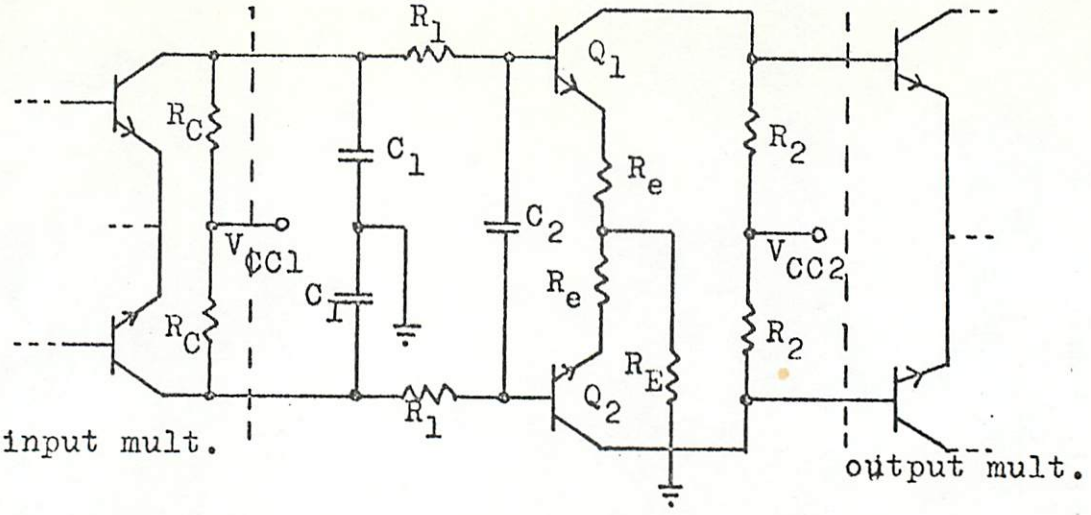


Figure 4.1 Single-stage lowpass amplifier.

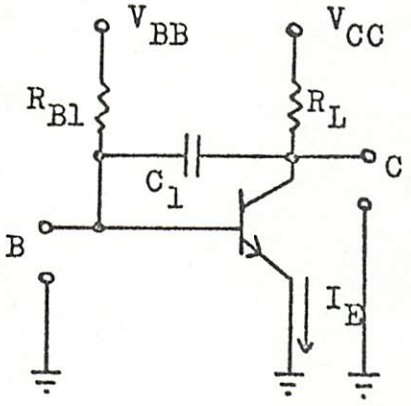


Figure 4.2 Single-stage capacitance multiplier.

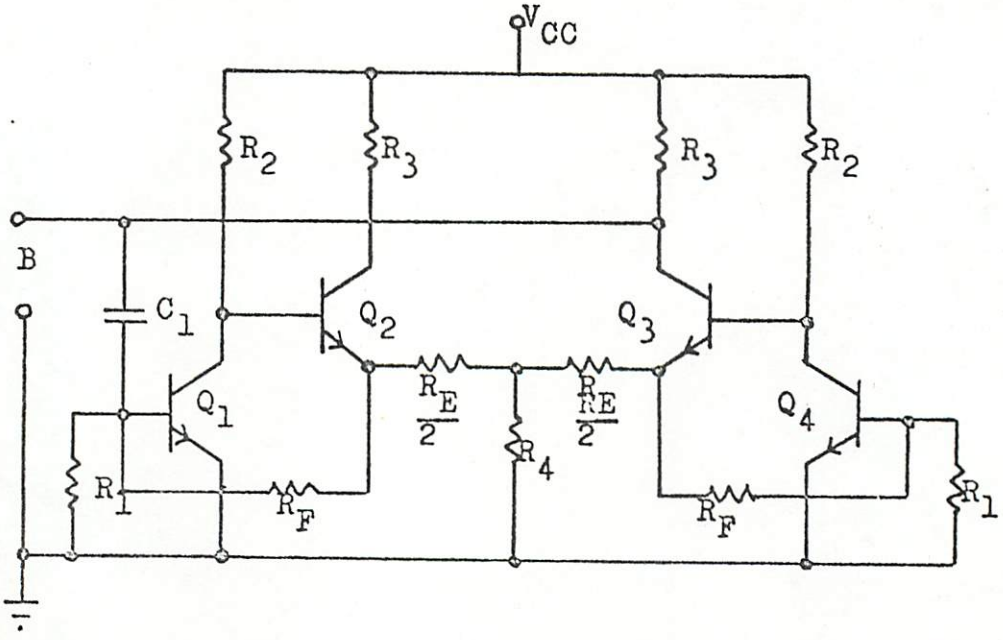


Figure 4.3 Two-stage capacitance multiplier.

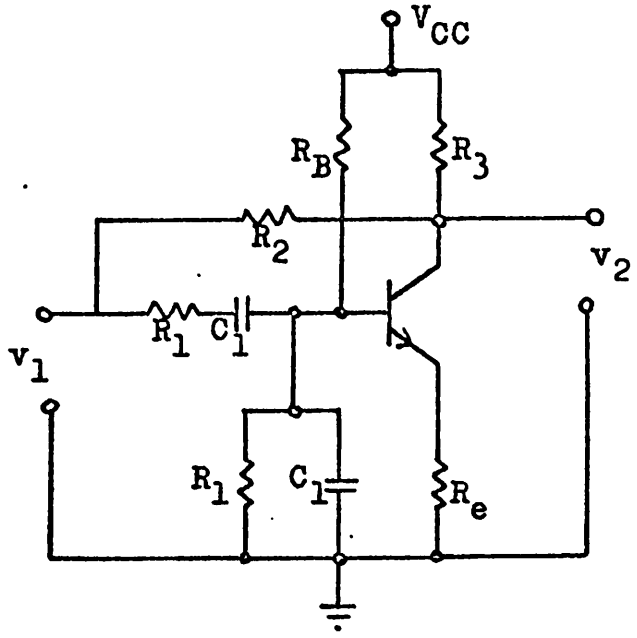


Figure 4.4 Active RC notch filter.

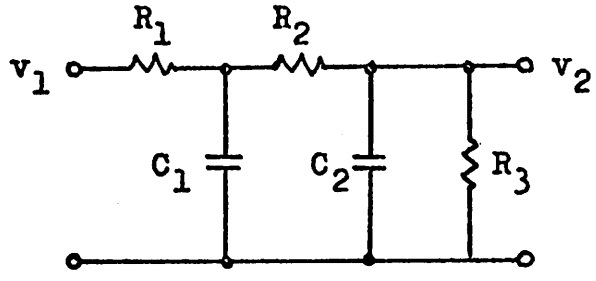


Figure 4.5 90^{o} phase-lag network.

CHAPTER 5. SYSTEM PERFORMANCE.

In this chapter the results obtained to this point are used to examine the ways in which the performance of a practical system differs from the idealized model presented in Chap. 2. Particular attention is paid to the sensitivity of various system parameters with temperature and a section is included in which the effects of negative feedback on performance are treated.

5.1 Dynamic Range.

The maximum allowed input signal to the system is determined partly by overload considerations and partly by a second-order effect arising out of any asymmetry in the input multipliers, which was described in Sec. 3.2. It was noted that a consequence of this asymmetry is the production of a dc signal in the lowpass stage whose amplitude varies as the square or higher order of the input voltage. This causes a spurious output signal, at the local oscillator frequency, which appears irrespective of whether the input signal is within the passband of the system or not. However, it was also noted in Sec. 3.2 that for the integrated multipliers used in this system, the asymmetry effect is only important

near overload levels in the circuit. On the basis of experimental measurements and the calculation of Sec. 3.2, the overload level for these multipliers is set at 20 mV sine-wave amplitude (40 mV pk-pk) and we require that the total input signal amplitude be greater than this.

The minimum detectable signal is set by noise and drift considerations and, in the following treatment, both of these effects are included. In contrast to conventional amplifiers, an appreciable amount of the noise at the output of this system is contributed by the output stages. This results from the fact that the narrow bandwidth of the lowpass stage limits the noise transmitted from the input multipliers, whereas the output multipliers are followed by a wideband amplifier.

We now proceed with a noise characterization of one channel in the system.

For a single common-emitter stage with a short-circuit collector load, it has been shown[25] that the equivalent input noise power due to both the transistor and the source resistance R_s may be represented by an equivalent noise resistance R_{nl} at the input.* Its approximate value is:

$$R_{nl} \simeq (R_s+r_x)(1+1/\beta_o) + 1/2g_m + (gm/2\beta_o)(R_s+r_x)^2\left[1+(\omega/\omega_{no})^2\right] \tag{5.1}$$

where $\omega_{no} \simeq \omega_t/\sqrt{\beta_o}$ is the noise corner frequency which is typically greater than 5 Mc/s. It is assumed in the calculations below that all frequencies of interest are below

---

* This expression does not include contributions due to 1/f noise.

ω_{no}. This expression is first applied to the input multiplier where a number of simplifications are possible. The short-circuit current gain β_o is first replaced by a_I^e corresponding to the source and load on this stage. Then, by a previous definition, $R_s+r_x = R_B$ and by condition (3.12), $g_m R_B/a_I^e \ll 1$ so that the last term in (5.1) may be neglected. Also, we note that $a_I^e \gg 1$. Finally, since this stage does not have the emitter grounded, a noise term due to the emitter supply resistor R_E should be added. Thus (5.1) becomes:

$$R_{n1} \simeq R_B + 1/2g_m + R_E \qquad (5.2)$$

The question of the noise bandwidth now arises. If the noise spectrum is "white" and the input multiplier is followed by a lowpass stage with bandwidth f_b, the following spectral regions are transmitted through the system: (0 to f_b) and ($f_o - f_b$) to ($f_o + f_b$). The former range is transmitted by the coefficient Q and the latter by a frequency translation through P. (See (3.15)). The mean square noise voltage at the output of the input multiplier $\bar{v}_{2n}^2$, which is represented as a noise voltage source in series with the load, is therefore:

$$\bar{v}_{2n}^2 \simeq 4kT\, R_{n1}(Q_1^2 + 2P_1^2)f_b \qquad (5.3)$$

For an estimate of a typical value we use the numerical example already introduced in which $P_1 = 14$, $Q_1 = 56$, $R_B = 100$ ohm, $g_m = 0.16$ mho, and $R_E = 300$ ohm. From (5.2), $R_{n1} \simeq 400$ ohm and for a typical $f_b = 10^4$ c/s, (5.3) yields

$\bar{v}_{2n}^2 \simeq 200\ (\mu V)^2$.

The lowpass stage has an equivalent input noise resistance R_{n2}, which is approximately the sum of its emitter resistance and the load resistance R_L of the input multiplier. The total mean square noise at the output of the lowpass stage is therefore approximately:

$$\bar{v}_{3n}^2 \simeq 4kT\left[R_{n1}H^2(Q_1^2+2P_1^2) + H^2R_{n2}\right]f_b \qquad (5.4)$$

where H is the voltage gain of the lowpass stage and the noise source, again, appears in series with load on this stage.

The output multiplier is followed by the summing amplifier whose bandwidth fs becomes the effective noise bandwidth of the output. The output noise due to the output multiplier alone $\bar{v}_{4n}^2$ is therefore:

$$\bar{v}_{4n}^2 \simeq 4kTR_{n3}(Q_2^2 + 2P_2^2)f_s \qquad (5.5)$$

where R_{n3} is the corresponding input noise resistance. The summation of all these noise components is represented by $\bar{v}_{5n}^2$ at the output of the channel and is given by:

$$\bar{v}_{5n}^2 \simeq 4kT\left[\left\{R_{n1}H^2P_2^2(Q_1^2+2P_1^2) + H^2P_2^2R_{n2}\right\}f_b + R_{n3}(Q_2^2+2P_2^2)f_s\right] \qquad (5.6)$$

Again, typical values are calculated for the numerical case above, with the following additional data taken from the circuits of Chap. 6: $P_2 = 2.2$, $Q_2 = 8.8$, $H = 0.5$, $f_s = 5\times10^5$ c/s. The result is $\bar{v}_{5n}^2 = 470(\mu V)^2$, i.e. 22 μV r.m.s.

and when the output of the 2 channels is summed, the mean square output noise voltage is approximately 1000 $(\mu V)^2$.

An expression for the total thermal drift in one channel is now derived, using the relations(3.14) and (3.16) of Chap. 3. Starting with the input multiplier, which has an external dc offset input V_{r1}, we have shown that the total equivalent imbalance voltage at the input is $(V_{r1} - \theta_1 - \phi_1)$. This produces a dc differential voltage $Q_1(V_{r1} - \theta_1 - \phi_1)$ at the output of this stage and, in turn, produces a dc input $HQ_1(V_{r1} - \theta_1 - \phi_1)$ to the output multiplier.* The total effective input offset at the output multiplier, including the effect of its own imbalance is, therefore, $HQ_1(V_{r1}-\theta_1-\phi_1)-\theta_2-\phi_2$. This offset produces, at the output of the channel, a spurious signal at the local oscillator frequency which cannot be filtered from the true output at the input frequency. Its amplitude V_d is:

$$V_d = 2P_2\left[HQ_1(V_{r1} - \theta_1 - \phi_1) - \theta_2 - \phi_2\right] \quad (5.7)$$

This can be set to zero at a given temperature, but not, in general, over a range of temperature.

The temperature coefficient of V_d is found by a differentiation analogous to that leading to (3.16), with V_{r1} held constant.

$$\frac{V_d}{T} = \frac{V_d}{Y}\frac{\partial Y}{\partial T} - \frac{2P_2}{T}HQ_1V_{r1}+2P_2HQ_1(V_{r1}-\theta_1-\phi_1)\left(\frac{1}{H}\frac{\partial H}{\partial T} + \frac{1}{Y}\frac{\partial Y}{\partial T} - \frac{1}{T}\right) \quad (5.8)$$

in which $Y = I_{em}R_L$ for both stages.

---

* It has been assumed here that the lowpass stage does not introduce an additional dc offset. A more general relation in which such an effect may be included is given in Sec. 5.4.

When the outputs of the two channels are summed, the respective imbalance output signals are in quadrature and, in the worst case, both have equal magnitudes so that the total output imbalance signal has a temperature coefficient: $\sqrt{2}(\partial V_d/\partial T)$.

To complete the dynamic range specification, we now consider the the upper limit on the signal level. This calculation is much simpler than the previous two. As the signal level increases, either the input multiplier overloads before the output multiplier, or vice-versa. (The lowpass stage and summing amplifier are designed so that they do not limit the dynamic range.) The factor that determines which stage overloads first is the midband gain of the lowpass stage H. To show this, let the input voltage which overloads the input multiplier be V_{1u} and the corresponding overload input to the output multiplier be V_{2u}. If the output signal from the two channels at the overload point is V_{3u}, we may write, at the center frequency of the filter:

$$V_{3u} = P_1HP_2V_{1u} \text{ if } P_1HV_{1u} < V_{2u}, \tag{5.9a}$$

$$= P_2V_{2u} \quad \text{if } P_1HV_{1u} > V_{2u}. \tag{5.9b}$$

In the special case where $P_1HV_{1u} = V_{2u}$, both expressions are identical and this can easily be shown to be the condition for maximum output signal. Therefore, to optimize the dynamic range of the amplifier, the gain of the lowpass stage is chosen to satisfy, at least approximately, this condition.

These three results are now combined in two system specifications: the maximum dynamic range (MDR) and the

thermal drift sensitivity S_d^T, whose definitions are:

$$MDR = \overline{V}_{3u}^2/2\overline{v}_{5n}^2 \quad \text{in dB.} \tag{5.10a}$$

$$S_d^T = \sqrt{2}(\partial V_d/\partial T)/V_{3u} \quad \text{in ppm/}^\circ\text{C} \tag{5.10b}$$

The numerical example begun above is completed with the calculation of these quantities. For $V_{2u} = V_{1u} = 20$ mV, we find, from (5.9b): $V_{3u} = 44$ mV. From above, $2\overline{v}_{5n}^2 \simeq 1000\ (\mu V)^2$, so that MDR $\simeq$ 58 dB. Typical measure values are 50 dB. Finally, (5.8) and (5.10b) are used to obtain $S_d^T \simeq 23{,}000$ ppm/°C. Because a worst case has been assumed here, typical sensitivities are a factor of two or more smaller than this and a method for reducing the drift sensitivity by a further factor of 20 is described in Sec. 5.4.

5.2 Sensitivity and distortion.

Sensitivity in its broadest sense includes the variation of system performance due to all changes in ambient conditions such as temperature, radiation flux, external supply voltage as well as internal variations due to aging. Of these, the effects of temperature are the most predictable and, in integrated circuits, are usually the most important. Therefore, the treatment of sensitivity is restricted to temperature effects, though the extension of these results to other sensitivities is not difficult. The four basic aspects of system performance are: gain, bandwidth, center frequency and distortion. As already noted, it is an important property of the frequency-translation technique that these quantities

may be treated independently.

The center-frequency sensitivity is almost purely a function of the type of local oscillator, since any loading interaction between the local oscillator and the rest of the system may be minimized by good design. If an integrated local oscillator [26, 27] is used, recent results indicate that frequency sensitivities as low as 100 ppm/^{o}C may be obtained, by using similar compensation techniques to those in active RC filters.[5]

The bandwidth sensitivity, which is essentially the sensitivity of the RC time-constants in the lowpass stages, has been commented on in Sec. 4.3, for the system without feedback, and the effects of feedback are noted in Sec 5.3. In general, the sensitivity is a function of the types of elements used in the RC networks, and not of the bandwidth itself.

In the signal path of the system there are at least four amplifying stages, as well as the passive networks, and the net midband gain sensitivity is related to the individual sensitivities of these stages. At room temperature (300^{o}K) the β_o of a transistor has a typical temperature coefficient[6] of +6000 ppm/^{o}C, while the temperature coefficient of g_m is -3300 ppm/^{o}C. The multiplier stages have g_m dependent transfer functions which thus have negative coefficients, but the gain of the linear stages tends to be more β_o dependent, resulting in a positive coefficient. At the center frequency, the overall system gain is, from (2.6), $K(\omega_o) = P_1 H(0) P_2 A$,

and its temperature coefficient S_K^T is:

$$S_K^T = \frac{1}{K}\cdot\frac{\partial K}{\partial T} = \frac{1}{H}\cdot\frac{\partial H}{\partial T} + \frac{1}{A}\cdot\frac{\partial A}{\partial T} + \frac{2}{I_{em}R_L}\cdot\frac{\partial(I_{em}R_L)}{\partial T} - \frac{2}{T} \qquad (5.10)$$

for $P_1 = P_2$. Since the algebraic sum of the terms in this expression may be set, by design, to zero over a finite temperature range, net gain desensitivity may be obtained. Note, however, that for a complete specification of the gain sensitivity, it is necessary to know the temperature variation of the local oscillator drive amplitude I_{em}, and this is a function of the type of local oscillator used.

We now consider the sensitivity aspects of distortion. Because the band-limiting filter in the summing amplifier removes frequency components which are well outside the passband of the system, we may restrict the treatment to distortion terms whose frequencies are close to the center frequency. These can result from three types of imperfection in the system: a difference between the transfer functions of the two channels (channel mismatch), a phase-shift between channels which is not exactly 90^o, and imbalance in the multipliers. The first two add to the output a distortion component at the frequency $(2\omega_o - \omega_i)$ while the last introduces a component at ω_o. The sensitivity of multiplier balance has already been derived in Sec. 5.1.

The channel mismatch sensitivity is conveniently obtained by defining a parameter $\gamma = P_{b1}H_b(0)P_{b2}/P_{a1}H_a(0)P_{a2}$, using the notation of Chap. 2. The ratio of the channel mismatch distortion component amplitude at $(2\omega_o - \omega_i)$ to the output at

ω_i is denoted D_m and is obtained by substituting in (2.5), giving:

$$D_m = \frac{1 - \gamma}{1 + \gamma} \tag{5.11}$$

In general γ has a sensitivity $S_\gamma^T = \partial\gamma/\partial T$, which arises out of a small differential gain sensitivity between the two channels. The temperature sensitivity of distortion due to channel mismatch is, therefore,

$$S_{Dm}^T = \tfrac{1}{2} D_m S_\gamma^T \tag{5.12}$$

The description of phase-shift effects is similarly treated by recalculating (2.5) for the case where the local oscillator input to the two channels is $I_{em}\sin\omega_o t$ and $I_{em}\cos(\omega_o t+\varepsilon)$, respectively. The distortion ratio due to the phase error is denoted D_ε and is thus found to be:

$$D_\varepsilon = \sin\varepsilon \tag{5.13}$$

From this we see that a phase error of 1^o produces an output at $(2\omega_o - \omega_i)$ whose amplitude is 1.8% of the true output at ω_i. The temperature sensitivity of ε has been described in Sec. 4.3 and is derived from (4.8). The temperature sensitivity of distortion due to phase error is thus:

$$S_D^T = D_\varepsilon S_\varepsilon^T \tag{5.14}$$

where $S_\varepsilon^T = \partial\varepsilon/\partial T$.

To conclude this section we consider the influence which the results obtained may have on the design of the system. The

magnitude of the drift sensitivity described in Sec. 5.1 is unpredictable because it depends on the degree of matching between transistor pairs obtained during fabrication. At the present state-of-the-art, this is a random parameter. In almost all cases, however, the degree of matching does lie within known limits and the numerical example which was calculated from (5.10b) is a worst case with respect to these limits. Beyond this, the performance of the system may be optimized by choosing the lowpass stage gain in accordance with the condition given, and by incorporating drift compensation as described below.

The distortion sensitivities to channel gain mismatch D_m and to phase-shift D_ε are fundamental and invariant quantities for a system without feedback. It is shown below, however, that they may be reduced by the use of feedback. The temperature sensitivity of channel gain match S_γ^T may be minimized by tight thermal coupling between stages, but its actual value is virtually unpredictable. Typical measured values are given in Chapter 6. On the other hand, S_ε^T has been calculated and may be minimized by time-constant compensation.

Finally, the gain sensitivity as given by (5.10) is controllable, since the system possesses a number of degrees of freedom in this respect. It is possible, in principle, to calculate the individual gain sensitivities, but in practice this calculation is very difficult. (Consider, for example the differentiation of (4.1) with respect to all sensitive parameters.) Computer studies and experimental measurements are more realistic approaches to this problem. Nevertheless,

the sensitivity of the system gain may be reduced to a low level simply by reducing its g_m and β_o dependence and may be reduced even further by the use of feedback.

5.3 Negative feedback.

Because of the nonlinear processes in this system, one cannot assume, a priori, that the application of negative feedback will improve its performance. However, it is shown here that such a technique does result in an improvement of most performance parameters. There is a variety of possible ways in which feedback may be applied, for instance, as local loops around the summing amplifier and lowpass stages, as a local loop around the lowpass stage-output multiplier cascade or as a single overall loop between input and output. All these possibilities have been investigated and the conclusion reached is that the last approach is to be preferred.

This feedback system is shown schematically in Fig. 5.1. The feedback network U is shunt-connected between the input bases and the output collector of the system. (Fig. 6.7 contains a more detailed diagram of the feedback connection). It is a simple RC highpass network with an assumed zero-phase, flat magnitude response to all frequencies within the passband of the system.

On the basis of the characterization of the forward path, it is proposed that the output voltage without feedback v_3 contains the desired frequency component at ω_i in addition to small distortion components at ω_o and $2\omega_o-\omega_i$. The first term

is found, from (2.6) to be $KV_{im}\sin\omega_i t$. The second is due to thermal drift, for which an expression is given in (5.7), and is represented as the sum of two components $V_{da}\sin\omega_o t$ and $V_{db}\cos\omega_o t$ from the respective channels. The other distortion term, resulting from channel mismatch and phase error is found from (5.11) and (5.13). The summation of these terms is therefore:

$$v_3 = KV_{im}\left[\sin\omega_i t + D_m\sin(2\omega_o-\omega_i)t + D_\ell\cos(2\omega_o-\omega_i)t\right] + V_{da}\sin\omega_o t + V_{db}\cos\omega_o t \qquad (5.15)$$

When the feedback loop is closed, the magnitudes of these terms are modified in a rather complex way, because of the interaction between the various frequency components. An approximate solution is obtained, however, if the distortion terms are small relative to the true output, by comparing coefficients of like frequency components around the loop. It is also assumed here that reverse transmission through the feedback network is negligible and that the feedback subtraction at the input is ideal. For convenience, a coefficient $w=(1+Ka)^{-1}$ is defined and the input voltage under closed-loop conditions is set at $V'_{im} = V_{im}/w$ so that distortion may be compared at the same output level. The closed-loop output voltage thus obtained, v'_3, is :

$$v'_3 \simeq KV_{im}\left[\sin\omega_i t + wD_m\sin(2\omega_o-\omega_i)t + wD_\ell\cos(2\omega_o-\omega_i)t\right] + wV_{da}\sin\omega_o t + wV_{db}\cos\omega_o t \qquad (5.16)$$

This result is a standard one in the respect that it shows the reduction and consequent desensitization of the midband gain and distortion level by the factor $(1+UK)^{-1}$. The system with feedback is therefore characterized by a transfer function $K' = K/(1+KU)$, and channel mismatch sensitivity $D'_m = D_m/(1+KU)$ and a phase error sensitivity $D'_\varepsilon = D_\varepsilon/(1+KU)$.

It appears, at first sight, that the drift sensitivity is also improved but there is another affect which nullifies this improvement. It may be seen from (5.8) and (5.16) that the closed-loop imbalance coefficient is $\sqrt{2}(\partial V_d/\partial T)/(1+KU)$. However, the maximum output signal V_{3u} is also reduced to $V_{3u}/(1+KU)$. To see this, we recall that for the system without feedback, the maximum input signal V_{1u} is independent of whether the input frequency is within the system passband or not. For this system with feedback, however, the overload point is frequency sensitive. Consider a signal of amplitude V_1 applied to the input bases. If its frequency is within the passband, the effective base drive is reduced by the feedback to $V_1/(1+KU)$. If, on the other hand, its frequency is well outside the passband, the loop-gain of the system is close to zero in magnitude and therefore the effective input voltage is approximately equal to the applied voltage V_1. Consequently, the system is more prone to overload for signal frequencies outside the passband and the maximum input signal must now be defined at some outband frequency. It thus has a value $V_{1u}/(1+KU)$. Similarly, the output overload level is now $V_{3u}/(1+KU)$ and the drift sensitivity S_d^T, as defined by (5.10b), is unchanged by the addition of feedback. In addition,

we note that the lowpass stage gain must be increased by the factor (1+KU) in order to satisfy the maximum dynamic range condition derived from (5.9). By an analogous argument, the MDR is also unchanged by the use of feedback. (See (5.10a).)

The experimental results reported in Chap. 6 confirm these predictions.

The effect of feedback on the bandwidth of the system is treated as follows. The transfer function of the lowpass stage, which has been assumed to have two negative-real poles, is represented as:

$$H(s) = \frac{H_o p_1 p_2}{(s-p_1)(s-p_2)} \tag{5.17}$$

From the general system transfer function as given in (2.6), we may write the open-loop transfer function K(s) as:

$$K(s) = 2A\, P_1 P_2 H_o \frac{2\left[s^2 - s(p_1+p_2) + (p_1 p_2 - \omega_o^2)\right]}{(s-j\omega_o-p_1)(s-j\omega_o-p_2)(s+j\omega_o-p_1)(s+j\omega_o-p_2)} \tag{5.18}$$

which has the desired two complex conjugate pole pairs. When the feedback network U, which is initially assumed to contribute no additional natural frequencies to the transfer function, is added and the feedback loop is closed, the closed-loop pole positions may be calculated by a fairly tedious manipulation of (5.18). The result is that the four poles have positions given by:

$$p_i = \frac{p_1+p_2}{2} \pm \frac{1}{2}\sqrt{(p_1+p_2)^2 - 4p_1p_2(1+K_oU)} \pm j\omega_o \tag{5.19}$$

where $K_o = 2A P_1 P_2 H_o$. The function $K_o U$ is recognized as the midband loop-gain in conventional feedback theory and the pole positions given by (5.19) may be plotted in root locus form with $K_o U$ as the parameter. This plot is shown in Fig. 5.2.

It may be seen from this plot that a variety of pole configurations may be obtained merely by choosing the loop-gain appropriately. In particular, a second-order Butterworth response is obtained for $(1+K_o U) = (p_1+p_2)^2/2p_1p_2$ and a pair of coincident conjugate poles for $(1+K_o U) = (p_1+p_2)^2/4p_1p_2$.

In the practical system there are, in addition to the dominant poles calculated here, a number of non-dominant poles which contribute an excess phase-shift to the midband loop-gain. As a consequence, the root-locus is perturbed, as shown by the dotted lines in Fig. 5.2 and the frequency response of the system is skewed about the center frequency. The magnitude of this effect may be estimated in the particular case of the above second-order system as follows: The transfer function $K(s)$ is assumed to be narrow-band, i.e. $|p_1|, |p_2| \ll \omega_o$, and a transformed frequency variable $s' = s - j\omega_o$ is introduced. Equation (5.18) may then be rewritten in the approximate form:

$$K(s') \simeq \frac{K_o p_1 p_2}{(s' - p_1)(s' - p_2)} \qquad (5.20)$$

This expression is next modified to include an excess phase-shift ξ at the center frequency ω_o, by multiplying it with $\exp(j\xi)$. The closed-loop poles of the system transfer function, in terms of the s'-plane, are now found from the zeros of $(1 + K(s')U \exp j\xi)$ and are given by:

$$p_i' \simeq \tfrac{1}{2}(p_1+p_2) \pm \tfrac{1}{2}(L^2-M^2)^{\frac{1}{4}}\left[\cos(\tfrac{1}{2}\arctan\tfrac{M}{L}) - j\sin(\tfrac{1}{2}\arctan\tfrac{M}{L})\right] \tag{5.21}$$

where $L=(p_1+p_2)^2-4p_1p_2(1+K_0U\cos\xi)$ and $M=4p_1p_2K_0U\sin\xi$. These poles may now be transformed into the original s-plane with the result:

$$p_i \simeq +\frac{p_1+p_2}{2} + \frac{(L^2-M^2)^{\frac{1}{4}}}{2}\left[\cos(\tfrac{1}{2}\arctan\tfrac{M}{L}) \pm j\sin(\tfrac{1}{2}\arctan\tfrac{M}{L})\right] \pm j\omega_0$$

$$\text{and } \frac{p_1+p_2}{2} - \frac{(L^2-M^2)^{\frac{1}{4}}}{2}\left[\cos(\tfrac{1}{2}\arctan\tfrac{M}{L}) \pm j\sin(\tfrac{1}{2}\arctan\tfrac{M}{L})\right] \mp j\omega_0 \tag{5.22}$$

This calculation shows that the result of an excess phase-shift is a change in both the center frequency and bandwidth of the closed-loop transfer function. This effect may be minimized in practice by the introduction of a simple RC network into either the forward or feedback path which is adjusted to cancel the excess phase-shift within the narrow passband of the system.

To illustrate these results, a numerical example taken from the feedback system of Chap. 6. is presented (See Fig. 6.7). The open-loop, lowpass stage poles are at: $p_1 = -4 \times 10^4$ sec.$^{-1}$ and $p_2 = -1 \times 10^3$ sec.$^{-1}$ The midband loop-gain K_0U is 8.0 with an estimated excess phase-shift of 30° at the center frequency. Equation (5.21) is now evaluated with $L = 4.1 \times 10^8$ and $M = 6.4 \times 10^8$ with the result:

$$s' \simeq (0.80 - j0.66) \times 10^4, -(3.3+j0.66) \times 10^4 \text{ sec.}^{-1}$$

The corresponding bandwidth (at a center frequency of 1.6 Mc/s) is 2.5 kc/s, which compares well with the measured value of

2.56 kc/s. The corresponding shift in center frequency is approximately 1.4 kc/s., i.e. 0.09%.

Finally, we note that the system bandwidth is sensitive to loop-gain, in the feedback case, as well as to the open-loop pole positions. This may be advantageous in some circumstances because it introduces an extra degree of freedom in designing for a desensitized response[5]. In practice, however, moderate desensitivity may be obtained if the sensitivities of the open-loop pole positions and the midband loop-gain are individually minimized.

5.4 Drift Compensation.

At the expense of further adjustments in the system, the drift sensitivity may be reduced by direct compensation. In the original system, provision is made for a single dc offset adjustment at the input multiplier of each channel (See Sec. 3.4). We now provide two adjustments in each channel, one at the input multiplier and one at the output multiplier. For a given channel we denote these by V_{r1} and V_{r2}, respectively. Equation (5.7), which was derived for the single-offset case, is now rewritten as:

$$V_d = 2P_2\left[HQ_1(V_{r1}-\theta_1-\phi_1) + V_{r2}-\theta_2-\phi_2\right] \qquad (5.23)$$

For balance, this expression should be zero, a condition which may be set by a range of combinations of the two independent variables V_{r1} and V_{r2}. The temperature coefficient

of V_d is found, as in (5.8), to be:

$$\frac{V_d}{T} = \frac{V_d}{Y}\frac{\partial Y}{\partial T} - \frac{2P}{T}(V_{r2}+HQ_1V_{r1}) + 2P_2HQ_1(V_{r1}-\theta_1-\phi_1)\left(\frac{1}{H}\frac{\partial H}{\partial T} + \frac{1}{Y}\frac{\partial Y}{\partial T} - \frac{1}{T}\right) \tag{5.24}$$

where, again, $Y = I_{em}R_L$ for both stages.

By setting $V_d = \partial V_d/\partial T = 0$, (5.23) and (5.24) become two simultaneous equations in V_{r1} and V_{r2} which provide for the simultaneous balancing of the channel and the minimization of its thermal drift sensitivity. This condition is set in practice by trial and error adjustments over a temperature range. The adjustments converge rapidly to a minimum sensitivity condition which gives typical values of $S_d^T = 500$ ppm/°C.

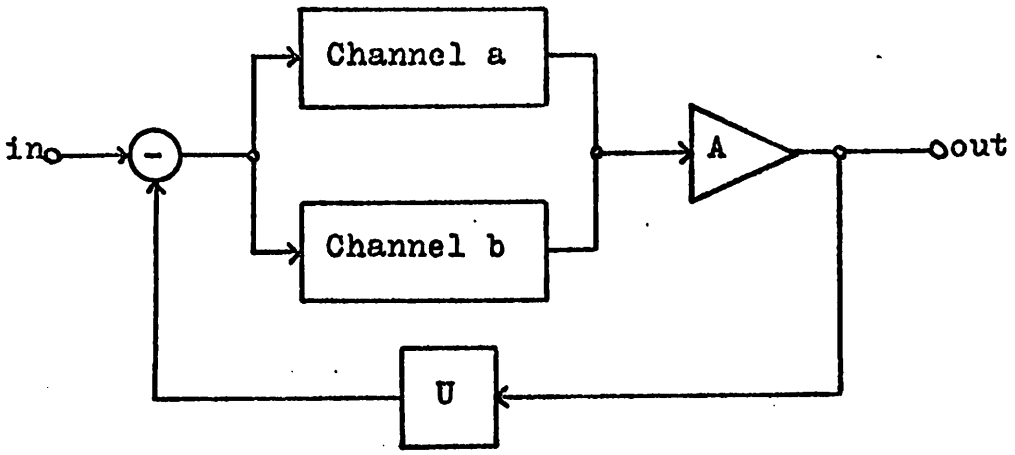


Figure 5.1 Schematic of single-loop negative feedback system.

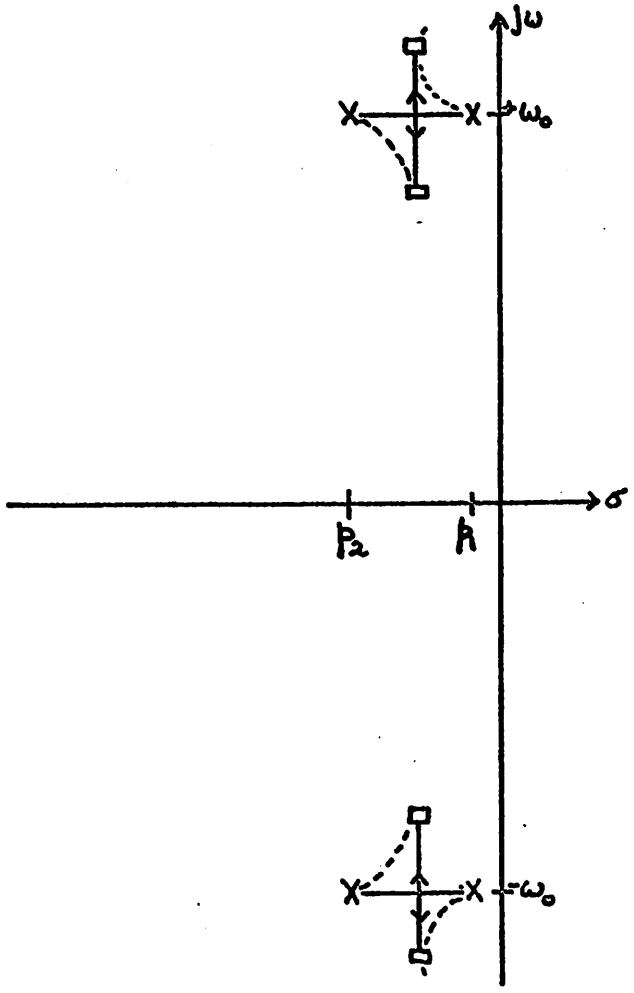


Figure 5.2 Root-locus plot for second-order filter with negative feedback showing ideal case (solid line) and excess phase case (dotted).

CHAPTER 6. EXPERIMENTAL STUDIES.

6.1 Integrated circuit design and fabrication.

With the exception of MOS capacitors, if required and a small number of adjustable components, all circuitry for the frequency-translation filter may be formed in monolithic silicon chips. The transistors are all double-diffused npn types and the resistors, which lie in the range 30 ohm to 3,000 ohm are formed simultaneously with the base diffusion of the transistors. The components are interconnected by conventional aluminum metallization and the dies are gold-bonded to TO-5 headers. A standard processing schedule, which is compatible with the capabilities of this laboratory, has been developed and is given in Appendix 1. Yields obtained with this process are typically 85% in terms of individual transistors for the small quantities which have been fabricated. Representative parameters of transistors within the integrated circuits are given in Table 6.1 and are comparable with those of commercial units, except for their high saturation resistance r'_c, which is due to the absence of a "buried" diffusion layer beneath the epitaxial layer[28].

As noted in Sec. 6.2 below, the selective amplifier system

is divided, for experimental convenience, into smaller functional blocks which are fabricated separately and mounted on separate headers. The basic electronic principles of their design have been developed during the course of the previous chapters and the following points, which relate these principles to the integrated circuit parameters are set out:

(a) The maximum allowed circuit area, from photographic limitations, is a 70 mil square. For more precise registration, however, this is reduced to a 35 mil square so that multiple-image masking may be used*.

(b) The transistor geometry is based on a 2 mil square emitter, with one or two base contacts, depending on the area available. The standard geometry is shown in plan and cross-section in Fig. 6.1.

(c) Pairs of transistors which should be matched with respect to each other (e.g. the multiplier pairs) are formed as close to each other as possible and care is taken to make their geometries nearly identical.

(d) The design-center base sheet-resistance is 150 ohm/sq. Resistors of more than one square are formed as rectangular strips 1 mil wide, while small resistors of a fraction of a square are made on square diffusion patterns, with strip contacts.

---

* The multiple-image masking scheme is designed to reduce errors resulting from the step-and-repeat stage in the photographic reduction.[29] The full mask area is divided into four quadrants in which the isolation, base, emitter and contact patterns are placed in order. The same mask is used four times, with an appropriate shift relative to the wafer each time, so that the patterns retain their correct relative positions, irrespective of any errors in the step-and-repeat.

(e) The isolation walls are 1 mil wide, normally, and are separated from the circuit elements by no less than 2 mils on all sides.

(f) The area of metalization is minimized to reduce stray capacitance and external connections are brought out to bonding pads.

The three types of functional blocks fabricated for this system are: the input multiplier, the lowpass gain stage-output multiplier cascade, and a general-purpose two-stage differential amplifier. The dc voltage levels in these stages are chosen so that they may be direct-coupled in cascade to form a system channel. The voltage drop across each stage is therefore minimized in order that the total system may operate from a moderate supply voltage. A collector-emitter voltage of not less than 2.5 volt is maintained for each transistor.

Three experimental versions of the input multiplier have been fabricated and the best is shown in the photomicrograph and schematic of Fig. 6.2. The 30 ohm base resistors are chosen as a compromise between reasonable input impedance level and the condition given by (3.12). The emitter resistor of 300 ohm supplies a total dc current (I_E) of 4 mA to the two transistors, and the local oscillator signal is applied at the common-emitter node E. The 1,400 ohm collector resistors are chosen to provide an adequate transfer gain and to maintain a sufficiently high impedance level for the convenient realization of the lowpass filter. If these resistors are made too large, however, the loading interaction with the

lowpass stage makes the gain of the latter very β_o dependent. As calculated in Secs. 3.2 and 3.3, the transfer coefficients are $P_o = 12.1$ and $Q_o = 48$ with a maximum frequency for correct operation of 1.6 Mc/s. The balance sensitivity S_b^T is, in a typical case, 10 $\mu V/^oC$. (See Sec. 3.2).

The lowpass gain stage-output multiplier block is shown in Figs. 6.3(a) and (b). The output multiplier operates under the same dc conditions as the input multiplier, though its emitter supply resistor is made external, because of the lack of available space on the circuit die. The lowpass gain stage is biased at an emitter current of 5 mA per side and has a voltage gain of 0.45 (See (4.1).).

The two-stage differential amplifier, shown in Figs. 6.4(a) and (b), was designed originally for capacitance multiplier studies (See Sec. 4.1). The transistors operate at 2 mA emitter currents and the resistors are chosen in accordance with time-constant requirements. This block is also used, after the removal of some interconnections, to provide extra gain in the lowpass stage for feedback studies.

6.2 System realizations.

As an initial check on the theory of the frequency-translation filter, a discrete-component realization was constructed. Measurements taken on this system confirmed the predictions for fixed-parameter operation and assisted in the design of the integrated version. Poor thermal coupling in

this prototype, however, did not permit the taking of meaningful sensitivity results.

Three integrated versions of the system, using the functional blocks described above, have been constructed. The summing amplifier in these versions, being peripheral to the basic operation of the system, is a discrete-component circuit, although it may also be integrated without difficulty.

The first version was designed to examine the performance and sensitivity of the system without feedback, and its circuit is shown in Fig. 6.5. The functional blocks in this diagram are marked as follows; input multiplier: IPM, lowpass amplifier: LPA, output multiplier: OPM, summing amplifier: SA and phase-shift network: PS. The integrated circuit blocks are shown enclosed by dotted lines. The input signal is applied through a resistive divider composed of the 3.3 kΩ series resistor and the 30Ω base resistors of the input multipliers. Each channel is direct-coupled and the potentiometers R_{v1} -R_{v4} provide balance and drift compensation as described in Sec. 5.3. The outputs of the two channels are summed across the 220Ω collector resistors.

The phase-shift network is dc isolated from the channels for simplicity, but in an integrated version the blocking capacitors may be replaced by breakdown diodes or isolating transistors. The potentiometer R_{v5} is provided to match the transfer functions of the two channels (See Sec. 4.3.), and the local oscillator signal- at an arbitrarily chosen frequency of 1.5 kc/s- is supplied by an external signal generator.

The summing amplifier, which is designed according to the principles of Sec. 4.2, uses a low-Q (<10) LC filter to attenuate the second-harmonic components in the output, but a pure RC filter is described below for the third version. The power-supply divider is conventional and is replaced in later versions by a simpler resistive divider. This version of the system is designated "A".

The second version, A', is the same as type A with the exception that the ceramic capacitors are now replaced by multiplied-capacitor blocks, as shown in Fig. 6.6. The bandwidth for this version is wider, because of the smaller RC time-constants, but in other respects the performance is similar to that of A.

The third version, shown in Fig. 6.7 and designated "B", incorporates a linear negative feedback loop, which provides a midband loop-gain $K_o U = 8$. The functional block notation is the same as in Fig. 6.5. The feedback loop is connected in shunt with the 3.3 kΩ input resistor, as shown, so that a current summation takes place across the 30 Ω input base resistors. The forward-path phase-shift is 180^o. As noted in Sec. 5.3, the addition of a feedback loop requires an increase in the lowpass stage gain and this is provided by an additional linear gain stage following the input multiplier. The resultant voltage gain ot the lowpass stage is 7. No drift compensation is incorporated in this version.

The phase-shift network is the same as that used for the "A" system and the revised summing-amplifier includes the RC

notch filter of Fig. 4.4 and two additional gain stages. The output stages incorporates a local shunt-feedback loop to reduce the output impedance.

The measured performance of these versions is set out in Table 6.2 and the frequency-response plot of Fig. 6.8.

Comparing the results in Table 6.2, we first note the differences in the gain and bandwidth of the systems which have merely resulted from design improvements. The gain of the feedback version is higher as a result of the additional stages in the lowpass and summing amplifiers, and its bandwidth has been reduced below that of systems A and A' to demonstrate the fact that this may be done without impairing sensitivities. The bandwidths of systems A and B agree to within 6% of the calculated values, which are 4.9 kc/s and 2.5 kc/s, respectively (See Sec. 5.3.).

It is seen that the dynamic ranges of A and B are approximately equal.

The power supply sensitivity is expressed as the ratio of the output signal (at ω_o) caused by multiplier imbalance, to the maximum output signal, for a one mV change in supply voltage. As expected, the application of feedback improves this sensitivity by about a factor of 10. The changes in the summing amplifier and lowpass stages between the two designs have changed the sign of the gain sensitivity. It may therefore be deduced that there exists a compromise design with a very low sensitivity, a fact which was predicted from (5.10). The difference in bandwidth sensitivity is not significant, because

of the errors inherent in such a measurement, but the improvement in the sensitivity of distortion, both to temperature and phase error; again illustrates the value of feedback in this system. The temperature sensitivity of distortion, calculated on the basis of phase error alone, is approximately 200 ppm/^{o}C, so that these results indicate that the dominant cause of the temperature sensitivity is a difference between the gain coefficients of the two channels. This is reasonable in view of the fact that the functional blocks for the two channels are mounted in separate headers.

The center frequency of system B may be varied by a few per cent without appreciably affecting the distortion level.

Finally, we note a dramatic difference in the multiplier balance sensitivity. The sensitivity predicted in Sec. 5.1 is 23,000 ppm/^{o}C, which is based on a worst-case assumption. In both systems, the sensitivity is considerably better than this value. The theory also predicted that feedback should have no effect on this parameter, but system A includes drift compensation, which is not included in B. This result therefore provides a measure of the effect of this compensation.

Table 6.1

Representative Integrated npn Transistor Parameters
($T=300^{o}K$)

Collector breakdown voltage V_{CBO} (V)	30 - 60
Collector leakage current I_{CBO} (nA) at V_{CB} = 10 V.	9 - 15
Emitter breakdown voltage V_{EBO} (V)	7 - 8
Emitter leakage current I_{EBO} (nA) at V_{EB} = 4 V.	100
Collector output capacitance C_{ob} (pF) at V_{CB} = 6 V.	6.5
Small-signal parameters at I_C = 3 mA, V_{CE} = 6 V.	
Low-frequency base-collector short-circuit current gain α_o.	65
Short-circuit c-e current gain 0 dB frequency f_t (Mc/s)	150*
Internal c-b feedback capacitance C_μ (pF)	5
Collector output resistance r_o (kΩ)	30
Collector series resistance r_c' (Ω)	700
Ohmic base resistance r_x (Ω)	140

* Based on extrapolation to $r_c' = 0$.

Table 6.2

Bandpass amplifier performance

Parameter	System A (without feedback)	System B (with feedback)
Center frequency f_o (Mc/s)	1.52	1.59
Bandwidth at -3 dB (kc/s)	4.61*	2.65
Bandwidth at -30 dB (kc/s)	120	37
Effective Q (f_o/BW)	328*	600
Max. output signal for 5% distortion (V pk-pk)	1.5	2.5
Sensitivity of multiplier balance to power-supply voltage (ppm/mV)	3000	300
Temp. sensitivity of gain S_K^T (ppm/°C)	+4100	-3000
Temp. sensitivity of bandwidth (ppm/°C)	-2000	-3000
Temp. sensitivity of distortion S_{Dm}^T (ppm/°C)	10,000	1000
Multiplier balance drift sensitivity S_d^T (ppm/°C)	500	10,000
Phase-error sensitivity of distortion D_ε (%/rad)	100	10
Temperature range for measurements (°C)	0 - 60	0 - 60

* For system A', with multiplied capacitors, the bandwidth and Q are 23 kc/s and 66, respectively.

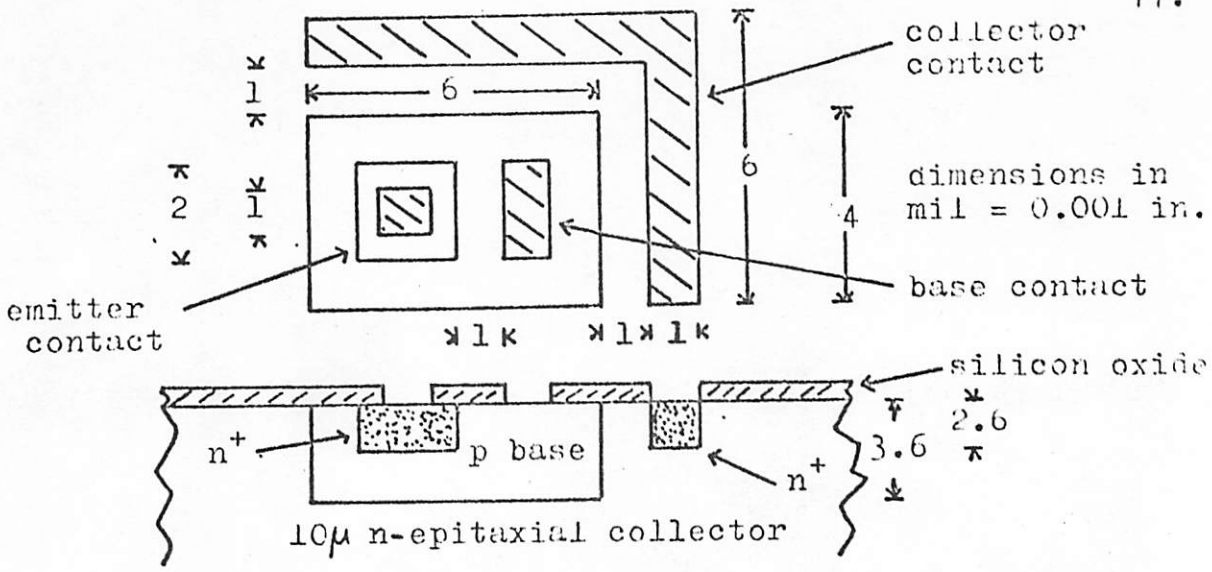
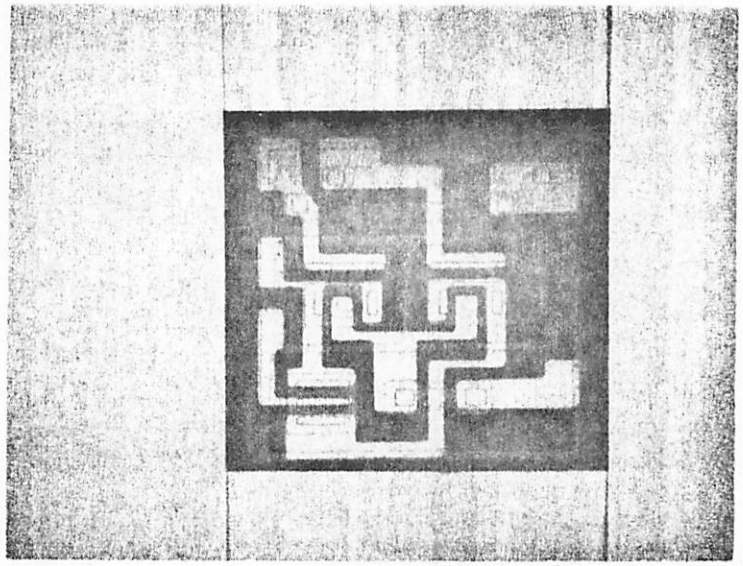


Figure 6.1 Plan and cross-section of typical npn integrated circuit transistor geometry.



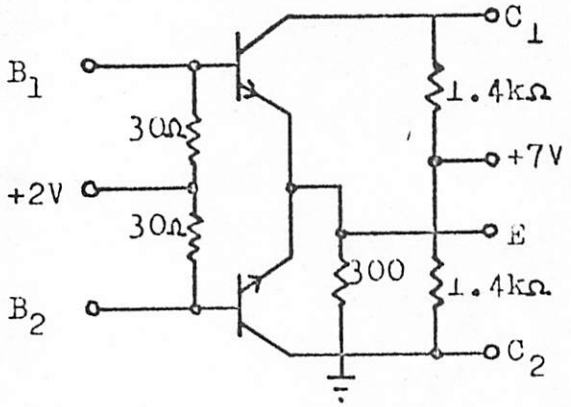
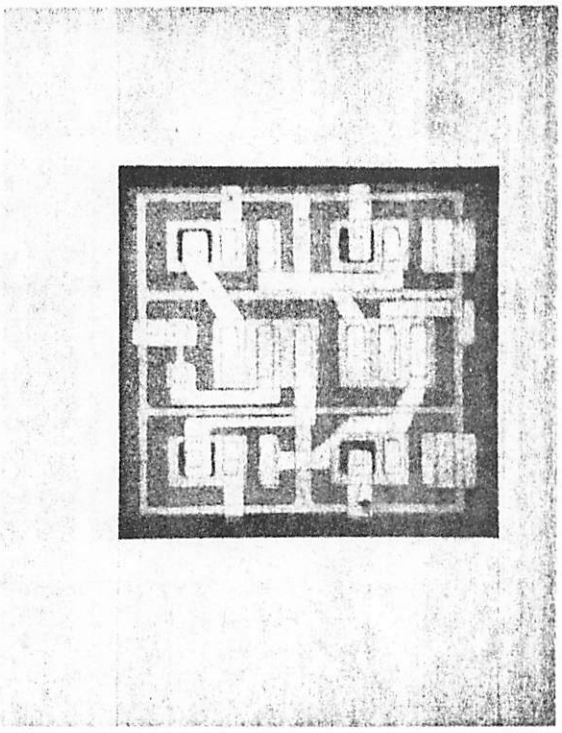


Figure 6.2 Photomicrograph and schematic of integrated input multiplier.



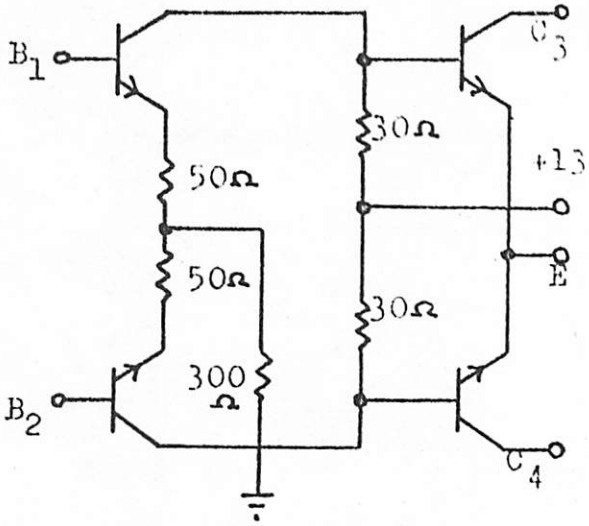
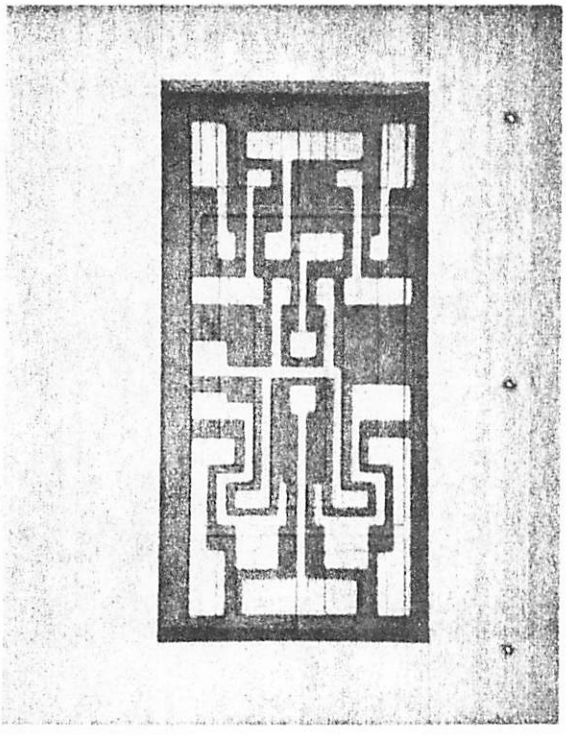


Figure 6.3 Photomicrograph and schematic of lowpass gain stage-output multiplier cascade



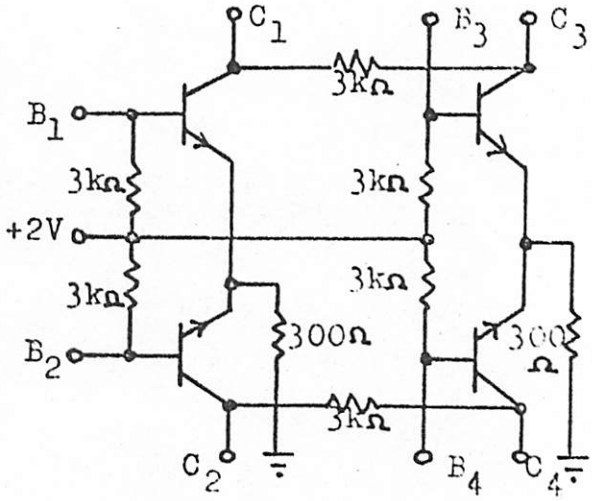


Figure 6.4 Photomicrograph and schematic of two-stage differential amplifier.

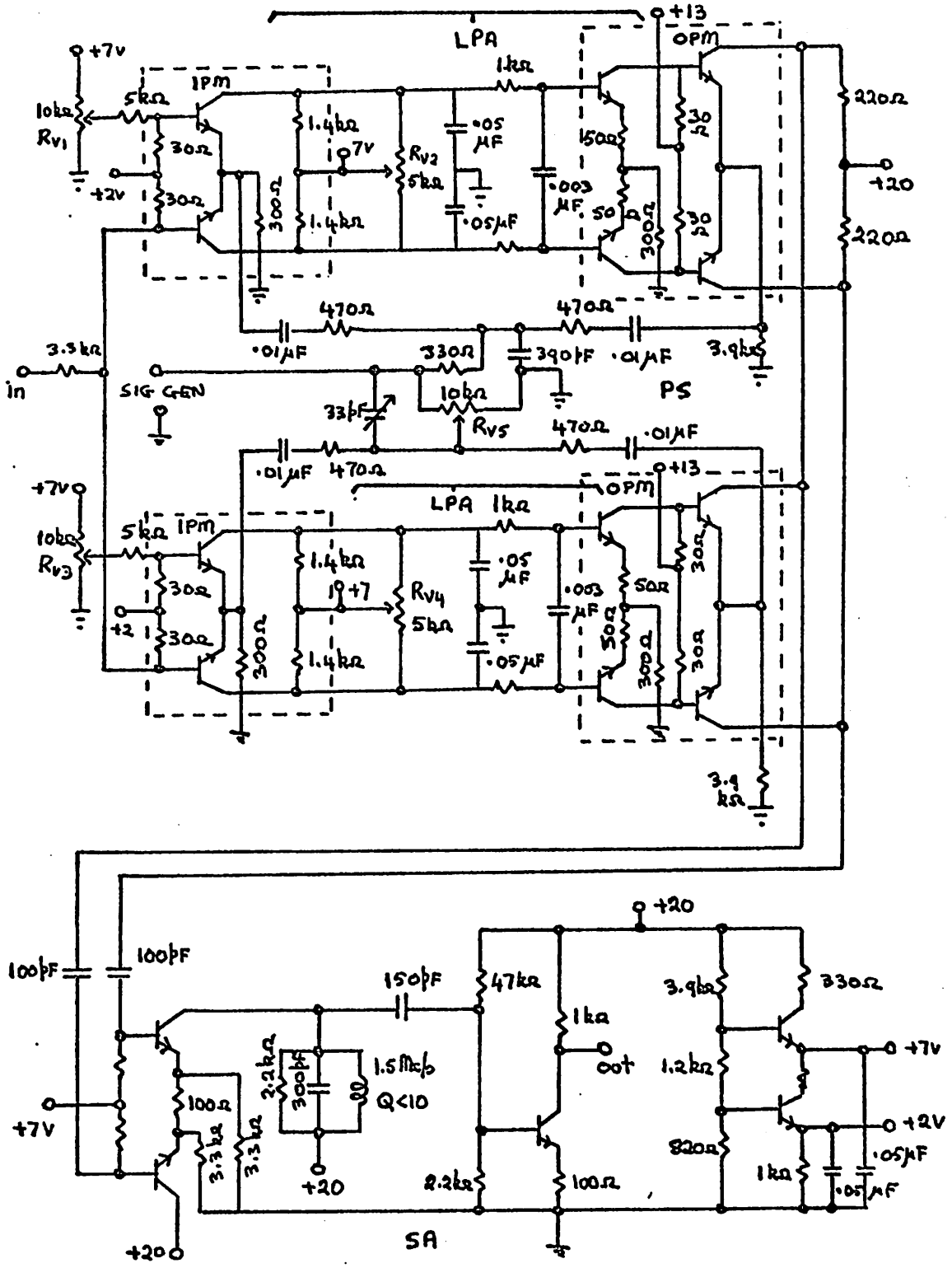


Figure 6.5 Prototype selective amplifier (System A).

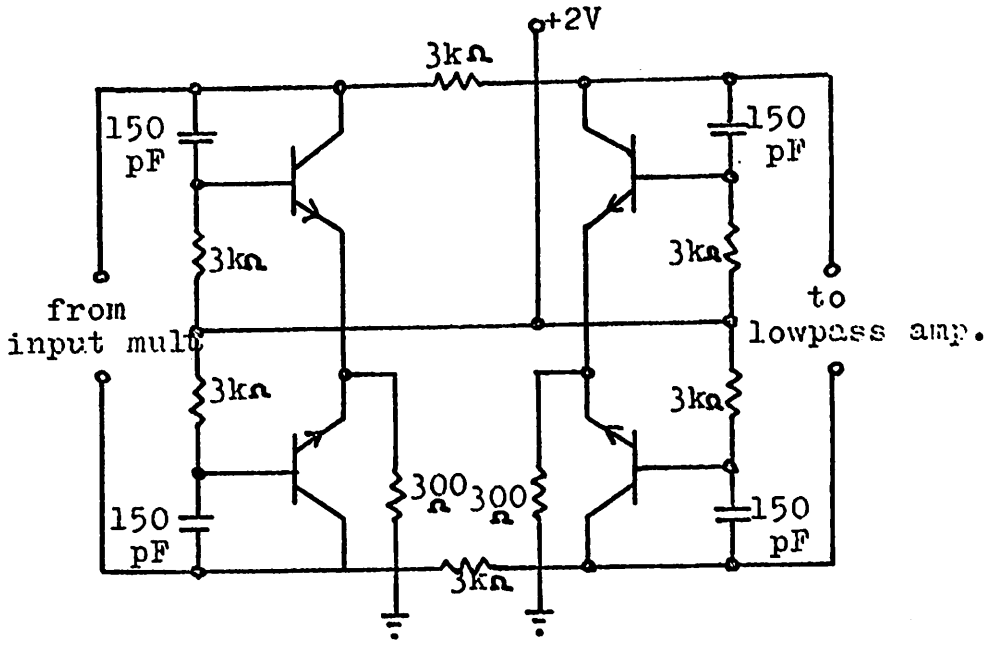


Figure 6.6 Multiplied capacitance stage (one channel) for System A'.

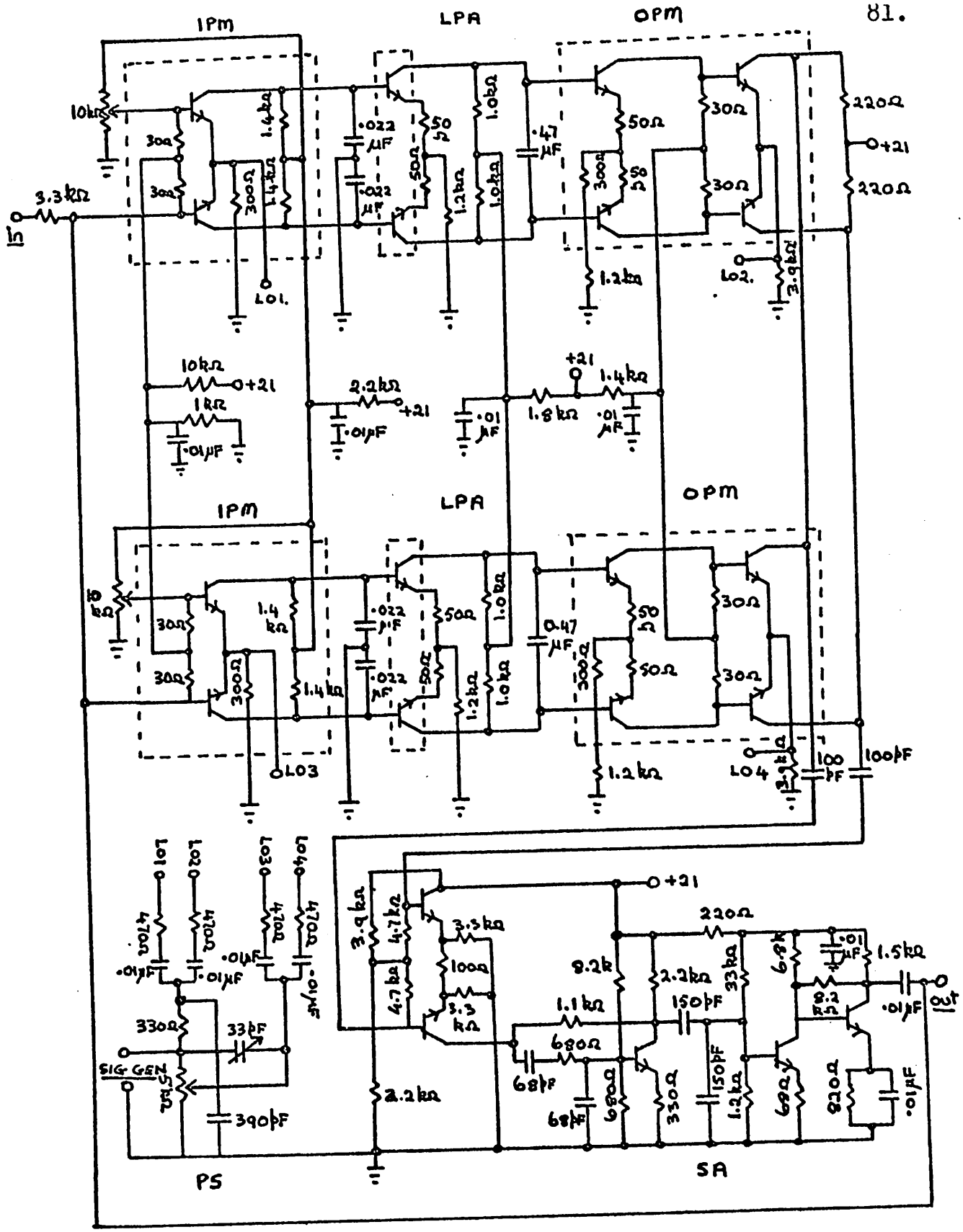


Figure 6.7 Prototype selective amplifier with negative feedback loop. (System B)

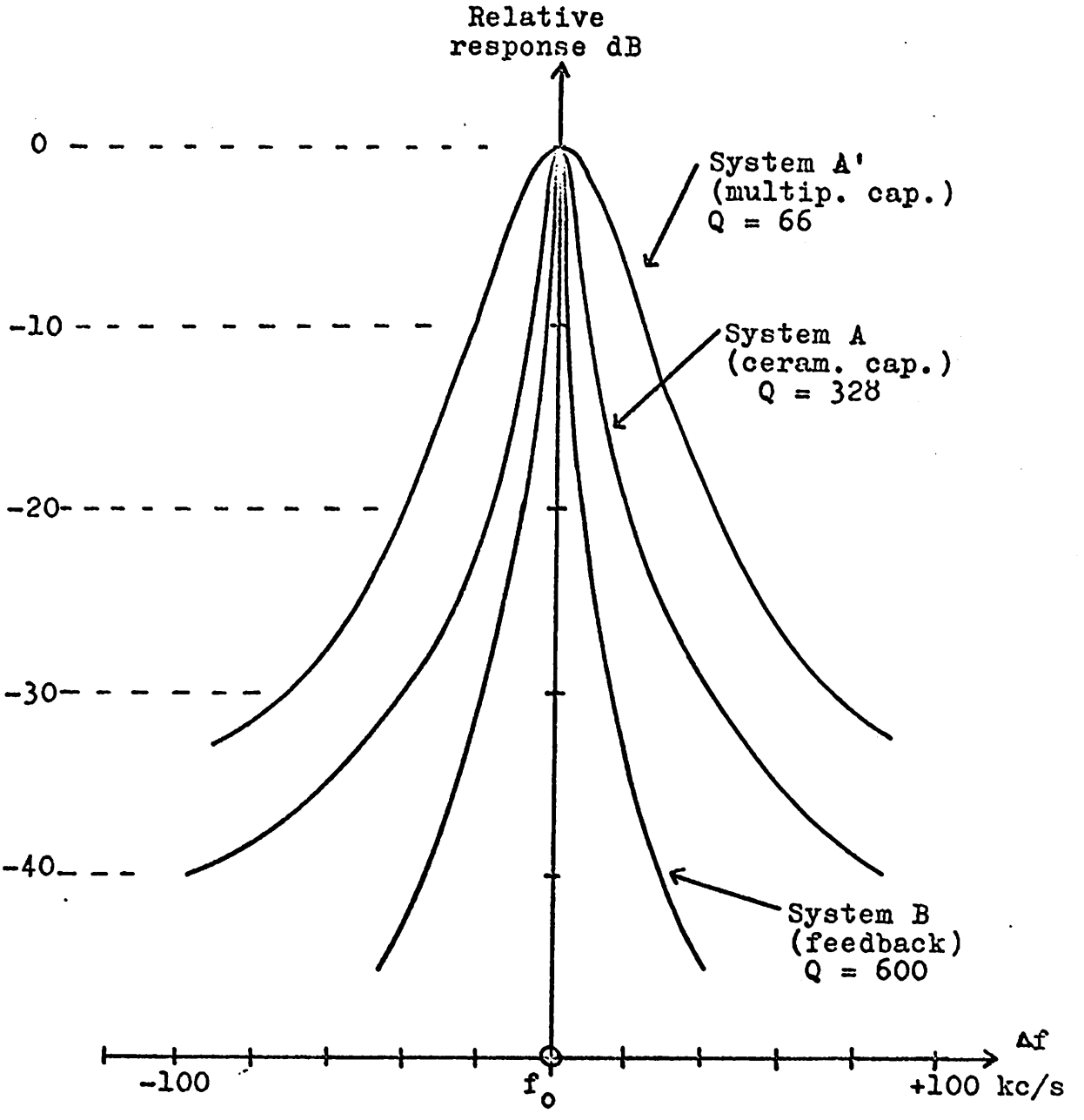


Figure 6.8 .Measured frequency response of three prototype amplifiers. ($f_o \simeq 1.5$ Mc/s.)

CHAPTER 7. SUMMARY AND CONCLUSIONS.

The greater part of this work has been concerned with one basic way of realizing a frequency selective amplifier characteristic in integrated circuit form. It has been seen that, although the frequency translation principle and its application to this problem is by no means new, the way it has been used in this work has some novel aspects. In particular, the development of a simple analog multiplier which can be realized in monolithic form has made possible the design of a circuit which overcomes many of the disadvantages inherent in the switching-type of frequency translation filter.

A detailed examination of the system has been presented and the results of this have led to an understanding of the design requirements which must be fulfilled to optimize its performance with respect to dynamic range and desensitivity.

Throughout the analysis it has been assumed that the ratio of the bandwidth of the filter to the center frequency is very small. This condition is not strictly necessary, however, though it makes the design of most parts of the circuit somewhat easier than in a broadband case. A broadband filter may be realized, in general, provided that the skirt selectivity of the lowpass stage is sufficent to validate the assumptions originally made in the analysis of Sec. 2.2. The most important uses of the frequency-translation filter, however, are in narrow-band

applications.

The experimental studies, which have involved the construction of three prototypes of this system in addition to the measurement of component characteristics, have supported the predictions made by the theory and have provided additional information on aspects of the performance which could not be predicted. A high-yield double-diffused device processing schedule (App. I) has been developed and has since been used with success by a number of workers in this laboratory.

The specific conclusions with regard to this selective amplifier are as follows: The first-order independence between gain, center frequency and bandwidth is its outstanding property which enables passbands with a wide variety of characteristics to be realized. Because there is virtually a one-to-one correspondence between the lowpass functions and the bandpass functions which they generate in the system, it should be possible to synthesize bandpass functions with greater accuracy than is possible in linear active RC schemes. If feedback is used, however, it is necessary that its effects on the transfer function be known precisely. Furthermore, the narrowness of the passband depends only on the availability of suitably long time-constants for the lowpass stage. Tuning of the amplifier is accomplished simply by varying the local oscillator frequency and this may be done over a small frequency range without any other adjustment in the system. Wider variations in center frequency, however, must be accompanied by a readjustment of the phase-shift

network, if distortion is to be minimized. Comparisons of the system with and without negative feedback show that the improvement of performance with feedback well justifies its use.

In the design of the system it is desirable that the circuit be realized on as few silicon dies as possible, and that these dies should be mounted in such a way as to obtain very close thermal coupling between them. Thermal tracking is also enhanced by minimizing the power dissipated in the components and, since three or more stages are direct-coupled in each channel, the voltage drop across each stage should be minimized. The basic design relations for the individual functional blocks have been given and it appears, on the basis of the parameters of typical integrated circuit components, that system center frequencies of up to 10 Mc/s may be obtained without difficulty. From the work reported here, it is not possible to predict, with any confidence, the ultimate performance of the system; but it is certain that a significant improvement in performance would result from the use of a smaller number of circuit dies and finer tolerances on circuit processing.

The main obstacle to monolithic realizations has been seen to occur in cases where the center frequency or bandwidth require the use of long time-constants in the lowpass and phase-shift sections. If these time-constants are beyond the range of simple MOS capacitor-diffused resistor products, either a multiplied-capacitance or hybrid construction may be

used.

In comparing this scheme with the linear active RC class of selective amplifiers, (in particular, the circuits described in Refs. 5 and 6) two notable disadvantages become apparent. In the first place, the frequency translation system is more complex, in terms of components, by a factor of 5 to 10. In the second place, it requires more adjustments than other types although, at the present state-of-the-art in integrated circuits, no high-performance, adjustment-free circuit is possible. For these reasons it is not realistic to imagine that this system is appropriate for the large number of "entertainment" and commercial applications whose relatively modest requirements may be met by simpler circuits. However, this system has properties which are, perhaps, unique in this area and in more demanding applications such as in intrumentation and communications, the special properties of the frequency translation filter may be used to advantage.

APPENDIX I: PROCESSING DETAILS.

The circuits described above are fabricated in 12 micron 1 ohm-cm n-type epitaxial silicon, commercially grown on a 5 ohm-cm p-type substrate. The processing schedule is given in Table A.1 and is based on a quantity of data accumulated in this laboratory as well as information in the literature*[30,31].

Before the first oxidation step, the wafers are cleaned twice in concentrated sulphuric acid at 95°C and then rinsed in 12% hydrofluoric acid at room temperature. Immediately before all other furnace steps, except aluminum sintering, the wafers are again rinsed for 10 sec. in 12% HF. The withdrawal of wafers from a furnace is always carried out in a dry N_2 atmosphere at a slow rate so that the cooling of the wafer from the hot-zone temperature to room temperature takes about 2 minutes. This step is believed to minimize the concentration of surface-states[32] in the wafer.

The choice of the isolation predeposit and drive-in conditions is influenced by two phenomena which have been observed after the wafer has been processed at high temperatures. When the boron predeposit is carried out at temperatures above 1000°C, the unoxidized silicon frequently

---

* Ref. 30 contains a useful bibliography on this subject.

develops etch-pits which are distributed randomly across the wafer. These resemble a "criss-cross" pattern of scratches a few microns in length, aligned in specific crystalographic directions. Since halides are known to etch silicon at high temperatures,[33] it is assumed that the traces of iodine, produced by the decomposition of the BI_3 doping source in the furnace, are responsible. Although etching can be inhibited by the addition of 1% oxygen to the furnace gas, the problem has been avoided in this case by performing all predeposits below 1000°C. The second phenomenon is the appearance of black irregular spots, typically 5μ across, on the unoxidized silicon during the isolation drive-in (1175°C). These spots, whose exact composition is not known, are very resistant to acid etches and are believed to contain boron or nitrogen compounds of silicon*. Although their presence does not result in any detectable degradation of final circuit performance, they can interfere with subsequent masking operations. Therefore, the exposed silicon is protected during the drive-in by the growth of an oxide film which is kept very thin to minimize gettering of the boron.

After step 7, the emitter drive-in, contact windows are opened on the wafer and the transistors are probed. At this stage the base width is approximately 2 micron and the typical β_o range is 8 - 15. The desired base width of 1 micron is then obtained by a series of trimming diffusions during which the e-b junction is driven down into the wafer at a faster rate

---

* See note on "boron-film" in Ref. 33.

than the c-b junction. The trimming procedure consists of a series of short non-oxidizing diffusions, after each of which the β_o of selected transistors is checked. The trimming is continued until a minimum β_o of 40 is obtained and normally takes 30 to 60 minutes beyond the initial drive-in. Because of uncertainties in junction-depth measurements, the exact base-width is not know, but it has been found that the value of $_o$ obtained by probing provides an indirect but reliable measure of this parameter. This is due to the fact that β_o in these devices is limited by the emitter injection efficiency which, in turn, is related to the base width*. (See Appendix II).

In the aluminum sintering step the withdrawal of the wafers from the furnace is begun 10 minutes after loading and continues over a period of 20 minutes, after which the wafers are almost at room temperature. The necessity for such an annealing process has not been demonstrated conclusively, however, and some workers consider the speed of withdrawal to be unimportant.

Representative parameters of bipolar transistors fabricated according to this schedule have been given above in Table 6.1. Specimens of these transistors have also been angle-lapped and examined with the scanning electron-beam microscope.+

---

* An independant series of experiments, performed by W.G Howard in this laboratory, has shown a strong correlation between β_o and base-width as measured by the sheet resistance of a large-area base diffusion buried beneath a large-area emitter diffusion.

+ These measurements were carried out in this laboratory by N.P McDonald.

Photomicrographs show diffusions of good planarity, but narrow base widths (in the range 0.5-0.7 micron) than those measured using the groove-and-stain technique.(approx. 1 micron.)

Table A.1

Summary of processing steps

Process	Temperature °C	Time and atmosphere*	Sheet resistance (ohm/square)	Junction depth + (micron)
1. Initial oxidation	1150°	30 min wet O_2 3 min dry N_2		0.4(oxide)
2. Isolation predeposit	985°	60 min N_n/BI_3	20-25	
3. Isolation drive-in	1175°	5 min dry N_2 5 min dry O_2 15 hr. dry N_2		>12
4. Base predeposit	950°	30 min N_2/BI_3	48-52	
5. Base drive-in	1175°	10 min wet O_2 110 min dry O_2 3 min dry N_2	150-180	3.6-3.8
6. Emitter predeposit	950°	30 min N_2/P_3N_5	9-12	
7. Emitter drive-in	1100°	10 min wet O_2 5 min dry N_2	4-6	
8. Emitter trim	1100°	2-5 steps of 15 min dry N_2		2.6-3.0
9. Aluminum sinter	525°	30 min dry N_2		

* All flow rates: 1 liter/min. Wet O_2 obtained by passing O_2 through DI water at 95°C.

\+ Measured by groove-and-stain technique.

APPENDIX. II

THE DEPENDENCE OF TRANSISTOR CURRENT GAIN ON BASE-WIDTH.

The dependence of the β_o of a double-diffused transistor on its geometry and doping levels is given approximately by:[35*]

$$\frac{1}{\beta_o} \simeq \frac{\rho_E X_B}{\rho_B L_{pE}} + \frac{X_B^2}{2L_{nB}^2} + F_s + F_{ER} \quad \cdot (A.1)$$

where ρ_E and ρ_B are the mean resistivities of the emitter and collector regions, respectively, X_B is the base-width, and L_{pE}, L_{nB} are the mean diffusion lengths of minority carriers in the emitter and base, respectively. The terms F_s and F_{ER} are surface recombination and emitter depletion-layer recombination terms, respectively. The first term of (A.1) represents the injection efficiency limitation on β_o and the second, the effect of bulk recombination. Calculations based on experimental data have led to the conclusion that with the doping levels which the process of App. I produces, β_o is injection-efficiency limited for base-widths greater than about 2 micron. For narrower base-widths, β_o is probably surface and bulk limited.

To illustrate this, the following two numerical examples, based on experimental results, are presented:

---

* For a more accurate expression, see Ref. 6.

(1) Base: Junction depth X_{BC} = 3.6 micron.

Surface concentration $N_{sB} = 5 \times 10^{18}$ atom cm^{-3}.

(See Ref. 31)

Emitter: Junction depth X_{EB} = 1.0 micron.

Surface concentration $N_{sE} = 5 \times 10^{19}$ atom cm^{-3}.

Base-width X_B = 2.6 micron.

Mean base resistivity ρ_B = 0.1 ohm-cm. (See Ref. 31.)

Mean emitter resistivity ρ_E = 0.005 ohm-cm.

The mean emitter diffusion length is estimated, on the basis of published results[36] as 1 micron.

Substituting these values in (A.1) yields the result:

$$\beta_o < 8.$$

which is in good agreement with measured values, which are in the range 5-10.

(2) The calculation is now repeated for a narrower base-width. In this case X_{CB} = 3.6 micron, X_{EB} = 2.6 micron and X_B = 1 micron. With the same surface concentrations as in (1), the mean resistivities become:

$$\rho_B = 1 \text{ ohm-cm}$$

$$\rho_E = 0.005 \text{ ohm-cm}$$

which lead to:

$$\beta_o < 210$$

Typical measured values in this case are in the range 35-100, which indicates the dominance of surface and bulk recombination.

REFERENCES

1. W.E.Newell: "Tuned integrated circuits - a state-of-the-art survey". Proc. IEEE, Vol. 52, pp. 1603-08. Dec. 1964.

2. J.G.Linvill: "Reactive filters" Proc. IEEE, Vol 42. p. 555, March 1954.

3. M.S.Ghausi, D.O.Pederson: "A new design approach to transistor amplifiers". IRE Trans. Circ. Th. CT-9, pp. 247-83 Sept. 1961.

4. W.M.Kaufman: "Theory of monolithic null device and some novel circuits.". Proc.IRE, Vol. 48, pp. 1540-45, Sept. 1960.

5. A.G.Thiele: "Optimum design of integrable bandpass amplifiers". Tech. Report. 60-456, Electronics Res. Lab. University of California, Berkeley.

6. A.A.Gaash: "Synthesis of inegrated selective amplifiers for specified response and desensitivity". Tech Report 65-31, Electronics Res. Lab., University of California, Berkeley.

7. A.A.Gaash, D.O.Pederson, R.S.Pepper: "Design of integrable desensitized frequency selective amplifiers". ISSCC Digest of Tech. Papers, pp. 34-35, 1966.

8. D.R.Curran, W.J.Gerber, D.J.Koneval, K.A.Pim: "Quartz uni-wafer filters, a thin-film device". IEEE Wescon Record Paper 18:2, Aug. 1964.

9. H.C.Nathanson, R.A.Wickstrom: "Aresonant-gate silicon surface transistor with high-Q bandpass properties". App. Phys. Letters, pp. 84-86, Aug. 15th. 1965.

10. W.E.Newell: "Novel circuit aspects of the resonant gate transistor". ISSCC Digest of Tech. Papers. pp. 62-63, 1966.

11. M.Schwartz: "Information Transmission, Modulation and Noise". McGraw-Hill, N.Y. 1959. p. 160 ff.

12. N.F.Barber: "Narrow bandpass filter using modulation". Wireless Engineer, Vol 24, pp. 132-34. May 1947.

13. C.A.Desoer: "A network containing a periodically operated switch solved by successive approximations". BSTJ Vol. 36, pp. 1403-28, Nov. 1957.

14. W.R.LePage, C.R.Cahn, J.S.Brown: "Analysis of a comb filter using synchronously commutated capacitors". AIEE Trans.(Commun. and Electronics), Vol. 72, pp. 63-68. March 1963.

15. L.E.Franks, I.W.Sandberg: "An alternative approach to the realization of network transfer functions: the N-path filter". BSTJ Vol. 39, pp. 1321-50, Sept. 1960.

16. A.Acampora, B.Rabinovici, C.A.Renton: "Generation of bandpass filters by switching techniques". Proc. IEEE Vol. 51, pp. 256-57, Jan. 1963.

17. J.R.Biard, W.T.Matzen: "Advanced functional electronic block development". Report RTD-TDR-63-4203. (Wright-Patterson AFB) pp. 264-98.. Jan..1964.

18. L.Brunetti: "A Fourier analysis of the N-path filter". M.S. Thesis, University of California, Berkeley. 1965.

19. T.B.Martin: "Circuit applications of the field-effect transistor". (Part 2.) Semiconductor Prod. Vol. 5. pp. 30-38, March 1962.

20. W.H.Highleyman E.S,Jacob: "An analog multipler using two field-effect transistors". IRE Trans. Comm. Syst. CS-10,pp. 311-17, Sept. 1962.

21. D.O.Pederson: "Electronic Circuits". (Prelim. Edition) McGraw-Hill N.Y. 1965, p. 20.

22. R.M.Warner, J.M.Fordemwalt: "Integrated Circuits". McGraw-Hill, N.Y. 1965, p. 194.

23. W.J.Cunningham: "Introduction to Non-Linear Analysis". McGraw-Hill, N.Y. 1958, p. 246.

24. R.M.Warner, J.M.Fordemwalt: op. cit. pp. 252-64.

25. D.O.Pederson: op. cit. p. 233.

26. G.D.Hachtel: "Integrated oscillators". Tech Report 65-32, Electronic Res. Lab. University of California, Berkeley, 1965.

27. G.D.Hachtel, R.S.Pepper: "Synthesis of near-harmonic oscillators". ISSCC Digest of Tech. Papers. pp. 40-41, 1966.

28. R.M.Warner, J.M.Fordemwalt: op. cit. p. 189.

29. G.W.Haines: "Photographic mask techniques". Internal Rept. Electronics Res. Lab., University of California, Berkeley, 1963.

30. S.F.Davis: "Design and fabrication of integrated circuit transistors". M.S.Thesis. University of California, Berkeley, 1964.

31. J.C.Irvin: "Resistivity of bulk silicon and of diffused layers in silicon". BSTJ, Vol. 41.pp.387-410, March 1962.

32. B.E.Deal, A.S.Grove, E.H.Snow: "Surface state charge density of thermally oxidized silicon: effect of oxidation conditions". Elec. Chem. Soc., Spring Meeting, (Cleveland, Ohio.) Paper 16, May 1966.

33. Research Triangle Inst: "Integrated silicon device technology". Vol. 10. (Chem. and metal. properties of silicon) Report AST-TDR-63-316 (Wright-Patterson AFB) 1964, p. 157.

34. ibid. p. 85.

35. R.M.Warner, J.N.Fordemwalt: op.cit. p. 102.

36. H.Kressel, H.S.Veloric, A.Blicher: "Design considerations for double-diffused silicon switching transistors". RCA Rev. Vol. 23. pp. 587-616. Dec. 1962.