FREQUENCY COMPENSATION TECHNIQUES FOR LOW-POWER OPERATIONAL AMPLIFIERS

by

Rudy G. H. Eschauzier
Philips Semiconductors, Sunnyvale, CA, U.S.A.

and

Johan H. Huijsing
Delft University of Technology, Delft, The Netherlands

KLUWER ACADEMIC PUBLISHERS
BOSTON / DORDRECHT / LONDON

Preface

"Frequency Compensation Techniques for Low-Power Operational Amplifiers" is intended for professional designers of integrated amplifiers and graduate students in this field. It serves as a guide to the frequency compensation of integrated amplifiers, emphasizing low-voltage and low-power solutions.

The basis of this text was written as a dissertation at the Delft University of Technology. The aim of the Ph.D. project was to investigate new frequency compensation techniques that were suited for the new generation low-voltage, low-power integrated amplifiers. During the four years of research it soon became apparent that, although frequency compensation is a subject that every integrated-circuit designer is likely to come across, surprisingly few attempts have been made to systematically address this subject.

The book aims at bridging the gap between the professional designer's needs and available techniques for frequency compensation. It does so in the first place by explaining existing techniques. Furthermore, the text covers several new techniques which are the direct results of the Ph.D. study. Examples of the latter are Hybrid Nested Miller compensation (Sect. 5.4), Multipath Miller Zero cancellation (Sect. 6.4) and Multipath Conditionally Stable compensation (Sect. 6.5). All compensation techniques are treated in a stage-number based order, progressing from a single transistor to circuits with six stages and more. Apart from discuss-

ing the mathematical basis of the compensation methods, the book provides the reader with the factual information that is required for practicing the design of integrated feedback amplifiers and many worked out examples. What is more, many bipolar and CMOS operational amplifier realizations, along with their measurement results, prove the effectiveness of the compensation techniques in real-life circuits.

In line with the current trends in analog circuit design, the text focuses on low-voltage, low-power integrated amplifiers. Many of the presented bipolar circuits operate at supply voltages down to 1V, while several CMOS amplifiers that function correctly just slightly above this voltage are demonstrated. The lowest measured power consumption amounts to 17μW for a class AB CMOS opamp with 120dB gain. Despite this attention to low-voltage and low-power, the frequency compensation strategies provided are universally applicable. The fundamental approach followed leads to efficient compensation strategies that are well guarded against the parameter variations inherent to the mass-fabrication of integrated circuits.

The book is divided into three main parts. The first part (chapters 1,2 and 3) is optional. The interested reader will find the basic theory of negative feedback amplifiers here. The second and the third parts, spanning chapters 4 through 7, establish the core of the work. Of these, the second part (chapters 4 through 6) introduces the various compensation techniques, including the theory and several simple examples, while the third (chapter 7) discusses the worked-out silicon realizations and the measurement results. Although the last two parts build upon the theory of part one, they are intended to be usable without prior reading of the first.

Ruud G.H. Eschauzier
Johan H. Huijsing

Delft, January 4, 1995.

Contents

List of Symbols

symbol	quantity	unit
β	backward gain	-
β_f	current gain of a bipolar transistor	-
Δ	system's determinant	-
Δ_{ij}	cofactor of the system's determinant	-
Γ	contour in complex plane	-
μ	charge carrier mobility	cm^2/Vs
ω	frequency	rad
ω_0	useful bandwidth	rad
ω_a	transit frequency of the gain asymptote	rad
ω_b	intercept frequency	rad
ω_c	secondary pole frequency	rad
ω_d	doublet frequency	rad
ω_t	transit frequency	rad
σ	real part of complex frequency	rad
τ	time constant	s
a	settling accuracy	-
a_1	first Taylor expansion component	V,A
A	forward gain	-
A_0	forward gain at D.C.	-
A_2	second order harmonic distortion component	V,A
A_m	ultimate gain limit	-
c	capacitance	F
c_l	load capacitance	F
c_m	Miller capacitance	F
c_p	interstage capacitance	F

C	symbol for capacitor	-
C_{ox}	specific capacitance MOS device	F
D_2	second order relative harmonic dist. comp.	-
D_I	total intermodulation distortion	-
D_{I2}	second order relative intermod. dist. comp.	-
D_T	total harmonic distortion	-
E	signal	V,A
$\hat{E}$	amplitude of sinusoidal signal	V,A
F	return difference	-
g_m	transconductance	Ω^{-1}
G	closed loop gain	-
GB	gain-bandwidth product	-
$H(s)$	arbitrary complex function	-
i	small-signal current	A
$\grave{i}$	small-signal current vector	A
$\overline{i_{eq}^2}$	squared equivalent input noise current	A^2/Hz
$\overline{i_{nb}^2}$	squared equivalent base noise current	A^2/Hz
$\overline{i_{nc}^2}$	squared equivalent collector noise current	A^2/Hz
$\overline{i_{nd}^2}$	squared equivalent drain noise current	A^2/Hz
I_b	base current	A
I	current	A
I_c	collector current	A
I_e	emitter current	A
I_d	drain current	A
I_s	source current	A
$j\omega$	imaginary part of complex frequency	rad
J	current density	A/m
k	Boltzmann's constant, $1.3805 \cdot 10^{-23}$	J/K
k	bandwidth reduction factor	-
k	normalized unity gain factor	-
K	gain constant	-
K_1	first unity gain factor	s^{-1}
m_{11}	(1,1) entry of the nodal-equations matrix	Ω^{-1}

M	nodal equations matrix	Ω^{-1}
M	symbol for MOS transistor	-
n	slope of gain asymptote	20dB/dec
N	number of stages	-
p	pole	s^{-1}
p_1'	bandwidth limiting pole	s^{-1}
P	power	W
P_s	supply power	W
$P(s)$	arbitrary complex function	-
q	electron charge, $1.6 \cdot 10^{-19}$	C
Q	symbol for bipolar transistor	-
R	symbol for resistor	-
s	complex frequency	rad
S_d	doublet spacing	-
S/N	signal-to-noise ratio	-
S_{ij}	plus or minus sign	-
S_A^G	sensitivity of G to A	-
T	return ratio	-
T	absolute temperature	K
T	symbol for generic transistor	-
T_s	settling time	s
u	transformed real part of frequency	-
v	small-signal voltage	V
v	transformed imaginary part of frequency	-
$\bar{v}$	small-signal voltage vector	V
$\overline{v_{eq}^2}$	squared equivalent input noise voltage	V^2/Hz
V	voltage	V
V_{dsat}	saturation voltage of a MOS device	V
$V_M(J)$	MOS voltage	V
V_{th}	threshold voltage of a MOS device	V
V_T	thermal voltage	V
w	constituent of an immittance	-
w	specific MOS channel width	m/A

W	width of MOS channel	m
x	gain margin	dB
y	small-signal admittance	Ω^{-1}
y	phase margin	180°
Y	admittance	Ω^{-1}
z	small-signal impedance	Ω
z	zero	s^{-1}
Z	impedance	Ω

1 *Introduction*

The rapidly increasing integration densities of I.C. technologies has forced some fundamental changes of direction in the field of analog integrated electronics. For years, the circuit design almost exclusively focussed on the improvement of the signal processing properties. Demands on accuracy, offset and noise, etc., led the appearance of new circuit generations. Although these properties remain of importance today, the design emphasis has shifted from the basic qualities towards the ability of circuits to operate at a low supply voltage, consuming a minimum amount of power. The lower supply voltage leads to drastic changes in the way circuits are built up, since many of traditional circuit solutions become obsolete and need replacement.

Associated with the transition to low-voltage low-power circuit design is a demand for frequency compensation techniques that are dedicated to the new context: compensation strategies which decrease the power consumption, but above all which are capable of handling the specific demands of the low-voltage amplifier topologies. A particular consequence of a low supply voltage is that the use of emitter and source followers and, to a lesser degree, cascodes is in many cases prohibited. This is due to the relatively high voltage that is required for their operation. Additional gain stages are necessary to counteract the gain loss caused by the absence of buffers. In order for frequency compensation methods to be useful they must be able to attack this high stage count.

Moreover, since many of today's applications are more demanding than ever before, they are not allowed to sacrifice basic signal processing properties. The main drive for the increasing circuit demands is the merging of analog and digital electronics, either in a system or on a single chip. The analog parts often serve as the head and the tail of an otherwise digital signal processing chain. The urge to match the high level of accuracy and dynamic range of the digital processing challenges the design of the analog circuits.

1.1. Low voltage and low power

Three apparent developments initiate the trend towards low voltage and low power. The first is the advent of complex battery powered portable equipment. To establish an acceptable operation period from a minimum sized battery, for applications such as note book computers and cellular telephones a low power consumption of the employed circuitry as well as a low minimum supply voltage are essential.

The second reason stems from the smaller feature sizes that lie at the base of the shrinking I.C. processes. The smaller dimensions lead to an increase of the electrical field strengths on the chip to a level where they start destroying the various components. The gate oxide of recent CMOS processes, for instance, are reported to break through at voltages of 3 volts and less. Compensating the effect of the smaller dimensions on the electrical fields requires lowering of the supply voltage.

The third and final development that underlies the shift towards low voltage and low power is related to the limited amount of power that a silicon chip can dissipate without overheating. As the density of the components on a chip increases, the power dissipation per component must decrease. This is to ensure that the maximum heat energy which the package can convey from the silicon die to the ambient is not exceeded. Keeping the total power dissipation under control requires power efficient circuit design.

1.2. Low-power frequency compensation

The trend towards lower power and especially lower voltage impels circuits that are able to operate within the new boundaries. In many cases this means that well-established circuit solutions need to be reconsidered. New circuit concepts must replace topologies that proved their effectiveness over the years. With this transition the techniques used for the frequency compensation of amplifiers also require careful re-examining.

When applying feedback to an amplifier without consideration of its frequency behavior, chances are that the circuit will fall into spontaneous oscillations. Frequency compensation measures intend to prevent this by shaping the frequency response in such a way that the circuit is stable under a large range of conditions. For general purpose amplifiers, such as opamps, these conditions include, apart from temperature and processing variations, different types of feedback networks and varying load impedances. The large number of variables involved has led opamp designers in the past to minimize the number of gain stages, keeping the frequency compensation as simple as possible. To maintain a sufficient amount of gain without affecting the frequency behavior, emitter and source followers were applied. An illustrative example is the well-known Darlington composite, comprising a common emitter gain stage preceded by an emitter follower. The Darlington behaves as a single transistor with a much higher current gain, reducing the internal loading of the circuit. Another classical approach to increase the gain of a circuit without deteriorating the frequency characteristic is the insertion of cascode stages. This technique is widespread for CMOS amplifiers, since the high output impedance of a cascode driving a MOS gate can accomplish very high voltage gains.

Unfortunately, voltage followers and cascodes bear a supply voltage penalty that render them undesirable and in some cases even unsuited for the new low-voltage low-power environment. The Darlington example shows this very clearly. The emitter follower causes the minimum supply voltage of the composite to rise, compared to the minimum for a single transistor, by its base-emitter voltage. The supply voltage penalty of a cascode transistor is related to the saturation voltage that is necessary to provide the device with enough head room for its operation. Bipolar transistors have a relatively low saturation voltage, so in this technology the increase of the minimum supply voltage is not much of an issue. CMOS cascodes, on the other hand, display higher saturation voltages and

their effect on the minimum supply voltage of an amplifier can be a serious point of concern.

The unfitness of voltage followers and cascodes in low-voltage circuits forces an increase of the number of gain stages (common emitter and common source configurations) to realize sufficient gain. It is this large number of stages that challenges the designer in finding new methods for the frequency compensation. For these methods to be applicable in general purpose circuits, they should match their fewer-stage predecessors in robustness against variations of the load impedance and the feedback network. But above all, they must at least be as efficient with respect to the supply power. To be able to compare the virtues of the different compensation techniques, this work follows the red line of the *bandwidth-to-power* ratio. Comparing the bandwidth of an amplifier to the power consumption necessary to realize it gives a valuable indication of the effectiveness of the used frequency compensation technique.

1.3. Organization of the work

Apart from this introduction, six chapters discuss the frequency compensation techniques for low-voltage and low-power operational amplifiers. As a mathematical basis, Chapter 2 starts off with an overview of the network theory associated with negative feedback amplifiers. Not only is attention paid to basic aspects such as sensitivity, noise and the impedances of feedback networks, but also to the estimation of non-linear distortion levels through the use of the linear theory.

Chapter 3 introduces the theory involved with the stability of linear time-dependent networks. The Nyquist criterion for stability is given, while the chapter further explains the background of two well-know quantitative measures for stability: the gain and the phase margin. Furthermore, the maximum amount of feedback that can be obtained from a particular amplifier setup is addressed.

Building upon the established theory, Chapter 4 arrives at the basic frequency compensation measures for amplifiers with one or two stages. The chapter begins with the single-stage case, soon to be followed by two-stage amplifiers. Two basic compensation strategies appear: parallel and Miller compensation. A comparison of these two techniques points out

that for most general purpose integrated amplifiers Miller compensation is the preferred method.

Chapter 5 extents the basic compensation techniques of the preceding chapter to amplifiers with three stages and more. The discussed compensation methods are: multi-stage parallel compensation, Nested Miller compensation, Reversed Nested Miller compensation and Hybrid Nested Miller compensation.

To overcome the bandwidth reduction that is typical for the (Nested) Miller compensation strategies, Chapter 6 introduces the Multipath technique. The Multipath Nested Miller compensation topology is investigated, followed by the Multipath Hybrid Nested Miller compensation technique. Further, the chapter proposes the Multipath Miller Zero Cancellation to eliminate the Right-Half Plane zero of Miller compensated amplifiers. A small detour from the absolutely stable path is provided by the final Multipath compensation technique, intended for conditionally stable amplifiers: Multipath Conditionally Stable compensation.

Chapter 7 demonstrates several practical implementations of the frequency compensation solutions covered by this work. Firstly, two precision operational amplifiers with Nested and Multipath Nested Miller compensation respectively are described, followed by an opamp with Multipath Miller Zero Cancellation. The third section presents two ultimate low-voltage CMOS opamps with (Multipath) Hybrid Nested Miller compensation. The fourth and final part of the chapter is reserved for a Multipath Conditionally Stable amplifier.

2 *Properties of Feedback Circuits*

The essence of feedback is to redirect part of the output signal back to the input. There the returned signal is compared to the incoming signal. Any inconsistency between the desired and the actual output value produces an error signal. The nature of the feedback loop helps reducing the error to a minimum. In this way, the loop accurately controls the output to produce an —amplified or otherwise processed— replica of the input signal.

The application of feedback in general improves the properties of a circuit at the cost of a reduced gain. Notable improvements are enhancement of the accuracy of the signal processing and an increased ruggedness against parameter variations. Furthermore feedback allows for extremely low or high input and output impedances, depending at which type of electrical signal is aimed. A third beneficial effect of feedback is its ability to reduce the levels of non-linear distortion.

The prime arguments for using feedback from a historical perspective are the increased accuracy and reduced sensitivity to parameter variation [1]. The improvement of these two properties very much depends, however, on which parameter is considered and where in the feedback loop it has its effect. Figure 2-1 shows a basic feedback system, which consists of the amplifier A and the feedback network β. The closed-loop gain G follows from the well-known expression

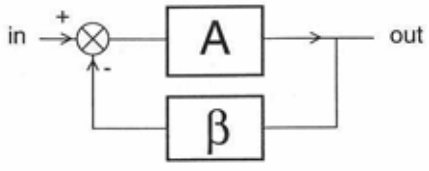


Fig. 2-1. *Basic feedback configuration with amplifier* A *and feedback network* β.

$$G = \frac{A}{1 + A\beta} \tag{2-1}$$

The quantity $A\beta$ represents the open-loop gain of the amplifier. It equals the transmission around the feedback loop. If $A\beta$ is much greater than 1, the feedback network β solely determines the overall transfer of Eq. 2-1.

$$G \approx \frac{1}{\beta}, \quad A\beta \gg 1 \tag{2-2}$$

The feedback reduces the gain, but in exchange the effects of parameter variations of the amplifier are eliminated. In general, the sensitivity factor S_A^G expresses the reduction of the variations. The sensitivity factor correlates the relative change of the closed-loop gain $G(A)$ to the relative change of the gain of the amplifier part A. In mathematical terms [2]

$$\frac{dG(A)}{G(A)} = S_A^G \cdot \frac{dA}{A} \tag{2-3}$$

The sensitivity of the basic feedback example follows from Eq. 2-1 and Eq. 2-3 and equals

$$S_A^G = \frac{dG(A)}{dA}\frac{A}{G(A)} = \frac{1}{1 + A\beta} \tag{2-4}$$

Since in the open-loop case the sensitivity S_A^G is equal to one, according to Eq. 2-4 the feedback reduces the sensitivity of the amplifier to variations of the forward gain by the factor $1 + A\beta$. For high loop gains $A\beta$ the reduction approximately equals the loop gain itself.

The amplifier's sensitivity to the feedback path β is less favorable. We can easily understand this from the fact that the feedback path grossly determines the gain of the amplifier. Since the closed-loop gain is inversely proportional to the gain of the feedback network, the sensitivity S_β^G approaches minus one:

$$S_\beta^G = -\frac{A\beta}{1 + A\beta} \approx -1 \tag{2-5}$$

The difference between the sensitivity of the closed-loop gain to the elements in the respective forward and backward paths indicates that we can only expect improvement of the performance when the feedback network β has better properties than the amplifier part A. Fortunately this will be the case for most circuits, since in practice the forward path relies on active devices, while for the majority of the applications a passive feedback network suffices. The passive components usually have significantly better properties regarding linearity and long-term stability.

The distinction between the effect of feedback on the forward and backward elements can be a severe complication in practical circuits. The difference renders it imperative to exactly identify the function of each of the components. In complex circuits, the classification can be vague, however.

A second aspect of feedback is the improvement of the input and output impedances of a circuit. In electrical networks four different feedback loops are possible, based on the nature of the signals at the input and the output. Figure 2-2 shows the basic circuits. We can easily identify the two distinctive parts of a feedback system: a forward gain block A and a feedback block β. The blocks represent two-ports, which permits two types of feedback connection at each side, accounting for the total of four different topologies.

The first type of feedback is the series connection, which appears, for instance, at the input of Fig. 2-2*a* and at the output of Fig. 2-2*b*. The feedback network is located in series with the amplifier terminals. At the output the feedback measures the current through the terminals, while the input voltage is compared to the voltage coming from the feedback net-

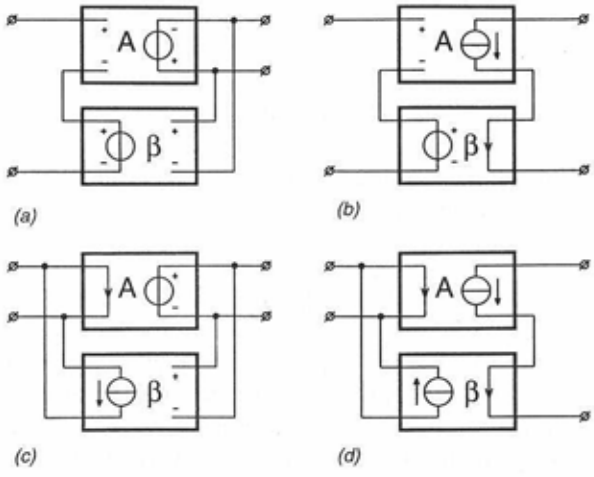


Fig. 2-2. *Basic feedback circuits with the feedback network connected as:* (a) *series-parallel,* (b) *series-series,* (c) *parallel-parallel,* (d) *parallel-series.*

work. Therefore series feedback at the input yields a voltage input, while when applied to the output terminals a current output results. The second type is parallel feedback. Examples are the input and the output of Fig. 2-2*d*. Parallel feedback has the opposite effect of series feedback, in that the current becomes the prime quantity at the input and the voltage at the output.

Based on the distinction between the two types of feedback, we can characterize the circuit of Fig. 2-2*a* as a voltage-to-voltage amplifier (series feedback at the input, parallel feedback at the output). Figure 2-2*b* is a voltage-to-current amplifier (series-series), 2-2*c* a current-to-voltage amplifier (parallel-parallel) and finally Fig. 2-2*d* represents a current-to-current amplifier with parallel connected feedback at the input and series connected feedback at the output. The controlled sources in their interior indicate the nature of each of the circuit blocks.

The impedance levels reflect the fact that the different feedback types turn either the voltage or the current into the primary signal. We can

expect that when the voltage becomes the controlled quantity at the output (parallel feedback), the impedance goes down, rendering the output voltage more independent from the load. Conversely, current being the fixed quantity, the output impedance is likely to go up. At the input things are the other way around: sensing voltages leads to a high impedance, whereas current sensing yields a low input impedance. It is commonly known that the feedback increases or reduces, depending on the type, the impedance levels of the basic circuits of Fig. 2-2 by the factor $1 + A\beta$ [1]. We can therefore summarize the properties of feedback in electrical networks as follows: series feedback increases the impedance by a factor $1 + A\beta$, while parallel feedback decreases it by the same factor. This conclusion holds for both the input and the output terminals.

Apart from the reduction of the sensitivity to parameter variation and improvement of the in and output impedances, feedback also helps to reduce signal distortion caused by the non-linearity of the active components. We can relate this subject to the stabilizing properties of feedback through the introduction of signal sources that model the signal components generated by the components' non-linearities. As long as the non-linearity is not too large, this simple approach leads to sufficiently accurate results.

The organization of this chapter is as follows. Section 2.1 presents a formal approach to feedback based on the fundamental quantity F, which denotes the so-called *return difference*. The section pays attention to the improvement of the accuracy and the sensitivity. Furthermore, it addresses the effect of feedback on the impedance levels at the input and the output of an amplifier circuit. Section 2.2 of this chapter is dedicated to non-linear distortion and noise. The section re-uses the sensitivity analysis of the first section for this new purpose. This approach again leads to a return difference based theory. Section 2.3 summarizes the conclusions.

2.1. Feedback in linear networks

The mathematical analysis of electronic feedback goes back to the beginning of this century and was founded by, among others, Nyquist and Bode [4], [1]. Their investigations are based upon the theory of linear networks. Although the linearity of many components —most notably transistors— leaves much to be desired, we can apply the linear network theory to phys-

ical circuits without too great an error. Under this assumption, a set of simultaneous linear equations fully describes the behavior of the circuit. Solving the set leads to the comprehensive solution of the voltages and currents in the circuit.

2.1.1. The solution of a linear network

Writing down the nodal equations of the network is an effective method for finding the relations that describe a circuit[1]. When a circuit comprises N nodes, it returns a set of N - 1 equations. Writing these down in the matrix form yields

$$\begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1\,(N-1)} \\ m_{21} & m_{22} & \cdots & m_{2\,(N-1)} \\ \cdots & \cdots & \cdots & \cdots \\ m_{(N-1)\,1} & m_{(N-1)\,2} & \cdots & m_{(N-1)\,(N-1)} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \cdots \\ v_{N-1} \end{bmatrix} = \begin{bmatrix} i_1 \\ i_2 \\ \cdots \\ i_{N-1} \end{bmatrix} \tag{2-6}$$

which we can also denote as

$$M \cdot \bar{v} = \bar{i} \tag{2-7}$$

where the vector $\bar{v}$ represents the nodal voltages and $\bar{i}$ the currents flowing into the nodes. The matrix equation expresses Kirchhoff's current law, stating that the sum of all the currents flowing into a node must equal the total current flowing out. As an example how the various impedances and controlled sources enter the matrix, consider the following example of the bipolar shunt stage of Fig. 2-3. The figure gives both the actual and the linearized small-signal circuit diagram. For simplicity, an ideal voltage controlled current source (transconductor) with input admittance y_p models

1. The mesh equations are also an acceptable choice. But since selecting a consistent set of meshes is a rather ambiguous task, the nodal equations are in general preferable.

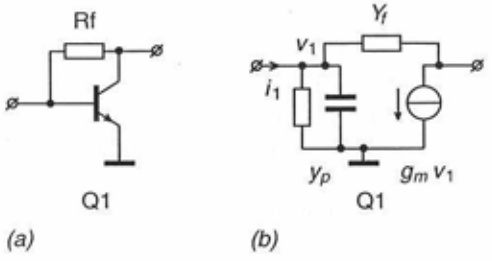


Fig. 2-3. (a) *Bipolar shunt stage and* (b) *the small-signal circuit diagram.*

the transistor. Notice that in the small-signal diagram is equally valid for a CMOS shunt stage if y_p consists of a capacitive part only. The matrix equation for the two nodal voltages is

$$\begin{bmatrix} y_p + Y_f & -Y_f \\ -Y_f & Y_f \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} i_{in} \\ -g_m v_1 \end{bmatrix} \tag{2-8}$$

or, moving the contribution of the controlled source into the matrix

$$\begin{bmatrix} y_p + Y_f & -Y_f \\ g_m - Y_f & Y_f \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} i_{in} \\ 0 \end{bmatrix} \tag{2-9}$$

We can find the output voltage as a function of the input current by applying Cramer's rule to Eq. 2-9 [5]. This results in

$$\frac{v_2}{i_1} = \frac{\Delta_{12}}{\Delta} = \frac{\det \begin{bmatrix} g_m - Y_f \end{bmatrix}}{\det \begin{bmatrix} y_p + Y_f & -Y_f \\ g_m - Y_f & Y_f \end{bmatrix}} = \frac{-g_m + Yf}{Y_f (g_m + y_p)} \tag{2-10}$$

In general, the set of nodal equations given by Eq. 2-6 has the following general solution [1]

$$\frac{v_j}{i_i} = \frac{\Delta_{ij}}{\Delta} \tag{2-11}$$

In this equation, v_j is the voltage at node j, i_i is the input current at node i and Δ is the determinant of the matrix. Symbol Δ_{ij} represents the cofactor of the entry at row i and column j. The cofactor is defined as the value of the matrix' determinant when row i and column j are omitted, multiplied by the appropriate plus or minus sign S. This sign depends on which are the removed row and column and is given by

$$S_{ij} = (-1)^{i+j} \tag{2-12}$$

Equation 2-11 expresses the transfer from an input current at node i to the voltage at node j. When the indices i and j differ from each other, the equation denotes a *mutual impedance*, while equal indices indicate a *self-impedance*. In the case of a self-impedance, the voltage of a particular node is related to the current flowing into the same node, which is just another way of expressing Ohm's law generalized for complex quantities. The determinant Δ in the denominator of Eq. 2-11 confirms the well-known fact that all transfers and self-impedances in an electrical network are subject to the same singularities. These singularities (the poles of the network) arise from the roots of the circuit's determinant.

2.1.2. Return difference and sensitivity

The general measure for the feedback in a network is the return ratio T. It is defined as the product of the gain of a particular element and the gain of the network around it. From the return ratio we can calculate the return difference F according to $F = 1 + T$. The return ratio and the return difference are similar to the quantities $A\beta$ and $1 + A\beta$ in a basic feedback system. An important difference is, however, that we can deduce the new parameters directly from the set of linear equations that describe the network. This aspect prevents the ambiguity of identifying each of the elements in the loop as part of the forward or the backward path. Furthermore, the return ratio T and the return difference F maintain their significance in systems with more than one feedback loop. In such multi-

loop systems the loop gain parameter $A\beta$ looses its meaning, since it is impossible to identify a single forward path A and a single feedback path β.

The return ratio of a particular element of a network obeys [1]

$$T = w\frac{\Delta_{kl}}{\Delta^0} \tag{2-13}$$

In this expression w is a constituent of the, mutual or self-, impedance under consideration, k is the input node of the impedance and l the output node. The denominator Δ^0 represents the determinant of the network when w is zero. Equation 2-13 directly follows from the definition of the return ratio, since w represents the gain of the element under consideration, while, according to Eq. 2-11, Δ_{kl}/Δ^0 is the gain of the network around it.

From Eq. 2-13, applying standard linear algebra, we can find the expression for the return difference [1]. The resulting equation is

$$F = 1 + T = \frac{\Delta}{\Delta^0} \tag{2-14}$$

where Δ denotes the system's determinant. Since the determinant corresponds to the characteristic polynomial, Eq. 2-14 states

> *The return difference for any element in a system is equal to the ratio of the respective values assumed by the characteristic polynomial when the specified element is included and when it has been removed from the system.*

We can use Eq. 2-14 to determine the return difference of the transistor Q_1 in the shunt stage circuit of Fig. 2-3. Substituting the proper values yields

$$F = 1 + T = \frac{\det\begin{bmatrix} y_p + Y_f & -Y_f \\ g_m - Y_f & Y_f \end{bmatrix}}{\det\begin{bmatrix} y_p + Y_f & -Y_f \\ -Y_f & Y_f \end{bmatrix}} = 1 + \frac{g_m}{y_p} = 1 + g_m z_p \tag{2-15}$$

The impedance z_p here denotes the reciprocal of the admittance y_p. Inspection confirms the result of Eq. 2-15, since the term $g_m z_p$ exactly gives the loop gain in the circuit.

The return difference on itself does not express the sensitivity improvement due to the feedback. One reason is that the relationship between sensitivity and feedback not only depends on the amount of feedback, but also on the location of the element in the loop. In the forward path the feedback reduces the effect of parameter variations to a great extend, but improvement is not likely for elements in the feedback path. A second reason is that the return difference does not involve a specific input or output of the system. Clearly, the sensitivity of a transfer with respect to an element must depend on, apart from the element considered, the transfer under investigation.

The expression for the sensitivity of the closed-loop gain G with respect to the parameter w, S_w^G is [1] [2]

$$S_w^G = \frac{1}{F}\left(1 - \frac{G_0}{G}\right) \tag{2-16}$$

In this equation G_0 is the direct transmission term. It represents the closed-loop gain when the element w has been omitted from the network. Equation 2-16 leads to a special case when the direct transmission term is zero. The sensitivity then approaches its maximum and equals the reciprocal of the return difference. In most amplifiers, the active components have no appreciable direct transmission, since removing them from the circuit breaks down the gain of the amplifier. Feedback reduces the sensitivity of these active parts by the return difference.

The transistor in the circuit of Fig. 2-3 is a typical example of an active part with zero direct transmission. The sensitivity to transconductance variations $S_{g_m}^G$ simply follows from the return difference of Eq. 2-15. To find the sensitivity of the transfer function with respect to the shunt resistor $S_{Y_f}^G$, we need the full expression of Eq. 2-16. The result is

$$S_{Y_f}^{G} = \frac{\det\begin{bmatrix} y_p & i_{in} \\ g_m & 0 \end{bmatrix}}{\det\begin{bmatrix} y_p + Y_f & i_{in} \\ g_m - Y_f & 0 \end{bmatrix}} = \frac{g_m}{g_m - Y_f} \approx 1 \tag{2-17}$$

This is in accordance to the expectations, since the shunt resistor y_f constitutes the feedback path of the amplifier.

2.1.3. The impedance of feedback circuits

The return difference also plays a major role in expressing the effect of feedback on the impedance levels in a circuit. Since the return difference is independent from the classification of the components as being part of the feedback or the forward path, such an approach leads to a generalized theory. It expresses the impedance between two arbitrary nodes as a function of a certain parameter, for instance the input impedance as a function of the transconductance of the output transistor.

The relation between the impedances with feedback z and without feedback z_0 is [1]

$$z = z_0 \frac{F(0)}{F(\infty)} \tag{2-18}$$

where $F(0)$ and $F(\infty)$ are the return differences as given by Eq. 2-14 when the terminals between which z is measured are respectively short-circuited and open-circuited.

Applying Eq. 2-18 to the input impedance of the shunt stage of Fig. 2-3 as a result of the feedback through the transconductance g_m requires the values of the input impedance z_0 of the circuit when $g_m = 0$, the return difference $F(\infty)$ when the input terminals are open and the return difference $F(0)$ when the input is short-circuited. The input impedance z_0 follows from inspection and equals $z_p = 1/y_p$. The return difference $F(\infty)$ is given by Eq. 2-15, while $F(0)$ equals one, as follows from setting $z_p = 0$ in Eq. 2-15. The input impedance therefore becomes

$$z_{in} = \frac{z_p}{1 + g_m z_p} \tag{2-19}$$

2.2. Non-linear distortion and noise

When the elements of an amplifier circuit are not sufficiently linear, we must take into account the non-linear distortion of the output signal. Although strictly speaking the linear network theory does not apply to these types of circuits, we can obtain workable results when the non-linearity is small. Fortunately, this appears to be the case for the majority of the amplifier circuits. For these circuits it is possible to model the distortion components by additive signal sources in the network. The sources generate the unwanted signals that are the direct result of the non-linear transfer curve of the components. The amplitude and the properties of the distortion signals very much depend on the input signal of the circuit.

Since a limitless number of input signals is possible, in the past several standard test signals have been defined. By examining the ratio between the desired and undesired output signals, we can obtain distortion figures that are a measure for the non-linearity of the circuit. The two most important figures are the harmonic and intermodulation distortion. The first bases itself on an input signal consisting of a single sinusoid. The latter assumes two input frequencies of different amplitude.

2.2.1. Harmonic distortion

Applying a single sinusoid to a non-linear circuit usually results in a spectrum that consist of a peak at the base frequency of the input signal plus several higher order components. The base frequency corresponds to the ideal output signal, while the higher order harmonics account for the non-linearity. The ratio of the total energy in the harmonics and the energy present in the base signal is the total harmonic distortion. If we designate the amplitude of each of the components of the output signal as A1, A2, A3, ..., where A1 is the amplitude of the base frequency, then the total harmonic distortion follows from

$$D_T = \frac{\sqrt{A_2^2 + A_3^2 + \ldots}}{A_1} \tag{2-20}$$

2.2.2. Harmonic distortion and non-linearity

To establish a relation between the non-linearity of a circuit —or a component in the system— and the resulting harmonic distortion, the starting point is the Taylor series expansion of the transfer function abound the quiescent point [6] [7]. Suppose that the transfer function is the function

$$E_o = f(E_i) \tag{2-21}$$

then we can write the function f as a Taylor expansion about the quiescent point $E_{oQ} = f(E_{iQ})$ as

$$E_o = E_{oQ} + (E_i - E_{iQ}) \cdot \left.\frac{dE_o}{dE_i}\right|_{E_i = E_{iQ}} + \frac{(E_i - E_{iQ})^2}{2!} \cdot \left.\frac{d^2E_o}{dE_i^2}\right|_{E_i = E_{iQ}} + \ldots \tag{2-22}$$

which we may rewrite as

$$E_o = E_{oQ} + a_1 (E_i - E_{iQ}) + a_2 (E_i - E_{iQ})^2 + \ldots \tag{2-23}$$

For a simple transfer, such as the exponential transconductance of a single bipolar transistor or the quadratic relation of its MOS counterpart, the constants a_1, a_2, a_3, ... follow from the proper differentiations of the analytical function. In the case that no analytical description is available, a simple curve fitting procedure readily establishes the constants. Anyhow, the first three or four terms in the series usually describe the transfer function adequately in the normal mode of operation.

Suppose that we apply a sinusoidal input signal around the quiescent point E_{iQ} to the system

$$E_i - E_{iQ} = \hat{E}_i \cos \omega t \tag{2-24}$$

Substituting the input signal of Eq. 2-24 into the Taylor series expansion of Eq. 2-23 and applying the well-known goniometric relation

$$2\cos a\cos b = \cos(a+b) + \cos(a-b) \tag{2-25}$$

results in the harmonic series for the output signal

$$\begin{aligned} E_o - E_{oQ} = &\left(\frac{1}{2}a_2\hat{E}_i^2 + \frac{3}{8}a_4\hat{E}_i^4 + \ldots\right) \\ &+\left(a_1\hat{E}_i + \frac{3}{4}a_3\hat{E}_i^3 + \ldots\right)\cos\omega t \\ &+\left(\frac{1}{2}a_2\hat{E}_i^2 + \frac{1}{2}a_4\hat{E}_i^4 + \ldots\right)\cos 2\omega t \\ &+\left(\frac{1}{4}a_3\hat{E}_i^3 + \frac{5}{16}a_5\hat{E}_i^5 + \ldots\right)\cos 3\omega t + \ldots \end{aligned} \tag{2-26}$$

which we can rewrite as

$$E_o - E_{oQ} = A_0 + A_1\cos\omega t + A_2\cos 2\omega t + \ldots \tag{2-27}$$

where the constants $A_0, A_1, A_2, \ldots$ represent the amplitudes of the harmonic components of the output signal[1]. From Eq. 2-27 we can determine the amplitude of the harmonic components as a function of the input signal's amplitude and the character of the non-linearity. Term A_0 indicates a D.C. component introduced to the output signal by the non-linearity. The term solely depends on even order harmonics, which non-symmetrical transfer functions introduce. It is caused by the fact that when the transfer function is not symmetrical around the quiescent point, the average value of the output signal will shift away from the quiescent output value E_o.

Substituting the constants $A_0, A_1, \ldots$ of Eq. 2-27 into the definition of harmonic distortion (Eq. 2-20) results in

1. Both [6] and [7] give —different— wrong values for the constants.

$$D_2 = \frac{A_2}{A_1} = \frac{\frac{1}{2}a_2\hat{E}_i^2 + \frac{1}{2}a_4\hat{E}_i^4 + \ldots}{a_1\hat{E}_i + \frac{3}{8}a_3\hat{E}_i^3 + \ldots} \approx \hat{E}_i\left(\frac{a_2}{2a_1}\right)$$

$$D_3 = \frac{A_3}{A_1} = \frac{\frac{1}{4}a_3\hat{E}_i^3 + \frac{5}{16}a_5\hat{E}_i^5 + \ldots}{a_1\hat{E}_i + \frac{3}{8}a_3\hat{E}_i^3 + \ldots} \approx \hat{E}_i^2\left(\frac{a_3}{4a_1}\right) \qquad (2\text{-}28)$$

$$D_4 = \frac{A_4}{A_1} = \frac{\frac{1}{8}a_4\hat{E}_i^4 + \ldots}{a_1\hat{E}_i + \frac{3}{8}a_3\hat{E}_i^3 + \ldots} \approx \hat{E}_i^3\left(\frac{a_4}{8a_1}\right)$$

Where D_2, D_3, ... represent the distortion figures for the various harmonics in the output signal. We can find the total harmonic distortion through

$$D_T = \sqrt{D_2^2 + D_3^2 + \ldots} \qquad (2\text{-}29)$$

The approximations in the right-hand terms of Eq. 2-28 bases itself on the assumption that the contributions of the higher order terms are negligible. In most practical cases this assumption will be justifiable. The magnitude of the harmonic components is proportional to the signal amplitude to a power one less than the number of the harmonic. In mathematical form we can write this general relation as

$$D_r \propto \hat{E}_i^{r-1} \qquad (2\text{-}30)$$

2.2.3. Intermodulation distortion

The measurement of intermodulation distortion involves inputting a signal that consists of two sinusoids of different frequency and amplitude. Because of mixing effects, the system will generate a multitude of distortion components at the output.

The fact that more than one frequency is present in the input signal complicates the specification of intermodulation distortion. More than one

definition is in common use. This section is limited to the type that defines a low frequency input signal at ω_a of high amplitude and a small signal with the relatively high frequency ω_b. The usual ratio of their magnitudes is 4:1. The choice of the input signals has the advantage that the harmonics of the high frequency signal ω_b fall outside the useful band of the system. Therefore the output power spectrum will have peaks at the following frequencies

$$\begin{aligned} &\omega_a, 2\omega_a, 3\omega_a, \ldots \\ &\omega_b, (\omega_b \pm \omega_a), (\omega_b \pm 2\omega_a), \ldots \end{aligned} \tag{2-31}$$

The second line of this equation denotes the intermodulation products that we are interested in. The first part symbolizes the harmonics of ω_a, which are of no immediate interest. The low value of the frequency ω_a, however, assures that they lie well below the intermodulation products with which we are concerned. Therefore, one can easily remove the harmonic signals by a high pass filter.

The relation between the non-linearity, the signal amplitudes and the intermodulation distortion follows from a similar analysis as for the harmonic distortion. The input signal is

$$E_i = \hat{E}_a \cos\omega_a + \hat{E}_b \cos\omega_b \tag{2-32}$$

where $\hat{E}_a$ is the amplitude of the low frequency signal and $\hat{E}_b$ the amplitude of the high frequency signal. In the special case considered here the amplitude ratio $\hat{E}_a/\hat{E}_b = 4{:}1$, but as we will see, this choice does not affect the outcome. Substituting the input signal into the Taylor series of Eq. 2-23 and performing a binomial expansion of the powers, we are able to calculate the intermodulation products. Since we are interested in the components around the high frequency signal $\hat{E}_b \cos\omega_b$, only terms in the result of the expansion need consideration of the form

$$\hat{E}_b \cos\omega_b (\hat{E}_a \cos\omega_a)^n \tag{2-33}$$

where n is an arbitrary power. The resulting intermodulation components lie at the frequencies $\omega_b \pm n\omega_a$. The general expression for the magnitudes of the intermodulation products that follows is

$$A_n = a_{n+1}\hat{E}_a^n\hat{E}_b \cdot \frac{n+1}{2^n} \tag{2-34}$$

where $a_n + 1$ is the $n + 1th$ coefficient of the Taylor expansion. Hence, the portion of the output signal we are interested is

$$\begin{aligned} E_o &= a_1\hat{E}_b\cos\omega_b + a_2\hat{E}_a\hat{E}_b\cos(\omega_b \pm \omega_a) \\ &\quad + \frac{3a_3\hat{E}_a^2\hat{E}_b}{4}\cos(\omega_b \pm 2\omega_a) + \ldots \end{aligned} \tag{2-35}$$

We can now define the intermodulation distortion of Eq. 2-35 in a fashion similar to the harmonic distortion given by Eq. 2-28

$$\begin{aligned} D_{I2} &= 2\frac{a_2\hat{E}_a\hat{E}_b}{a_1\hat{E}_b} = \frac{2a_2\hat{E}_a}{a_1} \\ D_{I3} &= 2\frac{3a_3\hat{E}_a^2\hat{E}_b}{4a_1\hat{E}_b} = \frac{3a_3\hat{E}_a^2}{2a_1} \\ &\ldots \end{aligned} \tag{2-36}$$

The factor 2 in each of the equations accounts for the two identical intermodulation products present below and above frequency ω_b. It is noteworthy that the intermodulation distortion does not depend on the amplitude of the high frequency signal at ω_b. The general relation between the distortion and the magnitude of the low frequency signal at ω_a is according to Eq. 2-36

$$D_{Ir} \propto \hat{E}_a^{r-1} \tag{2-37}$$

which is similar to Eq. 2-30 in the previous section. In analogy to the total harmonic distortion of Eq. 2-29, the total of all the components D_{Ir} is the sought for intermodulation distortion D_I

$$D_I = \sqrt{D_{I2}^2 + D_{I3}^2 + \ldots} \tag{2-38}$$

2.2.4. An example of distortion in a non-linear network

As an example of the two types of distortion covered in the previous sections, consider the voltage-to-current transfer of a bipolar transistor. This relation is furnished by the well-known expression

$$I_c = I_s \exp\left(\frac{qV_{be}}{kT}\right) \tag{2-39}$$

We can expand this exponential relation into a Taylor series around the base-emitter voltage under the quiescent condition V_{beq}

$$I_c = I_q + I_q \frac{V_{be} - V_{beq}}{V_T} + \frac{1}{2} I_q \left(\frac{V_{be} - V_{beq}}{V_T}\right)^2 + \ldots \tag{2-40}$$

where V_T abbreviates the thermal voltage kT/q. From Eq. 2-40 we can identify the Taylor coefficients a_1, a_2, ... as

$$a_1 = \frac{I_q}{V_T},\ a_2 = \frac{I_q}{2V_T^2},\ a_3 = \frac{I_q}{6V_T^3}, \ldots \tag{2-41}$$

To calculate the harmonic distortion, suppose that the amplitude of the input signal $\hat{E}_i = 10\text{mV}$. Then, with Eq. 2-39, the components of the harmonic distortion are

$$D_2 = \frac{\hat{E}_i}{2 \cdot 2V_T} = 10\%,\ D_3 = \frac{\hat{E}_i^2}{4 \cdot 6V_T^2} = 0.62\%, \ldots \tag{2-42}$$

Performing the calculations for the intermodulation distortion as given by Eq. 2-36 yields

$$D_{I2} = \frac{2\hat{E}_a}{2V_T} = 38\%,\ D_{I3} = \frac{3\hat{E}_i^2}{2 \cdot 6V_T^2} = 3.7\%, \ldots \tag{2-43}$$

To illustrate the effect of the magnitude of the input signal on the distortion figures, Fig. 2-4 plots the second and third order component of the

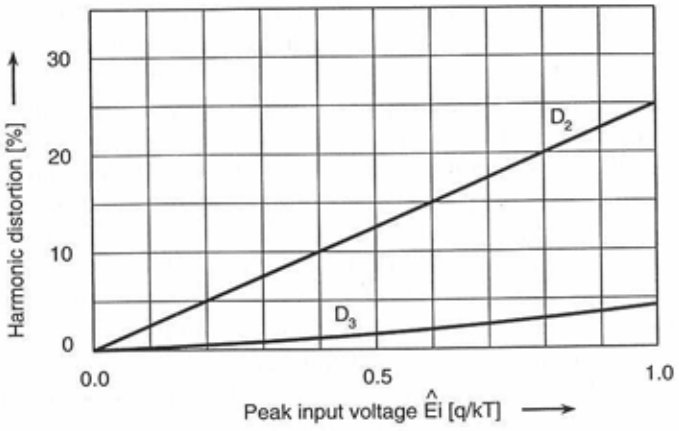


Fig. 2-4. *Second and third order harmonic distortion components as function of the input amplitude.*

harmonic distortion as a function of the amplitude of the input signal. The graph clearly shows that the second order component gives the greatest contribution to the total distortion. Considering practical input levels, the third order products lie over an order of magnitude below these second order contributions.

The importance of the second order components as part of the total harmonic distortion explains the popularity of symmetrical circuits for low distortion applications. The transfer function of these type of circuits is likely to be symmetrical about the quiescent point as well. This eliminates the even numbered terms of the Taylor expansion of Eq. 2-23, including the second. The second order term being removed, the third becomes dominant, lowering the total distortion figure significantly.

2.2.5. Modeling of non-linear distortion in linear networks

Only in a few very simple cases, such as the single bipolar transistor of the previous section, analytical distortion calculations of circuits lead to meaningful results. For more complex circuits we must revert to other, more simplified, means to incorporate the effect into the analysis. One approach is to model the distortion signals as signal-dependent additive sources in a linear network. That we can obtain accurate results in this way is not at all trivial. Non-linear distortion is by its nature a *multiplicative* effect. The distortion is introduced by a signal-dependent variation of the gain of a circuit. Modeling by *additive* sources is acceptable, however, as long as the "distortion-of-distortion" effects are small [1]. Equation 2-27 and Eq. 2-32 both indicate that the non-linear effects introduce distortion components on top of the expected linear output signal. By modelling these components as separate signal sources in the system, we can estimate their contribution to the output signal of the total system. We can consider the bipolar transistor discussed in the previous section, for example, as an ideal transconductor with an additional signal-dependent distortion source. Figure 2-5 illustrates this. The magnitudes of the signals

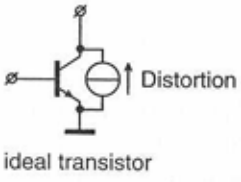


Fig. 2-5. *Bipolar transistor with current source modelling the distortion generated by the exponential characteristic.*

coming from the distortion source depend on the input signal and follows from Eq. 2-28 and Eq. 2-34 for harmonic distortion and intermodulation distortion, respectively.

By modelling the distortion as signal-dependent sources, we can investigate the effect of feedback on the non-linearity of a circuit. For this,

we must compare the distortion levels with and without the feedback. However, removing the feedback from the circuit will generally result in a much higher gain and therefore a higher output signal. In many practical situations the system will leave its assumed linear operation range and hit some physical limits, for example the supply voltage. We must therefore reduce the input amplitude to remain capable of accommodating the signal excursions in the system. The following assumes an input signal so much reduced that the output amplitudes with and without the feedback are equal. This condition assures that the system parameters will not change appreciably without the feedback. The expression for the distortion reduction that follows is [1]

$$\frac{D}{D_0} = F \tag{2-44}$$

where D is the distortion figure with and D_0 without feedback applied. Equation 2-44 simply states that the distortion reduction equals the return difference.

The ability to reduce the non-linear distortion is probably the prime argument for using feedback in electronic circuits. A single transistor, e.g. in the output stage, can easily produce tens of percents of harmonic and intermodulation distortion. As an illustration it is interesting to compare this number to the 16 bits resolution which is typical for most digital audio standards. Sixteen bits of resolution amounts to approximately 0.001% of summed noise and non-linear distortion. Therefore analog circuits that must perform up to this standard need an internal feedback of at least 10,000, or 80 dB.

2.2.6. Noise in feedback circuits

Equation 2-44, along with giving the relation between feedback and distortion, also suggests an approach for estimating the effect of feedback on the signal-to-noise ratio of a circuit. By replacing the distortion sources of the previous analysis by the internal noise sources, representing for example the thermal noise of the devices, we can find the total noise at the output. The return difference gives the ratio between the noise signals with and without the feedback. Despite the great similarity in their modelling, the effects of feedback on the noise and distortion performance differ widely. In case of distortion the undesired signal levels are reduced by the

return difference F. Although this reduction also applies to the noise levels, the *signal-to-noise ratio* does not improve. The feedback reduces the output signal by the same amount as the noise contributions. The effect on the signal-to-noise ratio is therefore given by

$$\frac{S/N}{S_0/N_0} = \frac{\left(\frac{S_0}{F}\right)/\left(\frac{N_0}{F}\right)}{S_0/N_0} = 1 \tag{2-45}$$

The distinction between the distortion and noise properties resides on the fact that the distortion levels depend on the signal magnitude, while the noise levels do not. This difference also accounts for the well-known fact that the signal-to-noise ratio is mainly determined by the properties of the input stage, while the distortion mostly originates in the output stage. The input stage normally carries the smallest signals, which are most easily corrupted by noise. The large signals in the output stage produce the greatest distortion contributions.

2.3. Conclusions

Feedback has a beneficial effect on such properties as accuracy, sensitivity to parameter variations, impedance levels and distortion. These improvements go at the cost of a reduced gain. One characteristic that does not improve is the signal-to-noise ratio, however. This chapter presented a return-difference based approach to the key aspects of feedback. The advantage of using the return difference over the $A\beta$ loop gain model to express the feedback lies in the elimination of the need to identify each of the components in the total feedback loop as a forward or a backward signal path. This is especially of importance in complex multi-loop systems, where it is impossible to pin-point a single forward and feedback path.

3 Stability of Feedback Circuits

Feeding back the output signal to the input does not in all cases establish the equilibrium state expressed by the basic input-output relation

$$G = \frac{A}{1 + A\beta} \tag{3-1}$$

We can understand this from considering the consequence of reversing the sign of the forward path. Setting its gain to $-A$ results in the following alleged transfer function

$$G = \frac{-A}{1 - A\beta} \tag{3-2}$$

Although Eq. 3-2 is a valid solution, there is no change for an actual circuit to arrive in this state. The circuit will amplify a small error between the actual and the desired signal, caused for instance by noise or drift, and produce a large output signal. Because of the positive feedback, this in its turn leads to a greater error, resulting in an even larger output signal. The system is said to be instable, since any infinitesimal deviation from the equilibrium condition will cause the output signal to explode. This suggests the following definition for the stability of a linear system: stability is the condition that a bounded input signal results in a bounded response.

It is well known that under certain circumstances amplifiers with negative feedback can also fall into instability. At a certain frequency the phase shift becomes 180°, reversing the feedback sign, and the circuit breaks into a spontaneous oscillation. The investigation into the stability of negative feedback amplifiers requires the modelling of the frequency dependency of the properties. Traditionally this is done by introducing the Laplace operator $s = j\omega$. This assumes a linear system, excited by complex exponential input signals. The transfer of the basic feedback system now becomes

$$G(s) = \frac{-A(s)}{1 + A(s)\beta(s)} \tag{3-3}$$

We can represent the transfer function of any linear system by the ratio of two polynomials in s [3]:

$$G(s) = \frac{p(s)}{q(s)} = \frac{K\prod_{i=1}^{M}(s+z_i)}{s^N\prod_{k=1}^{Q}(s+\sigma_k)\prod_{m=1}^{R}\left(s^2+2\alpha_m s+\left(\alpha_m^2+\omega_m^2\right)\right)} \tag{3-4}$$

The response of such a system to an impulse input signal $\delta(t)$ follows from the Laplace transformation of Eq. 3-4 and equals

$$c(t) = \sum_{k=1}^{Q} A_k e^{-\sigma_k t} + \sum_{m=1}^{R} B_m\left(\frac{1}{\omega_m}\right)e^{-\alpha_m t}\sin\omega_m t \tag{3-5}$$

For this result $N = 0$ is assumed. This condition implies the absence of poles at $s = 0$ which would lead to sustained step-like components in the step response. In the practice of electronic circuits this does not represent a limitation, since we can position the poles at $-\sigma_k$ arbitrarily close to zero.

If the quantities $-\sigma_k$ and $-\alpha_m$ in the arguments of the exponentials are greater than zero, the impulse response $c(t)$ will grow towards infinity at an exponential rate. Clearly, only negative values will result in a bounded output signal. Since the quantities are exactly the real parts of the

poles of Eq. 3-4, this observation leads to the fundamental condition for stability: *a feedback system is stable if and only if all the poles of the system transfer function have negative real parts.* In a graphical representation of the pole positions, this corresponds to the situation that all the poles reside in the left half of the complex s-plane.

The need to maintain a sufficient degree of stability poses fundamental limits on both the obtainable bandwidth and accuracy of feedback amplifiers. A good understanding of the mechanisms involved is essential to optimally exploit the potential of the available components. Especially for silicon technology —either bipolar or CMOS— distinctive limitations apply. This chapter addresses the fundamental limitations on the bandwidth and the gain, reverting to the Nyquist criterion for stability (Section 3.1). This fundamental theory also underlies the familiar concepts such as the gain and phase margin against regeneration and absolute and conditional stability. Several favorable open-loop frequency characteristics follow from the investigations. Section 3.2 addresses these along with the physical networks to implement the frequency responses. As it appears, most of desired frequency responses are not realizable in silicon I.C. processes because of the specific limitations of this fabrication method. Section 3.2 concludes by stating a 5-point list of demands for on-chip compensation strategies.

3.1. Stability analysis

Apart from the closed-loop criterion that no poles should reside in the right half of the complex s-plane, we would like to establish a requirement that decides on the stability of a circuit by examining the open-loop frequency response. The reason is that the positions of the poles before closing the loop are readily determined, since in general the system's transfer is known in a factored form. For the poles of the closed-loop system, on the other hand, we must solve the characteristic equation $1 + A\beta = 0$. If the order is too high, the solution of this polynomial equation is not at all trivial. Theoretically, fifth-order equations are the highest to yield an analytical solution. Mathematical practice restricts the order to as low as two. Higher order equations result in solutions that can fill dozens of pages, but

produce little insight. A criterion that is based on the open-loop properties of a feedback network effectively evades this complication, since the open-loop transfer function is typically available in a factored form.

3.1.1. Open-loop requirements for stability

Fundamental to the analysis of stability in terms of the open-loop characteristics is the Nyquist criterion. This criterion decides on the stability of a feedback system by examining the phase and the magnitude of the open-loop frequency response. It bases itself on Cauchy's theorem, which is concerned with the mapping of contours in the complex s-plane.

3.1.2. Cauchy's theorem

Cauchy's theorem describes the characteristics of the transformation of a closed trajectory in the s-plane by a complex function $H(s)$. This function $H(s)$ may be any complex function, but this chapter will investigate the characteristic equation of a feedback system

$$H(s) = 1 + P(s) = 0 \tag{3-6}$$

In a simple system $P(s)$ will be equal to the loop gain $A\beta(s)$, while in general it expresses the return ratio T. The function $H(s)$ represents the quantity $1 + A\beta(s)$, or the return difference F for the generalized case. Since the characteristic equation $H(s)$ has a finite numbers of poles and zeros, we may express $H(s)$ in the form

$$H(s) = 1 + P(s) = \frac{K\prod_{i=1}^{N}(s+s_i)}{\prod_{k=1}^{M}(s+s_k)} = 0 \tag{3-7}$$

Note that the zeros of $H(s)$ are the poles of the total feedback system, since the characteristic equation appears in the denominator of the transfer function.

Cauchy's theorem, also known as the principle of argument, relates the number of poles and zeros enclosed by a contour in the s-plane to the number of encirclements of the origin by the transformed contour in the $H(s)$ -plane:

If a contour Γ_s *in the* s*-plane encircles* Z *zeros and* P *poles of* H(s) *and does not pass through any poles or zeros of* H(s) *as the traversal is in the counterclockwise (positive) direction along the contour, the corresponding contour* Γ_H *in the* H(s) *-plane encircles the origin of the* H(s) *-plane* N = Z - P *times in the counterclockwise (positive) direction.*

As an example, consider Fig. 3-1. Part *a* shows the complex *s*-plane with

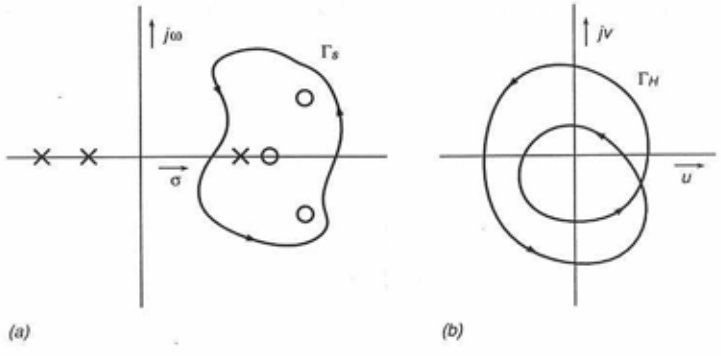


Fig. 3-1. *Example of Cauchy's theorem with* (a) *the s-plane and* (b) *the* H(s) *-plane.*

several poles and zeros. Contour Γ_s encloses three poles and one zero. Figure 3-1*b*, which depicts the $H(s)$ -plane, illustrates the result of traversing the contour Γ_s in the counterclockwise direction. The axes of the plot represent the variables u and jv, such that $H(s) = H(\sigma + j\omega) = u + jv$. The transformed contour is Γ_H. Since the number of zeros enclosed in the contour N is two greater than the number of poles P, setting the difference $N = Z - P = 2$, the number of counterclockwise encirclements of the origin in the $H(s)$ -plane by the transformed contour Γ_H is also equal to two. If the number of poles had been greater than the number of zeros, then the direction of the transformed contour Γ_H would have been in the opposite, clockwise, direction. In this case too, the number of encirclements of the origin would have been a measure for the difference between the numbers of poles and zeros.

3.1.3. The Nyquist criterion

Cauchy's theorem is useful for finding out whether poles are present in the right half of the complex s-plane by investigating the open-loop transfer function of the network. Since any poles appearing in the right half-plane indicate a instable network, this allows to predict the stability of a network by examining its open-loop frequency response. The theory was founded by H. Nyquist in 1932 and is therefore referred to as the *Nyquist criterion.*

In terms of the characteristic equation $H(s) = 1 + P(s) = 0$, no zeros of $H(s)$ should be present in the right half of the complex s-plane. Figure 3-2 shows the so-called *Nyquist contour* that encloses the entire

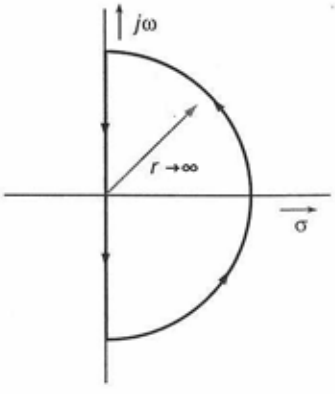


Fig. 3-2. *The Nyquist contour.*

right half-plane. The contour consists of a straight path from $-j\infty$ to $+j\infty$ and a semicircular path of radius r, where r approaches infinity, which closes it. Note that the straight path provides the familiar $H(s)$. To account for poles that lie exactly on the imaginary axis, the straight part of the contour may include infinitesimally small semicircular detours from the axis. According to Cauchy's principle of argument, the number of encirclements of the origin by the transformed contour Γ_H indicates the difference between the number of zeros and poles of $H(s)$ enclosed by the contour, i.e. in the right half-plane.

It is convenient to introduce a shift of variables into the characteristic equation $H(s)$. Let the function $H'(s)$ be defined as

$$H'(s) = H(s) - 1 \doteq P(s) \tag{3-8}$$

This change of variable translates the origin of the $H(s)$ -plane to the point (-1,0) in the $H'(s) \doteq P(s)$ plane. Therefore the number of encirclements of the origin in the $H(s)$ -plane by Γ_G corresponds to the number of encirclements of the point (-1,0) in the $P(s)$ -plane by Γ_P. The advantage of using the open-loop transfer function $P(s)$ is that it is typically available in a factored form.

The number of encirclements of the (-1,0) point in the $P(s)$ plane is equal to $N = Z - P$, the difference between the number of zeros and poles in the right half-plane of $H(s)$. But the poles of the characteristic equation $H(s) = 0$ are identical to the poles of the open-loop gain $P(s) = 0$:

$$H(s) = 1 + P(s) = 1 + \frac{N(s)}{D(s)} = \frac{D(s) + N(s)}{D(s)} \tag{3-9}$$

Hence, the number of poles of $H(s)$ in the right half of the s-plane, equals the number of poles of the open-loop gain $P(s)$ in the right half-plane. Further, the zeros of $H(s)$ correspond to the poles of the closed-loop system. So the number of zeros of $H(s)$ in the right half-plane, indicated by Z, equals the number of right half-plane poles of the total, closed-loop, system. We can therefore state the Nyquist stability criterion as follows:

> *A feedback system is stable if and only if, for the contour* Γ_P *the number of counterclockwise encirclements of the (-1,0) point is equal to the number of poles of* P(s) *with positive real parts.*

If there are no poles of $P(s)$ in the right half of the s-plane, which is a common situation, the Nyquist criterion reduces to

> *A feedback system is stable if and only if the contour* Γ_P *in the* P(s) *-plane does not encircle the (-1,0) point.*

Figure 3-3 gives an example. The figure represents a feedback system having loop gain

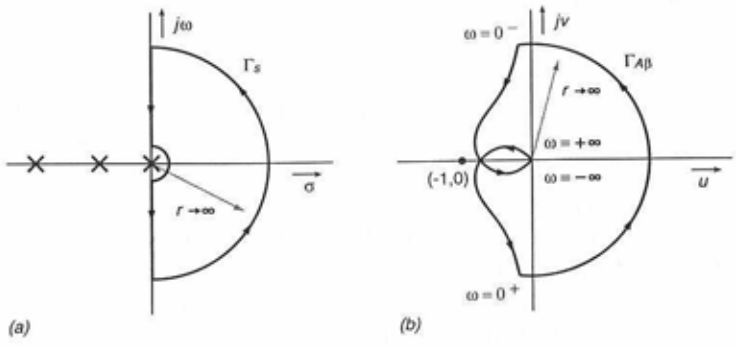


Fig. 3-3. *The Nyquist contour and its mapping into the* $A\beta\,(s)$ *- plane for a stable network.*

$$A\beta(s) = \frac{K}{s\,(\tau_1 s + 1)\,(\tau_2 s + 1)} \tag{3-10}$$

The open-loop gain $A\beta\,(s)$ contains three poles, which Fig. 3-3*a* shows in the complex *s*-plane. Also indicated in the figure is the Nyquist contour Γ_s with an infinitesimal semicircular detour around the pole in the origin. Mapping the contour Γ_s to the $A\beta\,(s)$ -plane yields the Nyquist diagram depicted in Fig. 3-3*b*.

Looking at the mapping in detail, the small semicircle in the *s*-plane transforms to the large semicircle in the $A\beta\,(s)$ -plane, while the large semicircle transforms to the origin. The portion of Γ_s that runs from 0^+ to $+\infty$ results in the part of $\Gamma_{A\beta}$ that corresponds to the well-known polar plot of the open-loop gain $A\beta\,(s)$. This is the case, because in this portion the relation $s = j\omega$ holds. The part of the Nyquist contour that covers the negative imaginary axis produces the complex conjugate of the polar plot.

Since there are no poles of the open-loop gain $A\beta\,(s)$ that lie in the right half of the *s*-plane, the fact that the point (-1,0) is not encircled by the contour $\Gamma_{A\beta}$ reveals that the closed-loop system is stable. By increasing the gain constant *K*, however, the network will become instable. The plot

in Fig. 3-4 demonstrates this. The increased gain scales up the contour $\Gamma_{A\beta}$

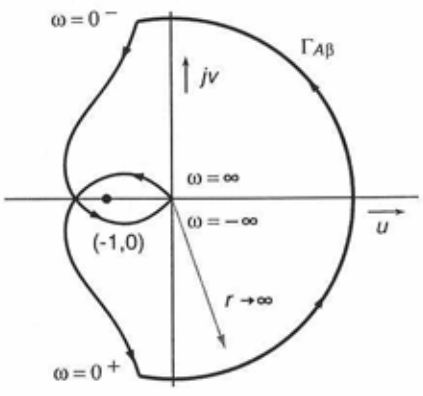


Fig. 3-4. *The Nyquist diagram of an instable network.*

which now encircles the (-1,0) point two times.

Adding the complex conjugate of a polar plot of the open-loop response, one can construct the Nyquist diagram of any system. The polar plot and its conjugate together form the Nyquist diagram. Only when poles are present in the origin of the s-plane, which causes asymptotes along the axes of the $A\beta\,(s)$ -plane, the Nyquist diagram requires additional circle segments of infinite radius to close the contour $\Gamma_{A\beta}$.

3.1.4. Conditional stability

Depending on the shape of the Nyquist diagram, we can make distinction between absolutely and conditionally stable networks. The difference bases itself on the effect of gradually reducing the loop gain. The absolutely stable network will remain stable, while the conditionally stable — also known as Nyquist stable— system will become instable in that situation. The Nyquist diagram of Fig. 3-3 is an example of absolute stability. Decreasing the loop gain, which corresponds to reducing the diagram, will never result in the critical point (-1,0) being encircled by the contour $\Gamma_{A\beta}$.

An example of a conditionally stable system is given in Fig. 3-5.

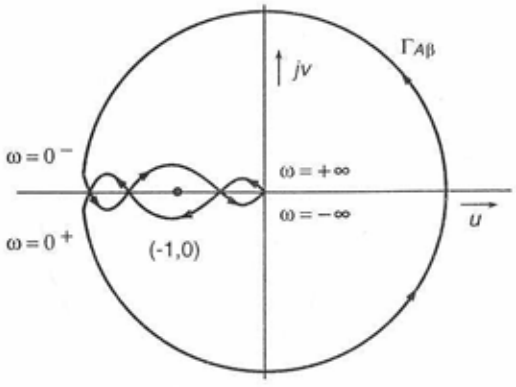


Fig. 3-5. *The Nyquist diagram of a conditionally stable network.*

The net number of encirclements of (-1,0) is zero (clockwise and counterclockwise encirclements counteract), hence the system in this form will be stable. Decreasing the gain moves the crossing in the contour to the left of (-1,0) in the positive direction along the real axis and the network will become instable.

When examining complex diagrams, such as the previous example, it can be difficult to count the number of encirclements. A convenient method in these cases is to consider the interior and exterior of the contour $\Gamma_{A\beta}$. The interior of a contour is defined as the area which lies to the left hand side of the locus when traversing in the counterclockwise direction. According to this rule a feedback system is stable if and only if the point (-1,0) lies in the exterior of contour $\Gamma_{A\beta}$ in the $A\beta\,(s)$ -plane.

3.1.5. Gain and phase margin

In many situations it is desirable to express the degree of stability by a quantitative measure. The Nyquist diagram lies at the base of two widespread figures: the gain and the phase margin [1]. Both figures judge the stability by defining the distance of the mapped contour $\Gamma_{A\beta}$ to the critical

point (-1,0). The gain margin x does so by indicating how much one can increase the gain constant K before the contour first encircles the (-1,0) point. It is defined as the reciprocal of the gain $|AB(j\omega)|$ at the frequency at which the phase angle reaches 180°.

The phase margin y is defined as the phase angle through which one must rotate the $|AB(j\omega)|$ locus in order that the unity-magnitude point $|AB(j\omega)| = 1$ passes through the (-1,0) point in the $AB(j\omega)$-plane. The measure is therefore equal to the amount of phase lag required for the system to become instable. We can evaluate the phase margin from the Nyquist diagram by drawing a unit circle centered around the origin. The difference between the phase angle at which the locus $|AB(j\omega)|$ crosses the unity circle and the real axis of the diagram gives the phase margin.

Both gain and phase margin are powerful in quickly determining the merits of a feedback system. The next section will establish quantitative data for the respective margins of electronic amplifiers, which will serve as a guideline for designing amplifier circuits throughout this work.

3.2. Maximum obtainable feedback

An important question before examining various compensation strategies is how much feedback we can obtain over which bandwidth with a given set of physical components. That there must be a limitation in this respect, is clear from the fact that any active component, be it a bipolar transistor, a CMOS transistor or even a vacuum tube, has a certain amount of phase lag at high frequencies associated with it. This phase shift cumulates when we connect a number of gain stages one after the other. Hence, cascading gain stages to improve the accuracy simultaneously reduces the phase margin of the circuit. This implies some kind of trade-off between accuracy and bandwidth. The investigation into this trade-off leads to the theory of the maximum feedback absolutely stable amplifier.

Apart from the absolutely stable approach, this section also probes into the properties of conditionally stable solutions. Although, for most applications absolute stability is a condition *sine qua non*, the improvement in the useful bandwidth that conditional stability can realize, sometimes makes it worthwhile to take these type of frequency responses into consideration. Typical applications are in the field of audio amplifiers and filters.

3.2.1. Maximum feedback absolutely stable amplifiers

The maximum feedback amplifier intends to realize the gain A_0 across as large a useful band as possible, while the phase margin should not become smaller than y. Figure 3-6*a* shows the (incomplete) Bode plot. In this plot,

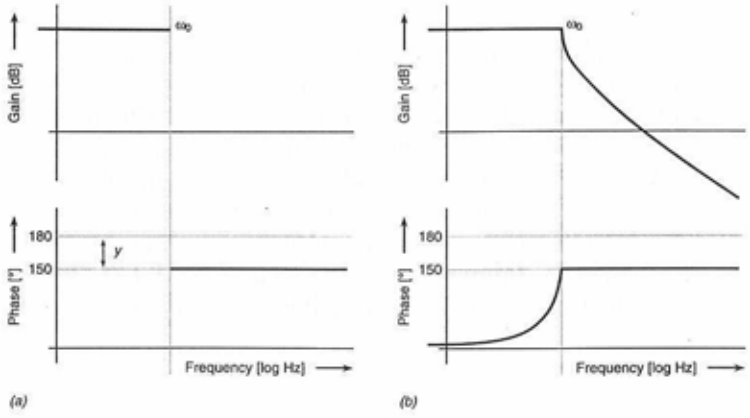


Fig. 3-6. *Bode plots showing* (a) *the starting conditions of the loop gain* Aβ *of a maximum feedback amplifier and* (b) *the total frequency response.*

the straight line in the gain plot from D.C. up to ω_0 illustrates the gain requirement, while the phase plot starting at ω_0 gives the demand on the phase margin. With the Bode relation that couples the imaginary part of a transfer function above a specific frequency with the real part below this frequency [1], we can uniquely construct the missing parts of the plot. The resulting transfer function has the form

$$A + jB = A_0 - 2\,(1 - y)\log\left(\sqrt{1 - \frac{\omega^2}{\omega_0^2}} + j\frac{\omega}{\omega_0}\right) \qquad (3\text{-}11)$$

where A and B are respectively the real and imaginary components of the loop gain $A\beta$ and y equals the phase margin in units of 180°. Figure 3-6*b* provides a plot of A and B for the choice $y = 1/6$, corresponding to a 30° phase margin.

In Eq. 3-11 we still have to determine the gain factor A_0. Further, the equation is very optimistic in that it assumes a phase characteristic that sustains a constant value below 180° up to infinite frequencies. In practice parasitic effects will limit the obtainable bandwidth. At high frequencies the parasitic capacitances will control the response, rendering it impossible to accurately approach the desired characteristic anymore. In general, at sufficiently high frequencies the frequency response of any amplifier approaches an asymptote. This asymptote is defined the *asymptotic gain limit* or the *gain asymptote*. Capacitors (intended or parasitic) determine the position of the asymptotic gain limit, which eventually fixes the maximum amount of gain obtainable from a circuit.

Figure 3-7*a* serves as an example of the asymptotic gain limit. The figure presents an amplifier, consisting of three cascaded gain stages and a feedback network N in parallel with the bypass capacitor C_b. Although the circuit comprises CMOS transistors, bipolar devices meet the needs equally well. Fig. 3-7*b* is the small-signal circuit diagram of the amplifier. The parallel combinations of the capacitors and the resistors $c_{p2}//r_{p2}$-$c_{p4}//r_{p4}$ model the frequency behavior of the forward circuit, while the combination $c_l//r_l$ symbolizes the load impedance. In the case of a CMOS amplifier, resistor r_{p4} solely stems from the output resistance of the signal source, while in the bipolar case it also incorporates the input resistance of transistor M_3. The function of the bypass capacitor C_b is to provide an efficient path for high frequencies from the output to the input. At sufficiently high frequencies the loop gain will depend only upon the various capacitors in the circuit, without regard to the resistors and the feedback network N. The asymptotic gain magnitudes of the respective transistors are $g_{m1}/(\omega c_{p2})$, $g_{m2}/(\omega c_{p3})$ and $g_{m3}/(\omega c_{p4})$. If we further assume that capacitor C_b and c_l are both significantly larger than c_{p4}, we can find the total asymptotic gain limit by multiplying the terms for each of the transistors. The result is

$$|A\beta| = \frac{g_{m1}g_{m2}g_{m3}}{\omega^3 c_{p3}c_{p2}c_l} \tag{3-12}$$

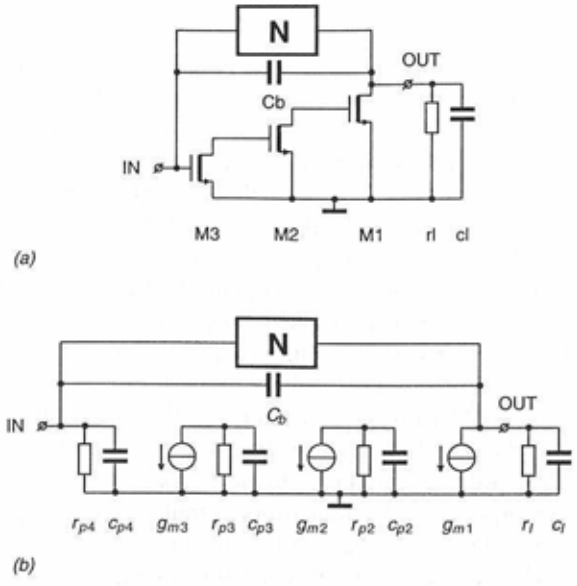


Fig. 3-7. *Typical circuit demonstrating the gain asymptote.*

The fact that we have assumed C_b to be larger than the input capacitor c_{p4} does not present a serious limitation to the general validity of the expression. When the circuit does not meet this condition, the combination C_b/c_{p4} will act as a capacitive voltage divider, merely introducing a frequency independent correction factor into Eq. 3-12. The reason that we assumed c_{p4} to be smaller than the load capacitor c_l, is that in this case c_{p4} does not contribute through the feedback network to the total capacitance seen at the output.

In a Bode plot the gain asymptote of Eq. 3-12 represents a straight amplitude characteristic with a slope of 60dB per decade. We can find the frequency ω_a where the gain asymptote crosses the 0dB axis by equating the right-hand term to one.

$$\omega_a^3 = \frac{g_{m1}g_{m2}g_{m3}}{c_{p3}c_{p2}c_l} \tag{3-13}$$

Equation 3-13 has a remarkable similarity with the expression for the transit frequency ω_t of a single transistor g_m/c_{gs}. In fact, when there is no load capacitor and the input capacitor c_{p4} is the only contribution to the output capacitance, the transit frequency ω_t of the total amplifier in terms of the transit frequencies of the individual transistors boils down to

$$\omega_t^3 = \frac{g_{m1}g_{m2}g_{m3}}{c_{p2}c_{p3}c_{p4}} = \omega_{t1}\omega_{t2}\omega_{t3} \tag{3-14}$$

where $\omega_{t,1}$ - $\omega_{t,3}$ are the transit frequencies of the respective transistors.

Equation 3-14 expresses the case of an ideal feedback amplifier, where the three stages only contribute three dominant poles. In practical situations it is possible that the feedback network itself introduces additional poles in the loop gain. This is likely, for example, when we omit the feedback capacitor C_b in Fig. 3-7 from the circuit. If the feedback network N does not have a direct path for high frequencies, additional poles will appear in the transfer around the loop. Also any frequency-independent attenuation at high frequencies of the loop gain, for example through capacitive voltage dividers such as C_b/c_{p4} in the figure, degrades the performance of the circuit. Such attenuation also occurs in non-unity gain configurations. In these circuits the feedback network will reduce the signal around the loop by the reciprocal of the desired gain. In all these cases, the resulting gain asymptote will lie below the optimum of Eq. 3-14.

Although the relations for the asymptotic gain limit were derived from the three stage example circuit, we can easily apply the results to an arbitrary amplifier with N stages. With N stages, the upper limit for the slope of the gain asymptote is $20N$dB. This value again relies on the assumption that the amplifier is ideal, meaning that the feedback network or any other part of the circuit does not introduce any additional poles and that there is no attenuation for high frequencies. The 0dB frequency ω_t, corresponding to Eq. 3-13 for the three stage situation, generally equals the product of the transconductances divided by the products of capacitances connected at the collectors/drains of the amplifier stages. In determining the capacitance at the output, we must make allowance for the

loading of the amplifier by the input impedance through the feedback network.

The asymptotic gain limit represents the upper bound for the bandwidth of any feedback amplifier. In practice no multi-stage circuit will ever attain this upper bound. We can easily see this from the phase shift associated with the slope of Eq. 3-12, giving the gain asymptote of the example. The phase shift amounts to 270°, hence closing of the feedback loop will lead to instability. The circuit requires a step back in unity-gain frequency to attain a stable behavior.

Figure 3-8 presents the bandwidth reduction caused by the intro-

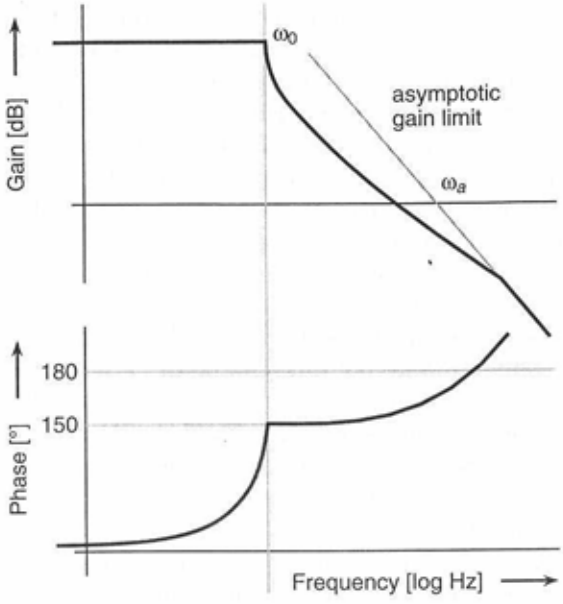


Fig. 3-8. *Bode plot showing the effect of the asymptotic gain limit on the phase margin of a maximum feedback amplifier.*

duction of the gain asymptote in the frequency response of the maximum feedback amplifier. The light line adds the asymptote to the plot of the

ideal characteristic. Since at high frequencies the gain asymptote represents the upper limit of the gain, the net frequency response will follow the solid curve for frequencies below the intersection point, and the light line above that frequency. The additional phase lag of the gain asymptote drastically reduces the phase margin. At high frequencies, the phase lag of the Nth order asymptote approaches $90N°$. The gain asymptote affects the phase margin over the entire frequency range above ω_0.

To obtain a (reasonably) constant phase margin from a physically realizable amplifier, we must connect the maximum feedback loop gain curve to the gain asymptote in a slightly different way. Figure 3-9 shows

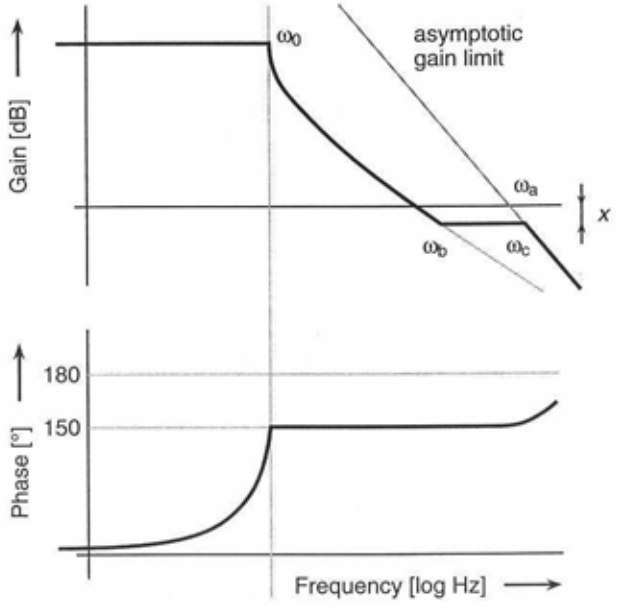


Fig. 3-9. *Realizable response of a maximum feedback amplifier.*

the new overall cut-off characteristic. It consists of the original theoretical characteristic from the edge of the useful band up to frequency ω_b. From ω_b the curve continues at a constant value, until it intersects with the gain

asymptote at frequency ω_c. The distance between the flat section of the plot lies and the 0dB-axis equals the predetermined gain margin of the amplifier.

We can regard the total frequency plot of Fig. 3-9 as the superposition of two separate semi-infinite line segments on top of the ideal frequency characteristic. This is illustrated by Fig. 3-10. The first segment

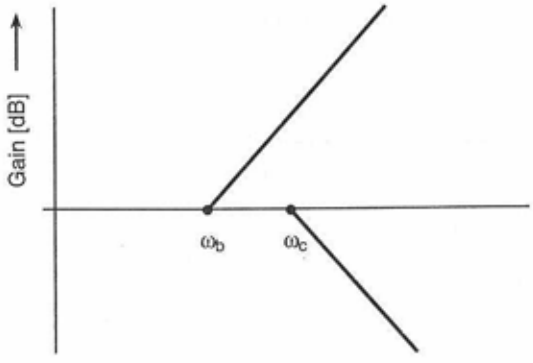


Fig. 3-10. *The two semi-infinite slopes that together reduce the phase shift in the passband of the maximum feedback amplifier.*

starts at frequency ω_b and has a positive slope that cancels exactly the slope at high frequencies of the ideal characteristic, i.e. 40 $(1 - y)$ dB/dec. The second segment starts at ω_c, and has a slope equal to that of the gain asymptote. The phase response at low frequencies of a semi-infinite slope is approximately proportional to the frequency and inversely proportional to the starting frequency of the slope. We can therefore choose the ratio of ω_b and ω_c in such a manner that the resulting phase shifts at low frequencies cancel out each other. We obtain the proper relations when the frequencies at which the slopes begin, are in the same ratio as the slopes themselves [1]. If we represent the slope of the asymptote, in units of 20 dB/dec, by n, this fixes ω_b in terms of ω_c by the relation

$$\omega_b = \frac{2(1-y)}{n}\omega_c \tag{3-15}$$

where y again denotes the phase margin in units of 180°. Further, it is possible to determine the frequency ω_c, since we know the frequency at which the gain asymptote crosses the 0dB level ω_a, as well as the slope at which the asymptote runs from ω_a to the frequency ω_c. Therefore, noticing that the gain difference at the frequencies ω_c and ω_a is exactly the gain margin, we can determine ω_c with respect to ω_a as

$$\omega_c = \omega_a 10^{\frac{x}{20n}} \tag{3-16}$$

where x is the gain margin in decibels. Finally, the maximum amount of gain in the useful band follows from [1]

$$A = A_m - (A_m + 17.4)\, y - \frac{n-2}{n}x - \frac{2}{n}xy \tag{3-17}$$

where

$$A_m = 40\log_{10}\frac{4\omega_a}{n\omega_0} \tag{3-18}$$

Frequency ω_0 in this equation is the frequency of the edge of the useful band. When inspecting Eq. 3-17 and Eq. 3-18, it becomes clear that A_m represents the ultimate limit on the gain, if the gain and the phase margin are zero. Requiring greater, more realistic margins reduces the obtainable gain. Since the cross product term xy is usually small, the total degradation in the feedback appears as the sum of the degradations due to the phase and the gain margin individually.

Depending on the requested phase margin, the slope of the high frequency portion of the maximum feedback Bode plot of Fig. 3-9 can lie anywhere between zero and 40dB/dec. Electronic circuits with lumped components have discrete poles and zeros, however, meaning that the asymptotic slope can only assume values that are a multiple of 20dB/dec. A constant slope of 30dB/dec, for example, is simply not realizable with a finite number of components. Therefore in practice one can at best approx-

imate the desired slope. The approach is to interlace poles and zeros on the negative real axis of the complex s-plane. With a logarithmic spacing, rates between 0 and 20 dB/dec are feasible. Figure 3-11 illustrates the

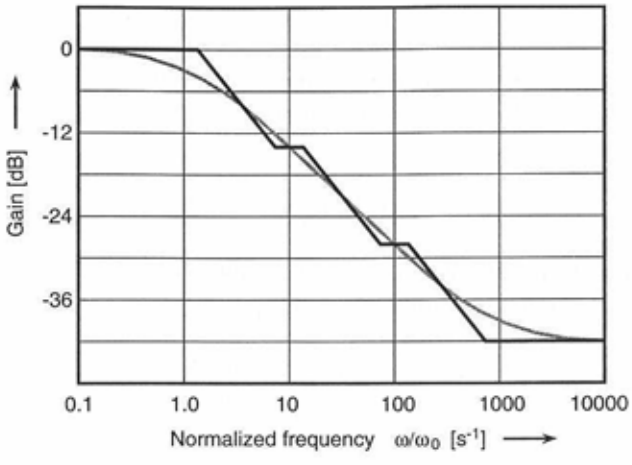


Fig. 3-11. *Bode plot showing the approximation of an arbitrary slope through pole-zero positioning. The dark line is the asymptotic approximation, the light line the actual response.*

pole-zero positioning. The poles frequencies are

$$\omega_0,\ p\omega_0,\ p^2\omega_0,\ p^3\omega_0,\ \ldots$$

and the zeros at

$$q\omega_0,\ pq\omega_0,\ p^2q\omega_0,\ p^3q\omega_0,\ \ldots$$

If both p and q approach one, so that the number of singularities becomes infinite, the asymptotic attenuation rate of the transfer function G becomes

$$\frac{d|G(j\omega)|}{d\omega} = -20\left(\frac{\log_{10}q}{\log_{10}p}\right)\text{dB/dec} \qquad (3\text{-}19)$$

The approximation holds quite well for p as large as 10, placing the poles one decade apart. We can obtain a gain slope of greater than 20dB/dec by placing additional poles at the beginning of the slope, $\omega = \omega_0\sqrt{q/p}$. Each pole contributes 20dB/dec to the total roll-off.

In the past many attempts have been made to realize amplifiers with maximum feedback properties. Especially in the early days of electronics, when the active parts lacked the degree of sophistication of present day monolithic components, designers spent much effort in approaching the theoretical limitations. It is surprising to see that, in fact, the results very much resembled the theoretical curve of Fig. 3-9. The general method to shape the frequency curve consisted of introducing shunt impedances between the amplifier stages. The impedances consisted of complex networks of capacitors and coils, obtained with help of the filter theory. Resistors were avoided, since the optimum of Eq. 3-17 prohibits the presence of resistive components of the shunt impedances outside the passband [1]. Figure 3-12 gives an example of the complexity of a typical interstage

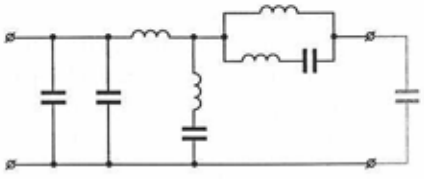


Fig. 3-12. *Typical interstage network for a maximum feedback amplifier.*

impedance [1].

3.2.2. Optimal frequency response for integrated amplifiers

Presently, the practical importance of the maximum feedback amplifiers is negligible. One of the reasons, already mentioned, is the improvement of the technology in which the circuits are realized. This development makes it less important to squeeze the maximum out of the components. The second is that the maximum feedback amplifier is very sensitive to variations of the components. One can easily understand this from the fact that the

amplifier is optimal with a certain set of parameters. Changes of one of the parameters, be it the transconductance of the active components, the load capacitance or any other determining factor, will subsequently result in a suboptimal configuration that very likely will not meet the specifications. This is an important point, since the matching of parameters that are not physically coupled as well as the absolute accuracy of the parameters are two weaknesses of silicon technology. Moreover, after-production trimming is only economically feasible in a very limited number of situations. The third and probably most important reason for the reduced interest in maximum feedback amplifiers is the necessity of large capacitors and inductors for the shaping of the frequency response. Components that silicon technology can implement only with great difficulty or even cannot implement at all.

The specific merits of monolithic integration demand frequency responses and compensation techniques that better exploit the advantages, while avoiding the drawbacks of the technology. The strong points include the availability of large numbers of active devices and accurate matching of coupled parameters. Also the frequency compensation must reflect that integrated amplifiers are universal building blocks for all kinds of feedback applications. The designer of an opamp does not know beforehand in what circuit the amplifier will end up. The lack of principal information, such as desired loop gain and the attenuation of the feedback network, is a serious handicap in making the subtle trade-offs that are associated with the maximum obtainable feedback response.

Summarizing, we can gather the demands on the frequency response of integrated amplifiers into two categories. Technological limitations govern the first:

1. *No inductors and only small capacitors.*
2. *Robustness against parameter variations.*

The second group of demands follows from the requirement of universal use:

3. *Absolute stability with sufficient phase and gain margin under a variety of load and feedback conditions.*
4. *As high a loop gain as possible at low frequencies.*

Point 4 reflects the fact that we cannot properly make the trade-off of Eq. 3-17 between the gain in the passband and the frequency to which the passband extends, without knowledge of the application. Therefore the choice is to provide as much gain as the circuit allows at low frequencies. It is clear that the resulting corner frequency of the passband will be correspondingly low. This is not as serious a drawback as it would seem at first glance, since the closed-loop bandwidth will lie well above this corner frequency. The magnitude of the loop gain only affects the accuracy of the final transfer function, reflected, for instance, by the non-linear distortion. The choice for a high loop gain at low frequencies therefore implies good accuracy at low frequencies, which is often more needed than accuracy in the upper portion of the frequency band.

An additional guideline stems from the fact that combinations of poles and zeros are highly undesirable in the open-loop transfer function. The reason is that such combinations —pole-zero doublets— cause so-called slow-settling components in the step response of the amplifier [8]. These components, signals that fade away slower than one can expect from the bandwidth of the amplifier, very much restrict the applicability of the amplifier in circuits where settling is of prime importance, e.g. sample-and-hold circuits. Section 6.2 addresses the background of pole-zero doublets in more detail. Since without interlaced poles and zeros the gain roll-off is a multiple of 20dB/dec and the asymptotic slope for high frequencies must be less than the limiting case of 40dB/dec, the only option is a first order, 20dB/dec, gain characteristic from the corner frequency upwards. This leads to the fifth and final requirement:

5. *An open-loop frequency response with a 20 dB/dec roll-off.*

Figure 3-13 shows the corresponding characteristic. The maximum D.C. gain that the number of stages in the circuit can realize sets the low frequency gain. From the corner frequency ω_0, the gain curve descents at a rate of 20dB/dec. The corresponding phase shift is 90°, indicating a phase margin of also 90°. At some frequency ω_c the gain curve will intersect with the asymptotic gain limit. At that point the desired gain is higher than that of the transistors driving into their own parasitic capacitances. Above the intersection frequency the gain curve will therefore follow the gain asymptote. Since the order of the gain asymptote is (ideally) equal to the

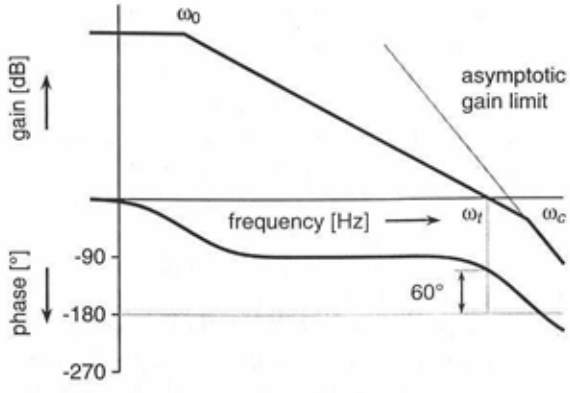


Fig. 3-13. *First order frequency response for use in integrated amplifiers.*

number of stages N, while the roll-off just below the intersection frequency is first order, a total of N - 1 poles will be present at the intersection frequency ω_c itself.

Following the same procedure as in the previous section, we can extend the region in which the phase margin equals 90° to a frequency very close to the point where the parasitic capacitances take over. But given the fact that a phase margin of 90° is far more than required in any application, a good alternative is to simply accept the poles that appear at high frequencies and adjust the unity-gain frequency ω_t in such a way that we obtain an acceptable phase margin. Figure 3-13 shows the result. The plot represents the characteristic of a three-stage amplifier, suggesting a third order gain asymptote and two poles at frequency ω_c. Aiming at a phase margin of 60°, the two poles may maximally contribute 30° to the 90° phase shift of the dominant pole at ω_0. Since the additional phase shift is 90° already at ω_c, we must choose the unity-gain frequency appreciably below the two poles. Section 5.1.1 and Section 5.2.1 will discuss the exact dimensioning of the frequency response. The analysis will show that slightly complex poles at ω_c reduces the ratio of ω_t and ω_c to approximately a factor 2, guaranteeing a unity-gain frequency close to the asymptote intercept.

The five basic demands on the frequency compensation of integrated circuits will serve as a road map to successful compensation strategies. Compliance with all five of them is a necessary requirement for the design of mass-production feedback integrated circuits. It is important to realize that many other considerations also attribute to the effectiveness of a frequency compensation technique. Aspects that the five demands do not express are, for example, how much power it takes to realize a given bandwidth or how complicated the resulting circuit will be. The list is definitely not exhaustive. The demands can serve, however, as the fundament from which to start the investigation into frequency compensation methods of integrated feedback amplifiers.

3.2.3. Conditional stability

The need for universal applicability limits the useful bandwidth of feedback that an amplifier can realize. If for certain specialized applications absolute stability is not an issue, other frequency responses are possible that more aggressively set the bandwidth of the useful band. Consider for example the Bode plot of Fig. 3-14. It shows two different characteristics,

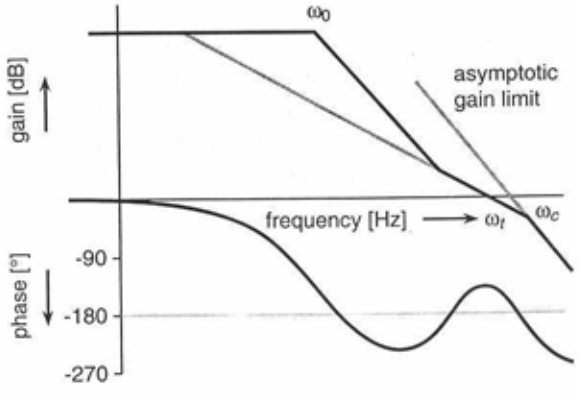


Fig. 3-14. *Bode plot demonstrating the enlarged useful bandwidth of a conditionally stable response.*

each having the same unity-gain frequency ω_t. The light curve depicts the first order characteristic of the absolutely stable amplifier of Fig. 3-13. The magnitude of the gain between ω_0 and ω_t is inversely proportional to the frequency. From that, knowing the low frequency gain A, we can find the useful bandwidth of the absolutely stable amplifier as

$$\omega_0 = \frac{\omega_t}{A} \tag{3-20}$$

The dark line shows a conditionally stable response. Three dominating poles appear at ω_0, followed by two zeros just below the 0dB frequency ω_t. As a result, the characteristic falls at the steep rate of 60dB/dec, but because of the two zeros it crosses the 0dB axis at a mere 20dB/dec. Although the phase shift becomes greater than 180° at certain frequencies in the passband, at the unity-gain frequency it is less than this critical number. The circuit is conditionally stable. The Nyquist diagram that corresponds to Fig. 3-14 is similar to that previously shown in Fig. 3-5. That the circuit is not absolutely stable becomes clear when we reduce the gain, resulting in the total gain magnitude plot shifting downwards. At some point in this process the unity-gain frequency, that reduces along with the gain, will enter the region where the phase margin is negative resulting in instability. The conditionally stable response offers much of the general applicability, since a reduction of the loop gain may be the result of an attenuation in the unknown feedback network.

The advantage of conditional stability becomes clear when we consider the useful bandwidth of the two solutions. Assuming that the first order section around ω_t is very small compared to the total roll-off characteristic, the useful bandwidth ω_0 of the conditionally stable amplifier equals

$$\omega_0 = \frac{\omega_t}{\sqrt[N]{A}} \tag{3-21}$$

where N is the number of stages and corresponding poles in the amplifier. In our example $N = 3$ and the roll-off is proportional to the third power of the frequency. Comparing Eq. 3-21 to Eq. 3-20, the increase of the useful bandwidth due to the conditionally stable response is $A^{1-(1/N)}$. Since the low frequency gain A is usually high, the improvement can be significant.

Suppose for example that in the three-stage case the gain in the useful band is 40dB, corresponding to factor 100. Then the conditionally stable frequency response increases the useful bandwidth ω_0 approximately 20 times.

The improvement obtained by the conditionally stable frequency response represents a special case of the attenuation integral theorem [1]. This theorem states that the total area under the gain magnitude plot is constant, as long as the added interstage shunt impedances have no resistive component at infinite frequencies. In this optimal case, the area is equal to that obtained from the parasitic interstage capacitances alone. In other words, there is a fixed gain fund that is available for distribution over the passband. It therefore pays off to reduce the gain above the useful band at a steep rate, as to ensure maximum use of the gain fund in the useful band.

Still, although the kind of improvements that are obtainable seem to make it worthwhile to reconsider the need for absolute stability, conditionally stable compensation schemes are restricted to a limited number of specialized applications. An example of their appearance is in audio amplifiers. There the useful band is limited to about 20kHz and a high loop gain is necessary to achieve low distortion figures. Section 6.5 will present a compensation technique that realizes a conditionally stable characteristic, while still meeting the four remaining demands. The rest of the work is fully dedicated to absolutely stable solutions.

3.3. Conclusions

This chapter introduced the Nyquist criterion for defining the stability of a circuit with respect to its open-loop properties. From the criterion we arrived at familiar concepts such as gain and phase margins. The asymptotic gain limit was introduced to establish the maximum obtainable feedback. This limit, set by the capacitors in the circuit, presents the upper bound of the attainable unity-gain frequency. To assure an appreciable phase margin the unity-gain frequency of any stable feedback amplifier must lie below that of the gain asymptote.

The freedom of component choice that the classical compensation practice allows, makes it possible to accurately realize frequency responses that achieve the maximum constant loop gain over the maxi-

mum useful bandwidth from a given set of active components. The fact that inductors of a useful value are not feasible in a monolithic integration technology greatly reduces the importance of this maximum feedback approach for present day circuits. The limitation on the use of coils, plus some additional characteristic of integrated circuit design, in Section 3.2.2 lead to five demands that must be fulfilled when realizing frequency compensation strategies for integrated circuits.

4 Basic Frequency Compensation of Integrated Circuits

Monolithic IC technology puts severe limitations on the options available for frequency compensation. A major restriction is the fact that inductors of any significant value cannot be integrated, while these components play an important role in the conventional compensation practice. This leaves capacitors as the only frequency dependent element to shape the frequency response. Unfortunately, even their use is limited to small values —much smaller than demanded by classical theory.

This chapter examines two types of frequency compensation methods: parallel and Miller compensation. We will investigate their fitness for integration. Important issues are the bandwidth-to-power ratio, noise, distortion and ruggedness against parameter variations. The traditional technique of parallel compensation appears to be of limited use for integrated circuits. Miller compensation, on the other hand, using local feedback to fix its properties, better complies with the five demands on the frequency compensation of integrated amplifiers which were introduced in Section 3.2.2.

4.1. The single-stage case

A particularly simple case of frequency compensation occurs for one-stage amplifiers. In this situation there is only one dominating pole and no additional measures are necessary to realize a first order frequency response. Consider the circuit of Fig. 4-1. It shows the MOS and bipolar situation of

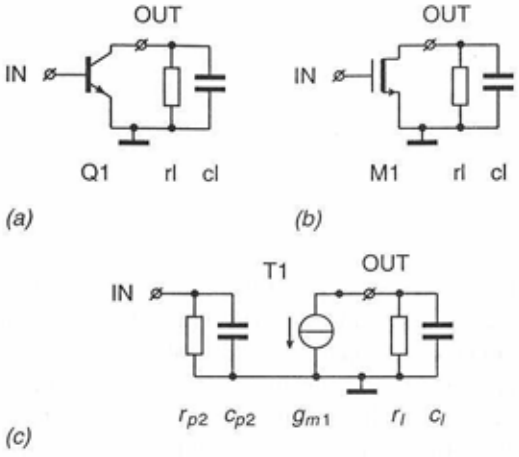


Fig. 4-1. *A one stage amplifier with* (a) *the bipolar circuit,* (b) *the MOS circuit and* (c) *the small-signal diagram.*

a single transistor, loaded by the parallel combination of a resistance and a capacitance. Figure 4-1*c* depicts the small-signal diagram. Evidently, in the MOS case the input resistance of the transistor approaches infinity. Figure 4-2 shows the Bode plot of the voltage-to-voltage transfer. The transconductance of the transistor and the value of the load resistance set the magnitude of the gain at low frequencies. The gain equals $A = g_{m1} r_l$. At the cross-over frequency of the load resistance and capacitance, $\omega_0 = 1/(r_l c_l)$, a pole arises and from there the gain plot descends at a 20dB/dec rate. We can find the unity-gain frequency as

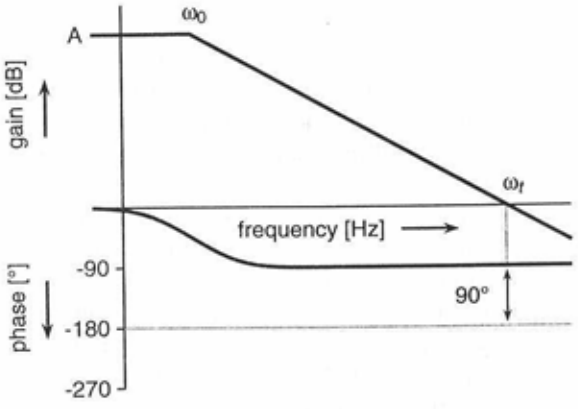


Fig. 4-2. *Bode plot of a single-stage amplifier.*

$$\omega_t = \frac{g_{m1}}{c_l} \tag{4-1}$$

Since there is only one pole in the circuit, the phase margin is close to 90°. Some additional phase lag will occur, due to the presence of higher order poles, e.g. from the base resistance and input capacitance of a bipolar transistor. This phase shift is normally small, however, and a single stage feedback circuit is likely to be stable under most circumstances.

It is interesting to investigate the situation when there is no capacitance connected to the output (collector or drain) of the transistor. According to Eq. 4-1, this would lead to an infinite bandwidth of the circuit. From a physical point of view this is quite impossible. The fact that the feedback network will also load the output limits the minimum load capacitance. Assuming for example a unity-gain configuration, where the output of the circuit is directly coupled to the input, then we must add the base-emitter (gate-source) capacitance of the transistor to the capacitance connected to the output. In this ideal situation, we can establish the unity-gain frequency established as

$$\omega_t = \frac{g_{m1}}{c_{p2}} \tag{4-2}$$

where c_{p2} is the input capacitance of the transistor. Notice that the unity-gain frequency of the amplifier equals the gain-bandwidth product GB of the transistor. Since a one-stage amplifier has a first order gain asymptote, this implies that the 20dB/dec slope of the frequency response exactly coincides with the gain asymptote.

Examples of feedback circuits with one transistor are the basic local feedback stages, such as the shunt and the series stage. But also the common collector or common drain stage (voltage follower) and the common base or common gate stage (current follower) consist of a single transistor with (unity-gain) feedback around it. In the latter two circuits it is a well-known fact that their bandwidths, which are determined by the unity-gain frequencies of the open loops, equal the gain-bandwidth products of the transistors. This corresponds to the prediction of Eq. 4-2.

4.2. Two-stage parallel compensation

In many applications the gain of a single transistor does not suffice. In those situations one must add extra stages to the circuit. Along with the introduction of the extra transistors, the issue of frequency compensation arises. The presence of two dominating poles introduces an asymptotic phase shift of 180°, which only in theory assures a stable amplifier. Of the two basic types of frequency compensation, parallel and Miller compensation, this section focuses on the first. Parallel compensation bases itself on the traditional technique of placing a parallel interstage impedance in the circuit. Since we do not have inductors at our disposal, the impedances will consist of combinations of capacitors and resistors alone.

As a starting point, Fig. 4-3 shows the simplified schematic of a two-stage amplifier. As before, the figure depicts both the bipolar and the CMOS variety. The small-signal diagram of Fig. 4-3*c* is valid for both circuits. For bipolar transistors, the resistance r_{p2} consists of the parallel combination of the output resistance of the transistor Q_2 and the base-emitter resistance of the output transistor Q_1. For the CMOS circuit of Fig. 4-3*b*, on the other hand, the resistance r_{p2} only models the output resistance of M_2. Since two single-transistor stages would yield a positive

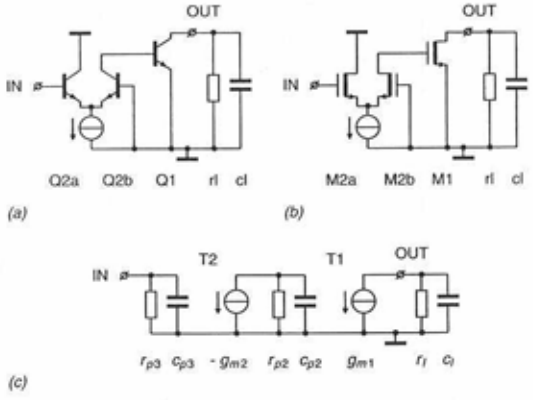


Fig. 4-3. *Simplified schematic of a two-stage amplifier.*

feedback sign, the input stage of the two-stage amplifier consists of a differential pair. The minus sign of the transconductance $-g_{m2}$ in the small-signal diagram accounts for the phase reversal of this stage compared to a single transistor. The absolute value of this transconductance equals half the transconductance of each side of the input pair. Hence

$$g_{m2} = \frac{1}{2} g_{m2a} = \frac{1}{2} g_{m2b} \tag{4-3}$$

Also the input impedance $r_{p3}//c_{p3}$ equals half that of the individual transistors T_{2a} and T_{2b}.

Figure 4-4 depicts the Bode plot of the circuit. The two stages together contribute two dominating poles at $p_1 = -1/(r_{p2}c_{p1})$ and $p_2 = -1/(r_l c_l)$, and the circuit is in need of compensation. Since the gain magnitude plot above the second pole defines the gain asymptote, we can draw the desired first order characteristic as the light line in Fig. 4-4. The frequency response guarantees a sufficient phase margin, since the unity-gain frequency ω_t lies well below the secondary pole frequency ω_c. The difference between the light and the solid line symbolizes the required effect of the frequency compensation.

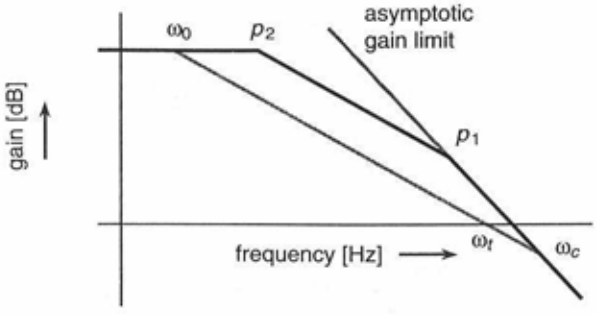


Fig. 4-4. *Bode plot of a two-stage amplifier. The solid line gives the frequency response without compensation, the light line indicates the best obtainable response with frequency compensation.*

Inserting a series combination of a resistor R_c and a capacitor C_c between the two transistors, as Fig. 4-5 shows, realizes parallel compensation of the circuit. These components, when of the right value, reduce the midband gain of the amplifier and transform the frequency response into a first order characteristic. The Bode plot of Fig. 4-6 illustrates this. The insertion of the resistor and the capacitor introduces two additional poles p_{c1}, p_{c2} and two additional zeros z_{c1}, z_{c2} into the Bode plot. The location of the poles is approximately

$$p_{c1} = -\frac{1}{r_{p2}C_c}$$
$$p_{c2} = -\frac{1}{R_c c_{p2}} \tag{4-4}$$

while the zeros follow from

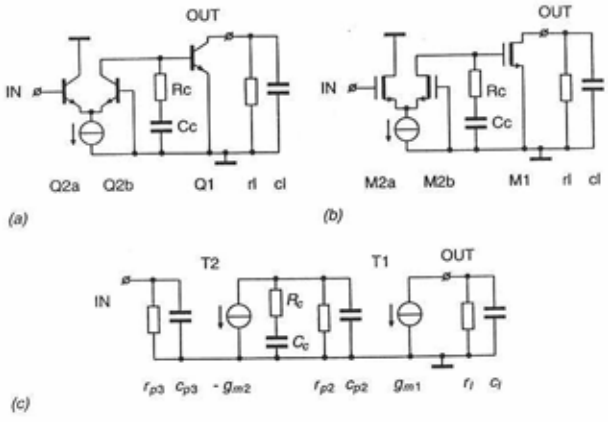


Fig. 4-5. *Two-stage amplifier with parallel compensation.*

$$z_{c1} = -\frac{1}{r_{p2}c_{p2}}$$
$$z_{c2} = -\frac{1}{R_c C_c} \tag{4-5}$$

As follows from inspection of Fig. 4-6, we can obtain the desired first order response if we place the compensation zero z_{c2} exactly on top of the pole p_2 of the uncompensated amplifier. This requires

$$R_c C_c = r_l c_l \tag{4-6}$$

In other words, the time constant of the compensation impedance must equal that of the load impedance.

The time constant of the compensation resistor R_c and the inter-stage capacitance c_{p2} (pole p_{c2}) determines the position of the secondary pole, which ultimately sets the obtainable unity-gain frequency. This leads to the second expression for the dimensioning of the parallel compensation circuit. A small resistor R_c results in a high secondary pole, but reducing

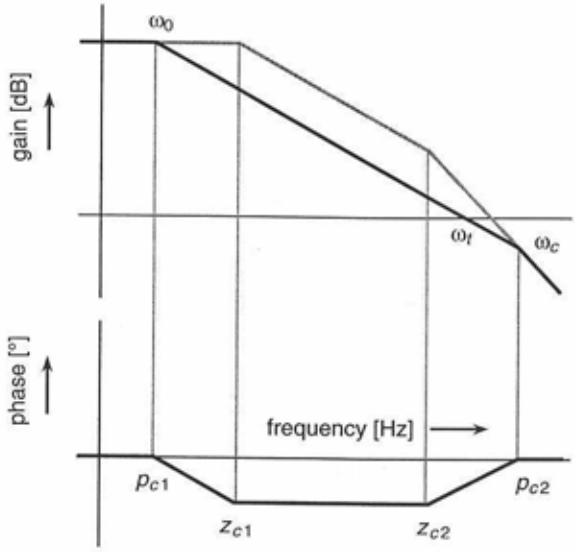


Fig. 4-6. *Bode plot showing the shaping of the frequency response by the parallel compensation components.*

R_c also reduces the unity-gain frequency, as it causes the gain plot to shift down. We can find the optimal value for R_c by noticing that, for a phase margin of 60°, the secondary pole must lie by a factor of two higher than the intended unity-gain frequency ω_t of the amplifier. Section 5.1.1 will elaborate on that.

$$p_{c2} = -2\omega_t \tag{4-7}$$

We can readily find the unity-gain frequency ω_t as $\omega_t = (g_{m1}g_{m2}r_l)/C_c$. Therefore, we are able to rewrite Eq. 4-7 as

$$\frac{1}{R_c c_{p2}} = 2\frac{g_{m1}g_{m2}r_l}{C_c} \tag{4-8}$$

which is yet another expression involving R_c and C_c. Solving the equations 4-6 and 4-8 yields for the values of the compensation components

$$R_c = \sqrt{\frac{c_l}{2c_{p2}g_{m1}g_{m2}}}$$
$$C_c = r_l\sqrt{2c_{p2}c_l g_{m1}g_{m2}} \tag{4-9}$$

and for the unity-gain frequency ω_t we find

$$\omega_t = \sqrt{\frac{g_{m1}g_{m2}}{2c_{p2}c_l}} \tag{4-10}$$

The unity-gain frequency that follows from Eq. 4-10 is a remarkably good result. Comparing it to the generalized Eq. 3-13, which gives the 0dB frequency of the asymptotic gain limit, indicates that the unity-gain frequency of the amplifier is only a factor $\sqrt{2}$ lower than the frequency where the gain of the asymptote becomes unity. This factor is a direct consequence of the desired phase margin. The conclusion is that parallel compensation in a two-stage amplifier achieves the theoretical maximum unity-gain bandwidth.

In the analysis of the parallel compensation we neglected the effect of the collector-base (drain-gate) capacitances. This very much simplifies the calculations, but is not always justifiable. Especially the feedback capacitance of the output transistor T_1 can influence the properties of the circuit. The reason is that, to the first order, the voltage gain of the stage T_1 amplifies the effect of this capacitance; the so-called Miller effect [9]. In cases where the voltage gain is large, this effect may easily dominate the effective capacitance seen at the base (gate) of T_1. In those situations the dimensioning equations we arrived at earlier do not apply anymore. Inserting a cascode transistor in the output terminal of the amplifier can overcome this problem. The low input impedance of the cascode stage reduces the voltage swing across the collector-base (drain-gate) capacitance and thus eliminates the Miller effect. This improvement is at the cost of additional die area and power consumption, however.

Unfortunately there are some disadvantages associated with the bandwidth capability of parallel compensation. These become apparent when putting the design equation Eq 4-9 to the test of the demands on the

frequency compensation of integrated circuits of Section 3.2.2. Parallel compensation fulfills some of the rules, but two are difficult to comply with. The first drawback is that the compensation capacitor C_c may have to assume large values. The second relates to the practical inability to control all the parameters on which the dimensioning depends. The large values of the compensation capacitor follow directly from the dimensioning equation. Suppose for example the following typical data: $r_l = 10\text{k}\Omega$, $c_l = 10\text{pF}$, $g_{m1} = g_{m2} = 4\text{mA/V}$ and $c_{p2} = 2\text{pF}$. In this list, the transconductances g_{m1} and g_{m2} correspond to bipolar transistors at a collector current of 100 µA. Substituting the values in Eq. 4-9, leads to a compensation capacitor of 250pF, which is too large in most cases to consider integration on a silicon die.

Under certain conditions the required capacitor C_c might boil down to an acceptable value. Especially when the load capacitance and the capacitance are small, which is likely when the output is to drive another on-chip circuit part only, the value of the capacitor will not be a limitation. But even under these circumstances, there remains the problem of guarding the amplifier against variations of the circuit parameters; the second drawback of parallel compensation. According to Eq. 4-9, the dimensioning of the compensation components depends on internal parameters, such as transconductances and parasitic capacitances, as well as the load that is connected to the amplifier. The latter observation implies that we must tailor the compensation network to a specific load capacitance and load resistance, which contradicts with the desired general applicability. Moreover, when the compensation components do not exactly match the impedance connected to the output, a pole-zero doublet appears in the passband. This complication, which gives rise to slow-settling components in the step response of the amplifier, originates from the introduction of the zero z_{c1} by the compensation network to cancel the pole p_{c1} of the time constant at the output. The exact positioning of the zero on top of the pole led to Eq. 4-6. Any change of the time constant of the load with respect to the compensation components introduces a closely spaced combination of a pole and a zero in the open-loop transfer function. In Section 6.2 we will investigate the effect of such a pole-zero doublet in more detail.

Another consequence of the large number of parameters that enter the dimensioning expressions is that the compensation heavily resides on quantities that lack physical correlation, and therefore have very poor matching properties. This makes it very difficult to predict the yield of the amplifier in large production volumes. The only way to be sure that most

of the realized amplifiers meet the specifications is an overly conservative choice of the compensation components. We can do this, for instance, by choosing a higher value than 2 for the gain margin factor that appears in the equations. However, such a conservative dimensioning counteracts the theoretically high bandwidth potential of parallel compensation.

4.3. Miller compensation

The great number of complications of the traditional parallel compensation when applied to mass-production integrated circuits has forced the appearance of other techniques which are better suited to realize the desired frequency response under the present conditions. Most of these methods base themselves on the use of feedback around single transistors in the amplifier.

An example of such a local feedback approach, which goes back to long before the integration of circuits on silicon, is the use of emitter or source followers in the output stage. The internal feedback of this configuration transforms the transistor into an almost ideal voltage buffer, with a high input impedance, low output impedance and a high bandwidth. Thus, the voltage follower effectively isolates the load from the rest of the amplifier and facilitates the frequency compensation. A drawback of this approach is that the high frequency advantages go at the cost of a reduced gain at D.C. under certain circumstances. Especially if the output faces a high resistive load impedance, the emitter (source) follower supplies considerably less gain than a common emitter stage in the same location. The trade-off between bandwidth and gain is very similar to the one that occurs in resistive broadbanding. Small resistors in parallel to internal capacitances to shift the corresponding poles to higher frequencies improves the phase margin, but the low-valued resistors drastically reduce the gain at low frequencies. Another disadvantage of the voltage follower is that its base-emitter (gate-source) voltage drop increases the minimum supply voltage at which the amplifier can operate. Since low voltage is a necessary requirement for a low power consumption, this can be a serious handicap for low-power circuits. The first section of the next chapter will further elaborate on this subject.

We can realize better amplifier characteristics by using capacitors in the local feedback path. An example of this approach is Miller compen-

sation, also referred to as pole splitting, which uses a capacitor as the feedback impedance of a shunt stage. Since the capacitor does not affect the D.C. gain of the transistor, this choice effectively ensures a high loop gain at low frequencies. Another option would be introducing an inductor in a series stage, but this is clearly not the preferable choice for integration on a silicon die.

The introduction of a Miller capacitor across the output stages Q_1 and M_1 of Fig. 4-3 leads to respective bipolar and MOS circuits shown in Figure 4-7. To investigate the behavior at frequencies above D.C.,

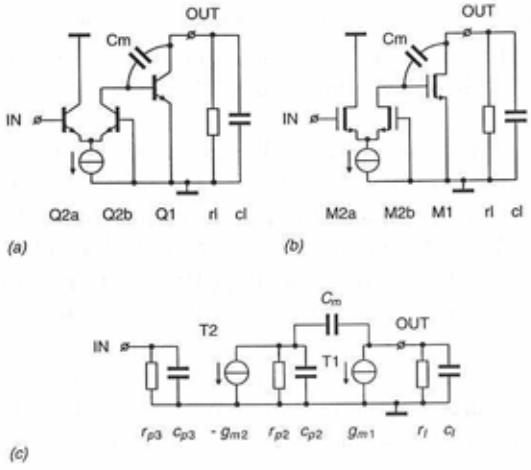


Fig. 4-7. *Two-stage amplifier circuit with Miller compensation.*

Fig. 4-7*c* depicts the corresponding small-signal diagram. As before, the diagram is valid for both circuits. We can introduce the zero gate-source conductance of a MOS device into the following analysis by making the proper choice for the impedances in the small-signal diagram. Solving the nodal equations leads to the following voltage transfer

$$\frac{v_{out}}{v_{in}} = \frac{g_{m1}g_{m2}r_{p2}r_l\left(1 - s\frac{C_m}{g_{m1}}\right)}{s^2 r_l r_{p2}(c_l C_m + c_l c_{p2} + C_m c_{p2}) + s(c_l r_l + C_m r_l + C_m r_{p2} + c_{p1} r_{p2} + C_m g_{m1} r_l r_{p2}) + 1} \tag{4-11}$$

which is a formidable result for such a small circuit. Inspection of Eq. 4-11 reveals that the transfer counts one zero and, because of the second order polynomial in the denominator, two poles. The location of the zero is

$$z = \frac{g_{m1}}{C_m} \tag{4-12}$$

which indicates a zero in the right half of the complex s-plane. Solving of Eq. 4-11 for the poles is more complex. Fortunately, we can greatly simplify the calculations by suggesting that the two poles are widely spaced. If we further assume that $r_{p1} \gg 1/g_{m1}$, which is equivalent to a high current gain, and $r_l \gg 1/g_{m1}$, then the roots of the denominator boil down to

$$\begin{aligned} p_1' &= -\frac{g_{m1}}{c_l\left(1 + \frac{c_{p2}}{C_m}\right) + c_{p2}} \approx -\frac{g_{m1}}{c_l} \\ p_2' &= -\frac{1}{r_l r_{p2} C_m g_{m1}} \end{aligned} \tag{4-13}$$

The approximation in the first equation is valid in the common situation that the load capacitance c_l and the feedback capacitor C_m are large compared to the interstage capacitance c_{p2}. One can easily confirm that under practical circumstances the poles p_1' and p_2' indeed lie far apart. Substituting the typical values of the previous section, for instance, results in the following poles: $p_1' = -240 \cdot 10^6$ and $p_2' = -250 \cdot 10^3$, which respectively corresponds to 37MHz and 40kHz. The separation is three orders of magnitude. This result supposes a compensation capacitor of 4pF. Figure 4-8 shows the effect of the capacitor C_m in a Bode plot. The light line indicates the frequency response of the amplifier without frequency compensation, while the black curve represents the response with the compensation capacitor. Inserting the capacitor splits the two original poles p_1 and p_2 and places them on their new positions p_1' and p_2'. The result is the much

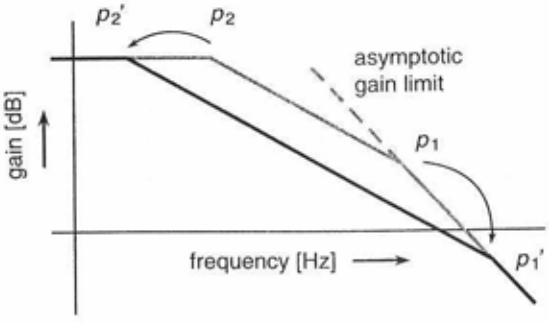


Fig. 4-8. *Bode plot showing the movement of the poles due to the Miller capacitor.*

desired first-order response, with p_2' as the dominating pole frequency, up to the secondary pole p_1'. For simplicity, Fig. 4-8 does not show the effect of the zero z in the right half of the complex s-plane. Section 6.4 will probe further into the consequence of the right half plane zero to the frequency response and obtainable bandwidth of a Miller compensated amplifier.

We can most easily calculate the unity-gain frequency from the location of the dominating pole p_2' and the D.C. gain of the amplifier. The latter follows from Eq. 4-11 by letting $s = 0$, and equals $g_{m1}g_{m2}r_{p2}r_l$. The result is

$$\omega_t = \frac{g_{m2}}{C_m} \tag{4-14}$$

This equation is particularly simple. Most of the parameters used for its derivation have vanished.

The elementary expressions Eq. 4-13 and Eq. 4-14 describing the Miller amplifier, suggest an alternative —intuitive— way to investigate the properties of capacitive feedback. Consider the schematic representation of the two-stage amplifier in Fig. 4-9. Here, ideal transconductors represent the transistors. Across the output stage T_1 again the capacitor C_m is

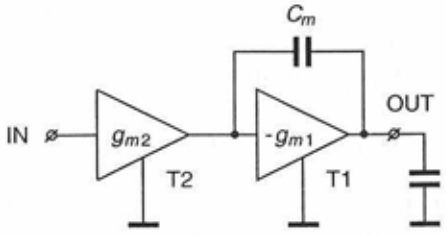


Fig. 4-9. *Miller compensated amplifier with idealized transconductors.*

connected. The combination of the gain stage T_1 and the capacitor constitute a basic integrator. Since the feedback around stage T_1 keeps the voltage at its input zero, all the current coming from the input stage T_2 flows into the feedback capacitor and forces the output voltage to change accordingly. Hence, the current-to-voltage transfer is $-1/(sC_m)$. Including the transconductance of the input stage we find

$$\frac{v_{out}}{v_{in}} = \frac{g_{m2}}{sC_m} \tag{4-15}$$

Setting the voltage gain to one, this expression leads to the unity-gain frequency of Eq. 4-14. We can further use Eq. 4-15 to determine the low-frequency pole p_2'. Figures 4-8 indicates that the flat gain section and the 20dB/dec slope sets this pole. The intersection of these two lines is where pole p_2' must be located. Since we found the D.C. gain to be $g_{m1}g_{m2}r_{p2}r_l$ and Eq. 4-15 defines the first order slope, one can easily confirm the result of Eq. 4-13.

Explanation of the high frequency pole p_1' requires considering the frequency where the gain around the loop of the integrator drops below one. Above this frequency the 20dB/dec roll-off of the integrator ends, since the feedback loses its effectiveness and cannot force the first order response on the input-output relation anymore. The transfer function will reduce to the (second order) gain asymptote of the circuit. The gain of a single transistor with a transconductance g_{m1} driving a load capacitance c_l

is unity at $\omega = g_{m1}/c_l$. Assuming that no attenuation of the signal occurs in the feedback network, this will also be the frequency of the secondary pole p_1'. The value g_{m1}/c_l matches the approximation of Eq. 4-13. In fact, the assumption that there is no attenuation, corresponds to the conditions that led to this approximation. The more complex expression for p_1' of Eq. 4-13 merely takes into account the attenuation caused by the capacitive voltage divider C_m and c_{p2} and the loading of the transistor by the feedback network.

For a minimum phase margin of 60°, the unity-gain frequency ω_t of the amplifier must be at least a factor two lower than the secondary pole p_1'. This fact plus the relations of Eq. 4-13 and Eq. 4-14 enable to set up the dimensioning formula for the two-stage local feedback amplifier

$$C_m \geq 2\frac{g_{m2}}{g_{m1}}c_l \tag{4-16}$$

where the equality holds in the limiting case of 60° phase margin. The corresponding unity-gain frequency is

$$\omega_t \leq \frac{g_{m1}}{2c_l} \tag{4-17}$$

Apart from the load capacitance, Eq. 4-16 only depends on the *ratio* of two transconductances. The nature of I.C. processing guarantees a near perfect matching of such a ratio. Depending on the technology, bipolar or CMOS, the matching typically ranges between 0.1% and 1%. Very predictable results are to be expected, even across fabrication batches. The inequality sign of Eq. 4-16 lies at the base of a second property of capacitive feedback compensation. When we assume the compensation capacitor C_m to be fixed, the equation states that the amplifier will be stable for any load capacitor smaller than a limiting value. In other words, if we design a feedback compensated amplifier for a certain load capacitance, it will also be stable with any capacitive load smaller than this value. As a final property, Eq. 4-16 shows that in general small capacitors suffice for compensation. Picking up the example from Section 4.2 again, local feedback compensation only requires a capacitor of 20pF. We can obtain an even

lower value when we chose the transconductance of the output stage higher than that of the input stage. Under that condition, the feedback capacitor will be in the order of several pF.

An important advantage of Miller compensation is that it requires no additional cascode stages to neutralize the effect of the collector-base (drain-gate) capacitances of the output transistor. The reason is that the relatively small parasitic capacitance is in parallel with the deliberate compensation capacitor C_m. This aspect prevents the extra die area and power consumption associated with the cascode transistors needed with parallel compensation.

4.3.1. Optimal dimensioning of the output stage and Miller capacitor

The value obtained for the compensation capacitor in Eq. 4-16 is based on the approximation in Eq. 4-13. This approximation on its turn originated from the assumption that there is no attenuation due to the capacitive voltage division of the feedback capacitor C_m and the interstage capacitance c_{p2}. Although this may be true in many practical situations, an interesting optimization problem occurs by taking into account that the interstage capacitance c_{p2} depends on the transconductance of the output stage T_1.

For bipolar transistors the *GB* product, which is nearly constant over a large range of operating currents, governs the relation between transconductance and input capacitance. The *GB* product equals g_m/c_{be} and therefore, neglecting the output capacitance of T_2, the interstage capacitance c_{p2} is proportional to the transconductance. The transconductance g_{m1} appears in the numerator of Eq. 4-13, where it is counteracted by the c_{p2} entry in the denominator. Since the transconductance linearly depends on the collector current, the question is: given a compensation capacitor C_m, what quiescent current results in the highest secondary pole frequency p_1' and hence the highest obtainable bandwidth?

The relation between the transconductance and the capacitance of a MOS device is directly connected to the geometry of the transistor. Disregarding the overlap capacitance, the gate-source capacitance c_{gs} is proportional to the area, i.e. the product of the width W and the channel length L. The transconductance, however, is proportional to the square root of the ratio of the width and the length; $g_m \propto \sqrt{W/L}$. Substituting these relations into Eq. 4-13, the first thing that we can conclude is that the optimum channel length L is clearly the minimum that the fabrication process permits. This choice guarantees both the highest transconductance and lowest

gate-source capacitance. When the channel length L is fixed at its minimum, both the numerator and the denominator of Eq. 4-13 are proportional to the transistor width W. Hence, for MOS we arrive at the following question: given a compensation capacitor C_m, what width of the output transistor yields the best bandwidth?

The optimization problems for the bipolar and the CMOS circuit are similar, but the results are quite different. The solution in either case requires substituting the transconductance and capacitance relations into the expression for the pole position. Differentiation to the collector current and the width of the channel respectively and equating to zero yields the desired result. The calculations become much simpler when we assume that the interstage capacitance c_{p2} is much smaller than the load capacitance. This is equivalent to neglecting the loading of the amplifier by the feedback network.

In the bipolar case it appears that there is no local maximum and the secondary pole monotonically rises with the collector current. The higher the current through the transistor, the higher the obtainable bandwidth. At high currents, however, when the base-emitter capacitance is much larger than the compensation capacitor, the improvement resulting from an increase of the collector current becomes very small. The pole function approaches an asymptote given by

$$\lim_{I_c \to \infty} p_1' = \frac{g_{m1} C_m}{c_{p2} c_l} = GB \frac{C_m}{c_l} \tag{4-18}$$

We arrive at a practical trade-off between the diminishing return on a current increase and the bandwidth when we choose the quiescent current such that $c_{be} = C_m$. In this case the pole position p_1' is a factor two lower than the limiting case of Eq. 4-18.

Unlike the bipolar circuit, the Miller compensated amplifier with CMOS transistors does have a well defined maximum for the bandwidth. The circuit reaches this maximum when we set the width W of the transistor so that

$$c_{gs} = C_m \tag{4-19}$$

This expression matches the compromise that at which arrived for the bipolar case. The maximum of Eq. 4-19 is relatively broad.

We can summarize the results of both technologies as follows: a well-balanced circuit with capacitive feedback for frequency compensation, requires that the compensation capacitor is of the same order of magnitude as the input capacitance of the compensated transistor. Under that condition, the secondary pole p_1' lies at half the maximum obtainable frequency given by $p_1' = g_{m1}/c_l$.

4.3.2. Current-mode Miller compensation

By interchanging the stages of the Miller compensated circuit of Fig. 4-7, we obtain an amplifier that accepts a current as the input signal and also delivers a current to the load. Figure 4-10 shows the circuit. We can simply

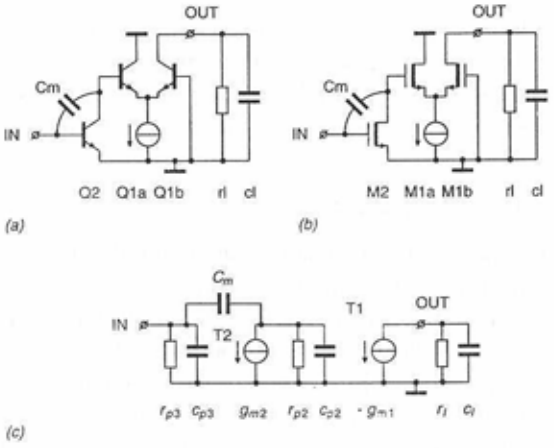


Fig. 4-10. *Simplified circuits of current-mode Miller compensation and the small-signal diagram.*

derive the dimensioning equations from the expressions for the voltage-to-voltage amplifier of Fig. 4-7. The basic observation is that the two amplifiers have identical loop gain characteristics when we connect their outputs to the respective inputs. This situation corresponds to unity-gain feedback. Comparing the parameters that determine the loop gains of the two cir-

cuits, we find that the function of g_{m1} and g_{m2} as well as that of c_l and c_{p2} have changed position. Therefore, we can reformulate the design equations of Section 4.3 into

$$C_m \geq 2\frac{g_{m1}}{g_{m2}}c_{p2} \tag{4-20}$$

$$\omega_t \leq \frac{g_{m2}}{2c_{p2}} \tag{4-21}$$

4.4. Parallel vs. Miller compensation

The two different types of frequency compensation, parallel and Miller, each have their own specific pros and cons. In the following we will compare the performance potential of the techniques. Points of attention are the bandwidth-to-power ratio, noise and distortion.

4.4.1. Bandwidth-to-power ratio

The expressions for the obtainable bandwidth of parallel and Miller compensation all involve the transconductances of the transistors. The higher the transconductances the higher the unity-gain frequency that the circuit can attain. Since the transconductances of bipolar and MOS devices directly depend on their bias currents, a relation exists between the power that the amplifier consumes and its bandwidth. This relation is expressed by the bandwidth-to-power ratio. Examining of this ratio for the idealized parallel and Miller compensated circuits of the previous sections gives an indication about the upper limit for the bandwidth that we can obtain from a fixed amount of power.

The dissipated power P of a two-stage amplifier equals the sum I_s of the two bias currents I_1 and I_2 times the supply voltage

$$P_s = V_s(I_1 + I_2) = V_s I_s \tag{4-22}$$

Since the unity-gain frequencies of the different compensation techniques rely on the transconductances of the transistors, we need a relation between the current in the device and the transconductances. The general relation is

$$g_m = g_m(I) \tag{4-23}$$

For a bipolar transistor, the transconductance is proportional to the collector current I_c. It obeys the well-known expression

$$g_m(I_c) = \frac{qI_c}{kT} = \frac{I_c}{V_T} \approx \frac{I_c}{26\text{mV}} \tag{4-24}$$

The approximation on the right of Eq. 4-24 holds at room temperature. The transconductance of a MOS device in strong inversion is proportional to the square root of the drain current

$$g_m(I_d) = \sqrt{2I_d \frac{W}{L} \mu C_{ox}} \tag{4-25}$$

where W is the channel width, L the channel length, μ the mobility of the charge carriers in the channel and C_{ox} the capacitance per unit area of the gate. The relation between the current and the transconductance becomes linear when we keep the current density in the channel constant by scaling the channel width W proportionally to the drain current. If the channel width per unit current is $w = W/I_d$, Eq. 4-25 turns into

$$g_m(I_d) = I_d \sqrt{2 \frac{w}{L} \mu C_{ox}} = \frac{I_d}{V_M(J_d)} \tag{4-26}$$

Although the structure of Eq. 4-26 very much resembles the expression for a bipolar transistor, Eq. 4-24, it is fundamentally different. The relation between the collector current of a bipolar transistor and its transconductance, given by the thermal voltage V_T, is fully governed by physical constants and the temperature. The voltage V_M of a MOS transistor on the other hand depends on the current density in the channel. This is reflected by the fact that the voltage V_M in the denominator on the right hand side is a function of the current density $J_d = I_d/W$. Furthermore, V_M depends on the device's geometry, processing constants and —through μ— the

temperature. In our analysis, however, for a given process and current density, the MOS expression Eq. 4-26 can be used in much the same way as its bipolar counterpart Eq. 4-24.

A. parallel compensation

Rewriting Eq. 4-10 to find the unity frequency of parallel compensation as a function of the transconductances of the transistors yields

$$g_{m1}g_{m2} = 2\omega_t^2 c_l c_{p1} \tag{4-27}$$

Substituting Eq. 4-23 into the left-hand term, we obtain

$$g_m(I_1)g_m(I_2) = 2\omega_t^2 c_l c_{p1} \tag{4-28}$$

For simplicity, two identical transistors are assumed. Since we are interested in the optimal unity-gain frequency at a given power (and hence a given total supply current), we need to know what division of the supply current over the two transistors gives the maximum left-hand term in Eq. 4-28.

One can easily verify that the product of any two identical monotonic rising function attains its maximum when their arguments are equal. In other words, for a maximum unity-gain frequency, the total supply current must divide evenly over the two transistor stages. The calculations involve expressing the two transistor currents I_1 and I_2 as a function of the total supply current and a division ratio r. Differentiation of the left-hand term of Eq. 4-28 to r and equating to zero leads to the desired result. In this derivation we assume that the interstage capacitance c_{p2} is independent from the current through the output transistor T_1. This seems to contradict the fact that the base-emitter capacitance of a bipolar transistor at a relatively high current is proportional to the current itself. When biased at a lower current, however, the base-emitter capacitance becomes constant. The capacitance of the base-emitter junction, which is independent of the collector current, dominates the diffusion capacitance of the transistor in this situation. Furthermore, the contribution of the parasitic collector-substrate capacitance of the input transistor T_2 to the total interstage capacitance c_{p2} will be large when the bias current of the output stage is low.

When we keep the current density of a MOS transistor constant, the gate-source capacitance is proportional to the transconductance. This is

the same relationship as for bipolar transistors. The linear dependency is due to the fact that increasing the width of the channel W to track a current increase, also enlarges the gate area and thus the capacitance. The gate-source capacitance follows from

$$c_{gs} = WLC_{ox} = I_d wLC_{ox} \tag{4-29}$$

From Eq. 4-5, which gives the obtainable unity-gain frequency for parallel compensation, we can conclude that increasing the bias current does not improve the unity-gain frequency of the amplifier when the interstage capacitance c_{p2} is linearly proportional to the transconductance of the output stage. Therefore, the situation that the fixed capacitances dominate over the variable capacitances, or at least are of the same order of magnitude, is favorable from the power point of view.

Since $I_1 = I_2 = (1/2)\, I_s$, we can rewrite Eq. 4-28 as

$$g_m^2\left(\frac{1}{2}I_s\right) = 2\omega_t^2 c_l c_{p1} \tag{4-30}$$

Substituting the expressions for the transconductance Eq. 4-24 and Eq. 4-26 and using the power equation of Eq. 4-22, we can finally determine the relations between the power and the bandwidth. In case of bipolar transistors we obtain

$$\frac{\omega_t}{P_s} = \frac{1}{V_s V_T}\frac{1}{\sqrt{2c_l c_{p1}}} \tag{4-31}$$

Equation 4-31 is the maximum *bandwidth-to-power ratio* of a two-stage bipolar parallel compensated amplifier. Since Section 4.2 indicated that the unity-gain frequency of parallel compensation approaches the theoretical maximum, the bandwidth-to-power ratio of Eq. 4-31 also is the limiting value for *any* two-stage bipolar amplifier. Clearly, the ratio is inversely proportional to the supply voltage: a lower supply voltage has no effect on the unity-gain frequency, but reduces the power consumption. Further, we can conclude from Eq. 4-31 that the load and intermediate capacitances are the only parameters of the circuit that determine the bandwidth-to-power ratio. This is an interesting result which makes it possible to easily test the unity-gain frequency of a particular amplifier against the optimum. In this way, we can use the bandwidth-to-power ratio as a figure of merit

for the performance. In the example of Section 4.2, we assumed an interstage capacitance c_{p2} of 2pF and a load capacitance of 10pF. Using the same values in Eq. 4-31 and further assuming a supply voltage of 3V, we obtain a bandwidth-to-power ratio at room temperature ratio of $\omega_t / P_s = 2.0 \cdot 10^{12}\ s^{-1}/W$ (= 323GHz/W).

The similarity of Eq. 4-24 and Eq. 4-26 suggests that the relation between the bandwidth and the power of a CMOS amplifier is of the same form as Eq. 4-31. The exact expression is

$$\frac{\omega_t}{P_s} = \frac{1}{V_s V_M(J_d)} \frac{1}{\sqrt{2 c_l c_{p1}}} \tag{4-32}$$

This is identical to the bipolar expression Eq. 4-31, except for the voltage V_M that takes the place of the thermal voltage V_T. Since V_M in Eq. 4-26 was found to be a function of the current density in the channel, it is interesting to investigate for which density it attains its minimum value, yielding the optimal bandwidth-to-power ratio.

Rewriting Eq. 4-26 to obtain V_M as a function of the current density J_d we obtain

$$V_M = \sqrt{\frac{J_d L}{2 \mu C_{ox}}} \tag{4-33}$$

According to Eq. 4-33, V_M becomes minimal in the limiting case that the current density J_d is zero. Under that condition the voltage V_M is zero too, which in Eq. 4-32 results in an infinitely great bandwidth-to-power ratio. Although being very favorable, evidently this situation cannot be realized in practice. At low currents the MOS devices will enter their subthreshold regions and display exponential voltage-current relations. The transconductances become proportional to the drain currents (and hence the current densities). The exact relation is [10]

$$g_m = \frac{I_d}{m V_T} \tag{4-34}$$

This equation very much resembles Eq. 4-24 which gives the relation for a bipolar transistor. The only difference is the occurrence of the constant m. This constant represents a worsening factor compared to the bipolar case.

It typically ranges between 1.1 and 1.5. Since Eq. 4-34 is independent from the current density, it is the upper bound for the ratio between the transconductance and the current of a MOS device. To obtain the maximum bandwidth-to-power ratio in strong inversion, the MOS devices should therefore be operated as close as possible to the subthreshold region. This requires setting the gate-source voltages V_{gs} approximately equal to the threshold voltage V_{th}.

B. Miller compensation

We can derive the bandwidth-to-power ratio of a two-stage Miller compensated amplifier from Eq. 4-14 and Eq. 4-17. For the transconductances, in the limiting case of 60° phase margin, holds

$$g_{m1} + g_{m2} = \omega_t (2c_l + C_m) \tag{4-35}$$

Section 4.3.1 demonstrated that the optimum Miller capacitor C_m is equal to the value of the interstage capacitance c_{p2}. In that situation the unity-gain frequency that follows from Eq. 4-14 reduces by a factor 2. Hence, we can write

$$g_{m1} + g_{m2} = 2\omega_t (2c_l + c_{p1}) \tag{4-36}$$

Following the same calculations as in the previous section, we can determine the bandwidth-to-power ratio. For bipolar transistors the result is

$$\frac{\omega_t}{P_s} = \frac{1}{V_s V_T} \frac{1}{4c_l + 2c_{p1}} \tag{4-37}$$

and for MOS devices

$$\frac{\omega_t}{P_s} = \frac{1}{V_s V_M(J_d)} \frac{1}{4c_l + 2c_{p1}} \tag{4-38}$$

Comparing these expressions to Eq. 4-31 and Eq. 4-32, which are the expressions for parallel compensation, we can conclude two things. First of all, the attainable bandwidth-to-power product of Miller compensation is lower than that of parallel compensation. This follows from the harmonic means in the denominators of Eq. 4-31 and Eq. 4-32, which will

always be smaller than the sum in the Equations 4-37 and 4-38. The cause of the limited bandwidth potential of Miller compensation lies in the fact that the gain asymptote of the output stage alone decides the attainable unity-gain frequency. With parallel compensation, the bandwidth is set by the gain asymptote of both transistors together. In cases with a large load capacitor, this two-transistor asymptote will lie above that of the single output transistor. In those situations the dimensioning of Miller compensated circuit must be adjusted to cope with the relatively low secondary pole p_1'. The result is that the total Bode plot of the open-loop gain shifts down. At high frequencies the magnitude curve will therefore deviate from the gain asymptote.

The second conclusion that follows from the comparison of bandwidth-to-power relations of parallel and Miller compensation is that, although Miller compensation is much more robust against variations of the load capacitance, the *maximum bandwidth-to-power ratio* of parallel compensation is less dependent on the load capacitance than that of Miller compensation. This becomes apparent from the c_l entry under the square root in the power-to-bandwidth ratios of parallel compensation, Eq. 4-31 and Eq. 4-32, while the load capacitor c_l enters the Miller expressions Eq. 4-37 and 4-38 directly. The difference again relates to the fact that in contrast to Miller compensation both transistors of the parallel compensated circuit contribute to the asymptotic gain limit. The gain asymptote of a single transistor is inversely proportional to the load capacitance c_l, while the asymptote of a two-stage amplifier depends on the square root of this quantity. Notice that the circuit with parallel compensation must always be adapted to any change of the load impedance. The Miller compensated circuit can handle most load variations without the need for readjustments.

C. Current-mode Miller compensation

The bandwidth-to-power ratio of current-mode Miller compensation follows from Eq. 4-37 and Eq. 4-38 and equals

$$\frac{\omega_t}{P_s} = \frac{1}{V_s V_T} \frac{1}{2c_l + 4c_{p1}} \tag{4-39}$$

and

$$\frac{\omega_t}{P_s} = \frac{1}{V_s V_M(J_d)} \frac{1}{2c_l + 4c_{p1}} \tag{4-40}$$

for bipolar and a MOS transistors respectively. The expressions includes the loading of the input of the amplifier by the load capacitance c_l. This loading also determines the optimal value for the Miller capacitor. According to the reasoning of Section 4.3.1, C_m should be of the same order of magnitude as the load capacitor c_l for the best bandwidth-to-power ratio. The factors 2 and 4 in the denominators of Eq. 4-39 and Eq. 4-40 have exchanged places compared to the corresponding equations of the voltage-mode circuits. In many cases the interstage capacitance c_{p2} is smaller than the load capacitance c_l, however. Therefore, in the limiting case of an extremely large load capacitance, a current-mode circuit attains a factor 2 higher bandwidth than the voltage-mode equivalent.

4.4.2. Noise

In most cases the use of (local) feedback does not change the noise properties much of a circuit. We can therefore expect that the noise performance of Miller compensation does not give in much to that of parallel compensation. Not withstanding this fact, it appears that the total integrated voltage noise over the bandwidth of a Miller compensated circuit is independent of the transconductance of the input stage and is dictated by the value of the compensation capacitor alone. In the following, we will examine the noise behavior of the various compensation methods.

A voltage and a current noise source at the input commonly model the noise performance of a circuit. The value of these sources is such that the actual circuit without the sources has the same noise properties as an ideal, noiseless circuit with the two sources at the input. The two input sources condense the contributions of all the noise sources in the circuit. The gain of the amplifier stages in most cases ensures that we only have to account for the noise contributions of the input stage.

Fig. 4-11 shows the small-signal diagrams of a bipolar and a CMOS transistor with their major noise sources. The noise sources of the bipolar transistor are i_{nb} and i_{nc}. Of these, source i_{nc} represents the shot noise on the collector current I_c. Source i_{nb} is the shot noise component of the base current I_b. The CMOS device only has one relevant noise source i_{nd}, accounting for the thermal noise of the channel. Because of the

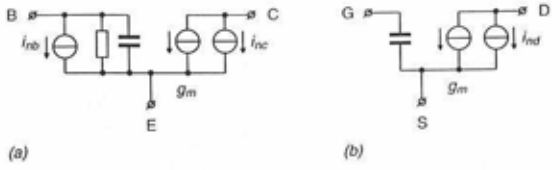


Fig. 4-11. *The small-signal diagrams of* (a) *a bipolar and* (b) *a CMOS transistor with the major noise sources.*

absence of a gate current, no appreciable current noise source is present at the gate. For simplicity we disregard the $1/f$ noise contributions, despite the fact that these can be significant in MOS circuits at low frequencies. Also, we neglect the noise contribution of the bulk resistance of the base, which is sometimes important in low-noise circuits. Notice that both the $1/f$ noise and the base resistance noise depend on the processing and are not fundamental. Theoretically, an advanced technology can make them arbitrarily small.

A. Bipolar

The collector noise i_{nc} and base noise i_{nb} obey

$$\overline{i_{nc}^2} = 2qI_c\Delta_f = 2kTg_m\Delta_f \tag{4-41}$$

$$\overline{i_{nb}^2} = 2qI_b\Delta_f = \frac{2kTg_m\Delta_f}{\beta_f} \tag{4-42}$$

where q is the charge of an electron, k is Boltzmann's constant, T is the temperature, Δ_f is the bandwidth of interest and β_f is the current gain.

The noise source i_{nb} already acts on the input of the amplifier and does not need any transformation. We can find the contribution of i_{nc} to the input noise by determining which input sources would result in the same current in the collector as the collector noise. The transconductance g_m

governs the relation between a voltage source at the input and the collector current. The equivalent input noise voltage of i_{nc} therefore becomes

$$\overline{v_{eq,c}^2} = \frac{2kT\Delta_f}{g_m} \tag{4-43}$$

Since a base current results in a collector current through the current gain β_f, the collector noise i_{nc} transforms into an equivalent noise current source at the input equal to

$$\overline{i_{eq,c}^2} = \frac{2kTg_m\Delta_f}{\beta_f^2} \tag{4-44}$$

The current gain of a practical transistor is always considerably greater than one. The result of Eq. 4-44 is therefore much less than the noise contribution of the base current of Eq. 4-42. Neglecting the $i_{ceq,c}$ term, the total equivalent input noise current becomes

$$\overline{i_{eq}^2} = \frac{2kTg_m\Delta_f}{\beta_f} \tag{4-45}$$

Equation 4-43 is the only contribution to the total equivalent input noise voltage v_{eq}, hence we obtain

$$\overline{v_{eq}^2} = \frac{2kT\Delta_f}{g_m} \tag{4-46}$$

At room temperature, kT amounts to approximately $4 \cdot 10^{-21}$ W/Hz.

B. CMOS

The expression for the drain current noise i_{nd} is

$$\overline{i_{nd}^2} = 4kTcg_m\Delta_f \tag{4-47}$$

In this expression c is a process dependent constant that, in practical cases, ranges between 1.3 and 2. We can transform the drain noise i_{nd} into an equivalent voltage noise source through the transconductance g_m. This is the only contribution to the total equivalent input noise voltage v_{eq}, which thus becomes

$$\overline{v_{eq}^2} = \frac{4kTc\Delta_f}{g_m} \tag{4-48}$$

Since the low-frequency current gain of a MOS transistor is near infinite, the drain current noise does not result in an equivalent current noise at the input and the total equivalent current noise is almost zero

$$\overline{i_{eq}^2} \approx 0 \tag{4-49}$$

We can use the results for the noise of a bipolar and a CMOS transistor to determine the noise performance of the parallel compensated amplifier of Fig. 4-5 and the Miller compensated amplifier of Fig. 4-7. In both cases only the input stages contribute appreciably to the equivalent input noise. Since the input transistors do not have any components around them affecting their noise properties, the general expressions Eq. 4-45, Eq. 4-46, Eq. 4-48 and Eq. 4-49 fully cover the equivalent noise sources at the input. As expected, there is no significant difference between parallel and Miller compensation in this respect.

An interesting expression for the noise of a Miller compensated amplifier follows from the observation that to maintain a constant unity-gain frequency ω_t, the transconductance g_{m2} of the input stage must be inversely proportional to the value of the Miller capacitor C_m. Since the total equivalent voltage noise at the input v_{eq} is directly proportional to the transconductance, this implies a relation between the noise of the circuit and the value of the Miller capacitor. This relation can be found by the following reasoning. To keep the frequency compensation optimal, increasing the Miller capacitor must be accompanied by a higher transconductance of the input stage. The increase of the transconductance in its turn reduces the input voltage noise. The relation between the noise and the value of the Miller capacitor for a bipolar circuit is

$$\overline{v_{eq}^2} = \frac{2kT\Delta_f}{C_m\omega_t} \tag{4-50}$$

while for a MOS amplifier holds

$$\overline{v_{eq}^2} = \frac{4kTc\Delta_f}{C_m\omega_t} \tag{4-51}$$

In both cases the total integrated noise voltage over the bandwidth ω_t of the amplifier is given solely by the value of the Miller capacitor. The transconductance of the input stage disappears from the expression. Setting $\omega_t = 2\pi\Delta_f$ in Eq. 4-50 and Eq 4-51 leads to

$$\overline{v_{eq}^2} = \frac{kT}{\pi C_m} \tag{4-52}$$

for the bipolar case and

$$\overline{v_{eq}^2} = \frac{2kTc}{\pi C_m} \tag{4-53}$$

for the MOS case.

Current-mode Miller compensation, shown in Fig 4-10, is hampered by the local feedback loop around the input stage. This feedback reduces the current gain and can therefore increase the contribution of noise sources further in the circuit to the equivalent input noise. Although these contributions may be small, one must pay careful attention to this aspect when designing current-mode Miller compensated amplifiers. We can find the equivalent input current noise of a bipolar current-mode Miller compensated amplifier as

$$\overline{i_{eq}^2} = \left(\frac{C_m}{c_{p2}}\right)^2 2kTg_{m2}\Delta_f \tag{4-54}$$

This value is in general significantly larger than the current noise of a single bipolar transistor, which was given in Eq 4-45. The reason is that the Miller capacitor brings back the current gain of the input stage from β_f to c_{p2}/C_m (for sufficiently high frequencies). Since the parasitic interstage

capacitance c_{p2} is generally smaller than the Miller capacitor, or at least of the same order of magnitude, the first term in the right-hand product of Eq. 4-54 will be larger than $1/\beta_f$ in Eq. 4-45.

The expression for the current noise in the MOS case of current-mode Miller compensation is

$$\overline{i_{eq}^2} = \left(\frac{C_m}{c_{p2}}\right)^2 4kTcg_{m2}\Delta_f \tag{4-55}$$

In Eq. 4-49 we found that the current noise at the input of a single MOS device is approximately zero. Therefore Eq. 4-55 clearly represents a deterioration of the noise characteristics.

The worsening of the noise performance that Eq. 4-54 and Eq. 4-55 express is a special case of the well-known rule that frequency compensation measures at the input of an amplifier increase the contribution to the equivalent input noise of noise sources further in the amplifier. This increase stems from the reduction of the input stage's gain that the frequency compensation causes.

4.4.3. Distortion

An important aspect of Miller compensation is that it employs local feedback to shape the frequency response into the desired proportions. The beneficial effect of feedback on the distortion suggests that the local feedback of Miller compensation reduces the non-linear distortion, as opposed to parallel compensation.

Assuming that all other components are ideal, the non-linearity of the active devices is what causes the distortion. By modelling the distortion components as separate signal sources, the small-signal diagram of Fig. 4-12 represents the two-stage amplifier of Fig. 4-3. The distortion sources are i_{d1} for the non-linear effects of T_1 and i_{d2} for T_2. The magnitude of these sources relative to the signal levels is proportional to the signal levels themselves. The output stage displays the greatest levels and therefore the contribution of the output transistor alone governs the total distortion of the circuit. In other words, the source i_{d1} almost exclusively generates the non-linear signal components at the output of the amplifier. Since the magnitude of this source is a function of the output current I_{out}, for the non-linear distortion d holds

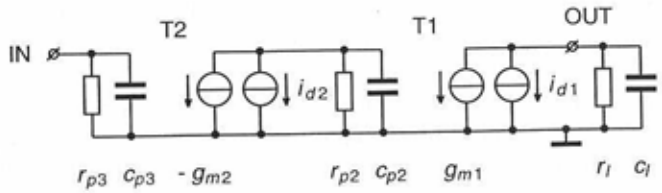


Fig. 4-12. *Small-signal diagram of a two-stage amplifier with the major sources of distortion.*

$$d = D(I_{out}) \tag{4-56}$$

The distortion function D depends on the exact nature of the non-linearity of the output device and on the type of distortion considered. Section 2.2.2 and Section 2.2.3 give the derivation of the functions D for the harmonic and the intermodulation distortion respectively. Depending on the output signal level, the distortion d typically ranges between 10% and 100%.

We found in Section 2.2.5 that the reduction due to the feedback of the distortion introduced by the output transistor equals the return difference F. The ratio of the circuit's determinants when the transconductance of the output transistor has its original value and when it is zero gives the return difference. Restating Eq. 2-14:

$$F = \frac{\Delta}{\Delta^0} \tag{4-57}$$

Since the output transistor T_1 is in the same loop as the Miller capacitor C_m, the return difference F of the transistor will be equal to the return difference of the capacitor. We can write the return difference as

$$F = \frac{\Delta}{\Delta^0} = \frac{\Delta}{\Delta_{ij}} \frac{\Delta_{ij}}{\Delta^0} \tag{4-58}$$

where i and j are arbitrary nodes in the circuit. In this expression, Δ_{ij} denotes the cofactor of the ith row and the jth column, first found in Section 2.1.1. It is equal to the determinant of the circuit's nodal equations matrix with the ith row and jth column removed, and multiplied by an appropriate plus or minus sign. If we choose i and j to be the input node and output node respectively, the ratio Δ_{ij}/Δ represents the transfer function T from the input to the output of the amplifier. At low frequencies, when the influence of the capacitor on the circuit is negligible, the cofactor Δ_{ij} will be equal to the cofactor without the capacitor, Δ_{ij}^{0}. Therefore at D.C. holds

$$F = \frac{\Delta}{\Delta^0} = \frac{\Delta}{\Delta_{ij}}\frac{\Delta_{ij}^0}{\Delta^0} = \frac{T_0}{T} \tag{4-59}$$

where T_0 is the transfer of the circuit without the Miller capacitor. When the transfer is an all-pole function, so that the cofactors merely represent gain constants, Eq. 4-59 will also be valid at frequencies above D.C. Hence, the gain reduces the distortion at each frequency by the same amount as the reduction of the gain. We can apply the results to the Bode plot of the amplifier. Because of the logarithmic magnitude unit, the ratio T_0/T in the expression is equal to the difference of the curves of the two transfers T_0 and T in the plot. In other words, the reduction of non-linear distortion of the Miller compensated circuit, equals the difference of the transfer with and without the frequency compensation in the Bode plot. In Fig. 4-13, the gray area indicates the difference. The figure shows that Miller compensation uses all internal gain to reduce the distortion of the output stage. The use of parallel compensation, on the other hand, reduces the gain in a particular frequency region, without enhancing the linearity of the amplifier. The distinction becomes sharper if we compare the distortion performance of both circuits in a closed-loop situation. Figure 4-14*a* shows the Bode plots of the parallel compensated amplifier without compensation T_0, with compensation T and with unity-gain feedback T_1. The overall feedback reduces the distortion of the circuit and the gray area now presents the distortion reduction of the total unity-gain amplifier. The difference of the magnitude lines of the compensated amplifier and the unity-gain circuit represent the reduction. In the situation of the Miller compen-

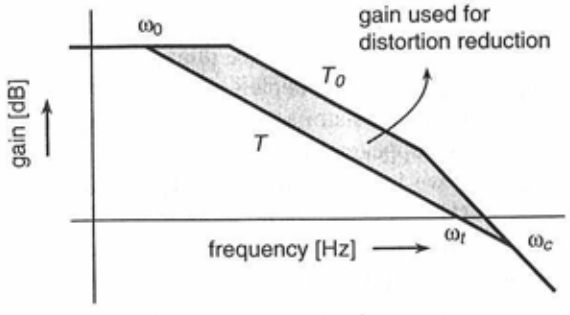


Fig. 4-13. *Bode plot indicating the distortion reduction due to the local feedback in a two-stage Miller compensated amplifier.*

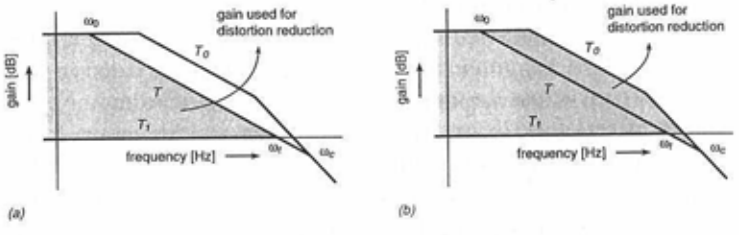


Fig. 4-14. *Unity-gain Bode plots showing the distortion reduction of* (a) *the parallel compensated amplifier and* (b) *the Miller compensated amplifier.*

sated amplifier in Fig. 4-14*b* the gray area is larger. It is bounded by the lines of the *uncompensated* circuit and the unity-gain plot. Clearly, all the gain present in the circuit contributes to the linearity.

The analysis of the distortion performance become more complicated when not only the non-linearity of the active components, but also that of the Miller capacitor is taken into account. The value of many inte-

grated capacitor types, especially those with a high specific capacitance, depends on the bias voltage between the plates. This is mainly caused by the narrowing and widening of a depletion layer between the two capacitor plates due to the electric field across it. Two types of capacitors that suffer from this form of non-linearity are junction and MOS capacitors. For the first type the presence of a depletion layer is obvious, since the depletion layer furnishes the insulator of the capacitor. For the second type —MOS capacitors— the appearance of a depletion layer is less clear. For a better understanding we have to look more closely into the constitution of the component. A MOS capacitor is based on the thin silicon oxide layer that is normally used for the gate isolation of a MOS transistor. The capacitor therefore consists of —top down— a polysilicon layer, serving as the first plate, a thin oxide layer as the insulator and a p or a n type region as the second plate. The material of the second plate corresponds to the respective channel implants for a NMOS or a PMOS transistor. It is in this 'channel' region that the voltage dependent depletion layer of a MOS capacitor occurs. Since the insulating oxide of the capacitor is thin, a small voltage across its terminals will already result in a high electric field at the surface of the second plate, just below the oxide itself. The doping of the channel region is relatively light, so that a broad, voltage dependent depletion layer arises which contributes to the value of the capacitor.

The non-linearity of the Miller capacitor directly affects the overall linearity of the amplifier, since the capacitor is located in the local feedback path across the output transistor. Although according to Section 2.2.5 the effect of the distortion signal components originating from the capacitor is reduced by the return difference F, the same holds for the base output signal. Since the ratio of the distortion signal components and the base signal determines the overall distortion figure, the distortion caused by the Miller capacitor is not reduced by the local feedback. For applications that require an extremely low distortion, it is therefore advisable to employ a voltage independent type for the Miller capacitor. The two interconnection layers of a double (or more) metal process offer this possibility. Furthermore, many I.C. processes have specific options for creating area efficient linear capacitors.

The changed position of the local feedback loop with the Miller capacitor C_m affects the distortion performance of current-mode Miller compensation as compared to its voltage-mode counterpart. Since the Miller capacitor is now connected across the input stage, the local feedback reduces the non-linearity of this circuit part, instead of the output

stage. The distortion of the input stage is generally low already, however, since the signals levels are small. The output stage, where most of the distortion occurs, experiences no beneficial effect of the local feedback. The current-mode Miller compensated amplifier will subsequently suffer more from non-linearities than the voltage-mode Miller circuit. In fact, not much improvement compared to parallel compensation will occur.

4.5. Conclusions

Two fundamentally different approaches for frequency compensation of electronic circuits exist. The first is parallel, the second Miller compensation. Parallel compensation is the traditional method for the frequency compensation of circuits built up with discrete components. The technique heavily relies on the use of inductors and large capacitors. Even though with parallel compensation the theoretical limit on the bandwidth can closely be approached, the dependency on non-integratable components limits its applicability in today's monolithically integrated circuits.

In an attempt to meet the specific needs of silicon I.C. technology, this chapter presented a special case of parallel compensation, using only capacitors and resistors. This approach demonstrated having a good bandwidth potential —although worse than the original technique with inductors. The large capacitor values needed for the compensation, however, renders parallel compensation less suited for use in integrated amplifiers. Other drawbacks are the sensitivity to parameter variations, the inevitable pole-zero doublet, the need for cascode stages and the poor linearity at high frequencies.

Although the bandwidth potential of Miller compensation is less than that of parallel compensation, the technique much better complies with the restrictions of I.C. circuit design. No inductors are required, while the Miller effect keeps the required capacitors small. Apart from this, a major advantage of Miller compensation is that the resulting circuits are inherently robust. Parameter variations of the process and changes in the load or the feedback have a limited effect on the properties of the amplifier. This is also reflected in the fact that pole-zero doublets do no occur, despite potential matching errors. The robustness is what makes Miller compensation a key player in the mass production of integrated amplifiers. Other advantages are that all the internal gain helps to reduce the non-lin-

ear distortion and the fact that local feedback with the Miller capacitor renders cascode stages obsolete.

5 Multistage Compensation Techniques

In many situations the gain of a two-stage amplifier is not sufficient and more stages are necessary. The traditional approach is to add current follower or voltage follower stages to the two-stage circuits of the previous chapter. These followers do not change the high frequency behavior appreciably, but can nevertheless improve the low-frequency gain. A serious drawback, however, is that the cascodes and emitter (source) followers increase the minimum supply voltage. Since low voltage is an essential step towards low power, a higher supply voltage will directly hamper the obtainable bandwidth-to-power ratio.

Figure 5-1 presents four common follower solutions derived from the two-stage Miller compensated opamp in Fig. 4-7. Figure 5-1*a* is the circuit that results when we introduce an emitter follower after the two-stage amplifier. The emitter follower decouples the intermediate stage Q_2 and the accompanying Miller capacitor C_m from the load. When the load consists of a large capacitance and a low resistance, this improves both the D.C. gain and the bandwidth of the amplifier. An example of this type of circuit is the renowned Fairchild μA741 operational amplifier [11]. The base-emitter voltage of the voltage follower Q_1 causes the impairment of the minimum supply voltage. The base-emitter voltage decreases the voltage swing at the output and, in order to compensate for this, we must raise the supply voltage by the same amount. Although Fig. 5-1*a* shows bipolar transistors, the circuit is equally suited for MOS devices.

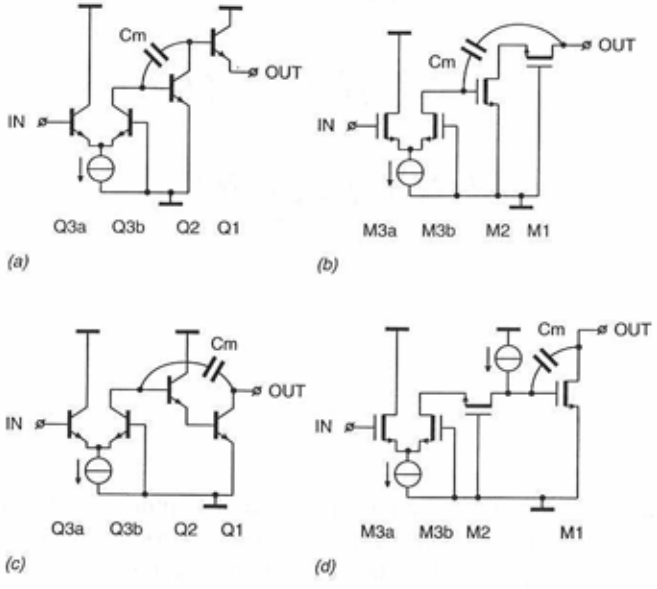


Fig. 5-1. *Conventional methods of inserting voltage and current followers into a two-stage circuit to increase the number of stages without increasing the frequency compensation complexity.*

Figure 5-1*b* is the result of inserting a cascode stage instead of an emitter follower after a two-stage Miller compensated amplifier. The circuit is shown using MOS transistors, since we can expect gain improvement only when the circuit is loaded by a large resistive load. The input stage of a subsequent MOS amplifier supplies such a load. The cascode reduces the output voltage swing by its drain-source voltage. This again demands a higher supply voltage than that of the two-stage circuit alone.

Introducing an emitter follower in between the input and the output stage of the two-stage amplifier leads to the Darlington topology of Fig. 5-1*c*. We can expect both gain and bandwidth improvement from this type of circuit. The increased interstage impedance to which the input stage supplies its current is responsible for the gain improvement. The

enlarged bandwidth results from the fact that the Miller capacitor C_m is no longer connected to the, relatively large, base-emitter capacitance of the output transistor Q_1. Therefore no voltage division occurs due to these capacitances and the secondary pole frequency p_1' approaches its theoretical maximum g_{m1}/c_l (c_l being the load capacitance). Clearly, these advantages go at the cost of a reduced voltage swing and consequently a higher minimum supply voltage. The circuit requires an additional base-emitter voltage to accommodate the emitter follower. The properties of the Darlington configuration are described in detail by the famous Solomon tutorial [11]. A serious drawback is the occurrence of the so-called 'output bump': a peaking effect at high output currents. An example of the use of the Darlington configuration is given in the discussion of two precision operational amplifiers with Nested Miller and Multipath Nested Miller compensation in Section 7.1. In Section 6.1.1 of the next chapter we will introduce two Darlington substitutes that subsequently evade the peaking associated with the traditional circuit and the increased minimum supply voltage.

As a final example of the conventional approaches to increase the number of stages of feedback amplifiers, Fig. 5-1*d* shows the outcome of inserting an interstage cascode into our two-stage amplifier. This method is extensively used in MOS amplifiers [12], since in this technology the high output impedance of the cascode and the purely capacitive gate-source impedance of the output stage yield the desired voltage gain. The circuit of Fig 5-1*d* includes a current source to explain the worsening of the minimum supply voltage. In order to maintain the voltage gain of the cascode, the impedance of the current source should at least be as high as the output impedance of the cascode Q_2. Otherwise, the current source will sink the current coming from Q_2, nullifying the effect of the cascode. In practice, this dictates implementing the current source with a cascode as well. The cascode current source, requiring head room of a saturation voltage more than a current source comprising a single transistor, causes the increase of the minimum supply voltage. Connecting the Miller capacitor of Fig. 5-1*d* to the source of the cascode transistor Q_2 instead of its drain can improve the obtainable bandwidth-to-power ratio. Section 5.3.2 presents the resulting circuit.

Avoiding the increase of the minimum supply voltage associated with the conventional circuits of Fig. 5-1 demands for alternative circuit solutions. An effective option is to use a cascade of gain stages (common emitter, common source) instead of the cascodes and the voltage followers

to realize sufficient gain. That such a chain of stages is not inhibited by limitations on the supply voltage can be seen from Fig. 5-2. The figure

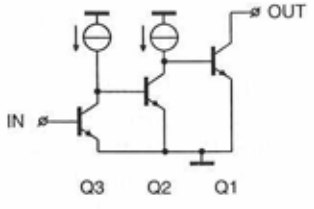


Fig. 5-2. *A basic amplifier embodying three gain stages without frequency compensation.*

shows an amplifier containing three gain stages without any measures for frequency compensation. Clearly, no restrictions apply for the supply voltage other than those of the equivalent two-stage circuit. The current sources in the figure comprise single transistors.

This chapter presents the multistage compensation techniques that allow to frequency compensate amplifiers with three gain stages and more. In accordance with the previous chapter, which gave the methods for two-stage compensation, we will introduce both multistage parallel compensation and multistage Miller compensation. The disadvantages of parallel compensation —large capacitor values, sensitivity to mismatch, dependence on the load impedance, the need for cascode stages and the weak distortion performance— become more prominent when the circuit comprises more than two stages.

Nesting Miller capacitor loops in an amplifier is a more effective method to compensate circuits containing three or more stages. The advantages of the Miller technique, most notably the robustness, distortion reduction and the small required capacitor values, remain valid. Many options are available for nesting Miller loops in a circuit. The result is a family of compensation methods for both bipolar and CMOS integrated amplifiers. The techniques covered in this chapter include Nested Miller compensation (Section 5.2), Reversed Nested Miller compensation

(Section 5.3) and, in Section 5.4, Hybrid Nested Miller compensation. Each of these techniques has its own specific field of application.

5.1. Multi-stage parallel compensation

Broadening the concept of parallel compensation to more than two stages requires additional capacitor-resistor pairs to shunt the interstage impedances. When we choose the values of the capacitors and the resistors correctly, the shunt impedances each introduce a zero to compensate for the pole at the output of the next transistor. A pole at a much higher frequency replaces the original pole. Further, the shunt impedance moves the pole of the interstage impedance at the location of the compensation network to a lower frequency. This pole-zero positioning is similar to the compensation of the two-stage amplifier in Section 4.2. For the purpose of analysis, Fig. 5-3 shows a three-stage parallel compensated circuit. Starting from

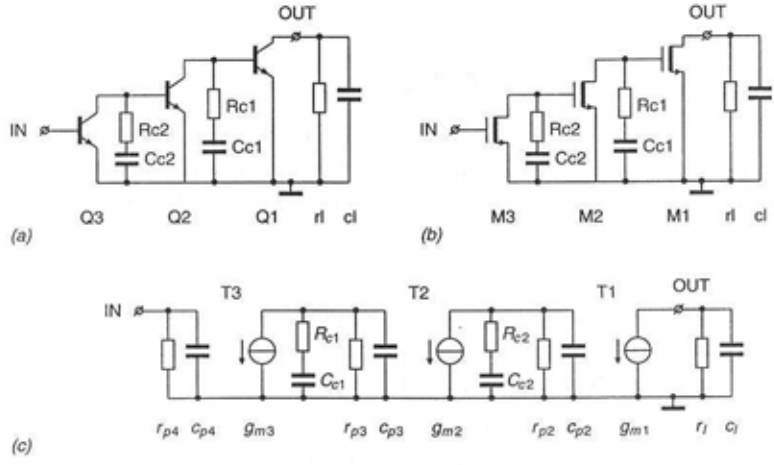


Fig. 5-3. *Three-stage amplifier with parallel compensation.*

the output stage backward, the zero of the compensation network C_{c1}/R_{c1}, must cancel the pole of the load impedance p_1. Hence

$$R_{c1}C_{c1} = r_1c_1 \tag{5-1}$$

The output pole is replaced by a pole at

$$p_1' = -\frac{1}{R_{c1}c_{p2}} = -\frac{1}{\omega_c} \tag{5-2}$$

This is a pole at a high frequency. It will eventually be one of the two secondary poles at ω_c that limit the attainable unity-gain frequency of the opamp. The compensation network further reduces the interstage pole p_2 to the new location p_2', which is

$$p_2' = -\frac{1}{r_{p2}C_{c1}} \tag{5-3}$$

Moving to the interstage shunt impedance C_{c2}/R_{c2}, we can repeat the same procedure. Only this time the newly placed pole p_2'' is the one that we must compensate by the zero of the compensation network. The relation that follows is

$$R_{c2}C_{c2} = r_{p2}C_{c1} \tag{5-4}$$

Again the shunt impedance introduces a high-frequency pole, this time at the position $p_2''' = -1/(R_{c2}c_{p3})$. Placing this pole on top of the secondary pole of the previous stage p_1', as shown in Eq. 5-3, requires

$$R_{c2}c_{p3} = r_{p2}C_{c1} \tag{5-5}$$

Finally, the compensation network C_{c2}/R_{c2} moves the corresponding interstage pole to a lower frequency p_3'. This pole will be the dominating pole frequency of the total amplifier. Its location is

$$p_3' = -\frac{1}{r_{p3}C_{c2}} \tag{5-6}$$

The Bode plot of Fig. 5-4. clarifies the positioning of the poles. Figure 5-4*a* shows the effect of the shunt impedance C_{c1}/R_{c1}. The light curve is the frequency response of the uncompensated amplifier. Since the circuit has three stages, we can distinguish three dominating poles. Inserting C_{c1} and R_{c1} moves the pole p_2 at the base (gate) of T_1 down to location

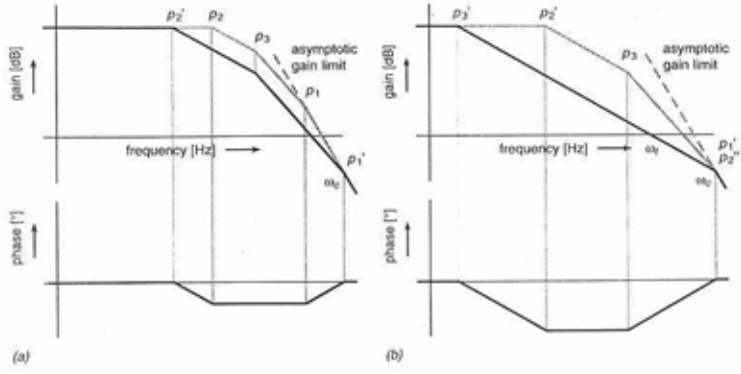


Fig. 5-4. *Bode plot of the pole-zero positioning of three-stage parallel compensation. Plot* (a) *shows the effect of* C_{c1}-R_{c1}, *plot* (b) *that of* C_{c2}-R_{c2}.

p_2'', while the pole p_1' at frequency ω_c replaces the output pole p_1. The latter pole is cancelled. The black line presents the resulting curve with two dominating poles p_2'' and p_3, and one secondary pole p_1'. Figure 5-4*b* shows the result of introducing C_{c2} and R_{c2}. This compensation impedance acts on the newly placed poles. It yields the desired 20dB/dec roll-off given by the black line in Fig. 5-4*b*. The compensation zero now cancels the pole p_2'', introducing the pole p_2''' at the secondary pole frequency ω_c. Furthermore, the shunt impedance moves down pole p_3 to the dominating pole frequency p_3'. The dashed line in both Bode plots represents the gain asymptote of the three transistor stages. We can repeat parallel strategy to compensate any number of gain stages.

The expressions for the dimensioning of the multi-stage parallel compensation follow from the observation that the unity-gain frequency ω_t of the total amplifier should be a factor k smaller than the position ω_c of the secondary poles p_1' and p_2'''. This is to ensure an acceptable phase margin. Therefore we have

$$\omega_c = k\omega_t \tag{5-7}$$

The bandwidth reduction factor k depends on the number of stages. Section 5.1.1 will elaborate on the value of the factor. The dimensioning now follows from Eq. 5-1, Eq. 5-2, Eq. 5-4, Eq. 5-5 and Eq. 5-7. Simultaneously solving these equations gives

$$\begin{aligned} R_{c1} &= \sqrt[3]{\frac{1}{k}\frac{c_l c_{p3}}{c_{p2}^2 g_{m1} g_{m2} g_{m3}}} \\ C_{c1} &= r_l \sqrt[3]{k \frac{c_l^2 c_{p2}^2 g_{m1} g_{m2} g_{m3}}{c_{p3}}} \end{aligned} \tag{5-8}$$

$$\begin{aligned} R_{c2} &= \sqrt[3]{\frac{1}{k}\frac{c_l c_{p2}}{c_{p3}^2 g_{m1} g_{m2} g_{m3}}} \\ C_{c2} &= r_l r_{p2} \sqrt[3]{k^2 c_l c_{p2} c_{p3} g_{m1}^2 g_{m2}^2 g_{m3}^2} \end{aligned} \tag{5-9}$$

The expression for the unity-gain frequency of the amplifier is

$$\omega_t = \sqrt[3]{\frac{1}{k^2}\frac{g_{m1} g_{m2} g_{m3}}{c_l c_{p2} c_{p3}}} \tag{5-10}$$

The unity-gain frequency of Eq. 5-10 is, apart from the bandwidth reduction due to k, equal to the unity-gain frequency of the gain asymptote. This suggests a bandwidth close to the theoretical limit, as was the case with two-stage parallel compensation in Section 4.2. Circuits with more than two stages cannot attain the optimum bandwidth anymore, due to the non-complex poles at ω_c. The poles cause a considerable phase shift that already impairs the phase margin at frequencies far below ω_c. Using inductors in the compensation networks to move the poles to complex positions paves the way to better results [1]. For obvious reasons, however, this peaking technique is no option for integrated amplifiers.

The complexness of Eq. 5-8, Eq. 5-9 and 5-10 leaves no doubt that the dimensioning of the multi-stage parallel compensation technique is even more dependent on the parameters of the components than the two-stage technique. In practice, it will be virtually impossible to predict the behavior of the circuit over different fabrication batches or over temperature. Another disadvantage is the value of the compensation capacitors. Especially the capacitor C_{c2}, or $C_{c(N-1)}$ in an N-stage amplifier, needs to be

large since this component together with the interstage resistance r_{p3} (r_{pN}) sets the dominating pole frequency of the circuit. An example is furnished by the following parameter values: $r_l = 10\text{k}\Omega$, $r_{p2} = 100\text{k}\Omega$, $c_l = 10\text{pF}$, $c_{p2} = c_{p3} = 2\text{pF}$, $g_{m1} = g_{m2} = g_{m3} = 4\text{mA/V}$ and $k = 4$. These values are similar to those used for the two-stage circuit in Section 4.2. The compensation capacitor C_{c2} that follows from Eq. 5-9 and the given parameters is 20nF, which is several orders of magnitude larger than the (already large) value found for the two-stage circuit. Increasing the number of stages of the parallel compensated circuit further enlarges the required compensation capacitance.

In the analysis of the parallel compensated amplifier the parasitic capacitances of the transistors between the collector and the base or the drain and the gate have been neglected. In many cases, however, these feedback capacitances will have a prominent effect on the transfer function of the circuit, due to the Miller effect. Therefore, cascode stages between the amplifier stages are necessary. Also, when the load impedance is high, we may have to insert a cascode into the output lead of the circuit. This addition has the disadvantage of increasing the power consumption and thus the bandwidth-to-power ratio. Further, the cascodes occupy scarce die area.

5.1.1. Bandwidth reduction of parallel compensation

The secondary poles at ω_c of the parallel compensated circuit ultimately limit the obtainable unity-gain frequency. The N - 1 poles —N being the number of stages— introduce a considerable phase lag at frequency below ω_c, which deteriorates the phase margin. The additional phase lag comes on top of the 90° phase shift of the dominating pole p_3'. For a phase margin of at least 60°, the phase lag at the unity-gain frequency ω_t of the secondary poles should therefore not exceed 30°. We made allowance for the phase margin in the analysis of the previous section through the introduction of the bandwidth reduction factor k. This factor relates the unity-gain frequency ω_t to the frequency of the secondary poles ω_c through $\omega_c = k\omega_t$, an expression which we found earlier as Eq. 5-7. At this point we would like to establish how much we must reduce the unity-gain frequency compared to the secondary pole frequency ω_c in order to obtain a sufficient phase margin. The phase shift at a certain frequency ω associated with a single pole generally follows from

$$\varphi = \operatorname{atan}\frac{\omega}{\omega_0} \tag{5-11}$$

where ω_0 is the frequency of the pole itself. We can summate the contributions of the N - 1 individual poles at ω_c to find the total phase shift at ω_t. Using the inverse function of Eq. 5-11, we can express the separation k of ω_c and ω_t as a function of the allowable phase lag

$$k = \frac{\omega_c}{\omega_t} \geq \tan\frac{\varphi}{N-1} \tag{5-12}$$

For a phase margin of 60°, a total phase lag φ of 30° is tolerable. If the number of stages of the amplifier N is three or more, which is the case under consideration, the argument of the tangent function is small. A first order approximation of the function suffices, and we obtain

$$k = \frac{\omega_c}{\omega_t} = \frac{6\,(N-1)}{\pi} \tag{5-13}$$

This is the desired result. Equation 5-13 clearly reveals that increasing the number of stages reduces the obtainable bandwidth. For example, in the three-stage amplifier of Fig. 5-3 the unity-gain frequency ω_t must be approximately four times lower than the position of the secondary poles at ω_c.

Equation 5-13 does not represent the minimum bandwidth reduction that one can obtain from an N-stage first order compensated amplifier. Better results are possible if we allow the secondary poles to be complex instead of being tied to the real axis of the complex s-plane. Moving the poles to complex positions is only possible with inductors as part of the parallel compensation impedances [1]. In the next sections we will see that local feedback techniques can create complex poles that yield an optimal open-loop frequency response, without the need for inductors.

5.2. Nested Miller compensation

We can modify the Miller compensation technique to compensate amplifiers with more than two stages. The philosophy behind this extension is different than that behind parallel compensation. Parallel compensation merely requires an additional transistor and a shunt impedance for every added stage. The structure is a repeating cascade of identical stages. Applying this procedure to Miller compensation is not possible. It would lead to a cascade of transistors with each having a Miller capacitor connected between the base (gate) and the collector (drain). The nature of the feedback around a Miller stage —a shunt-shunt connection— causes the input of the stage to accept currents, whereas the output signal is a voltage. Connecting two stages together therefore results in a signal mismatch that destroys the compensation properties of the Miller capacitors.

The best way to understand the expansion of Miller compensation to more stages is by inspection of the Bode plot shown in Fig. 5-5 of the

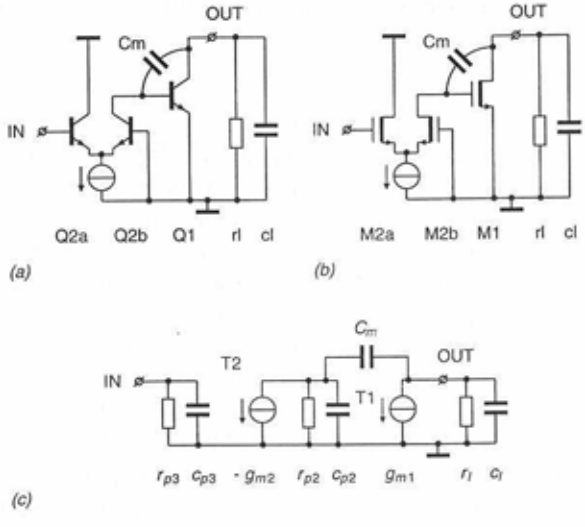


Fig. 5-5. *Two-stage Miller compensated amplifier.*

two-stage Miller compensated amplifier. The transfer function of the total circuit consists of one dominating pole p_2' and a second pole p_1' at a much higher frequency. Figure 5-6 shows this. Neglecting the non-dominant

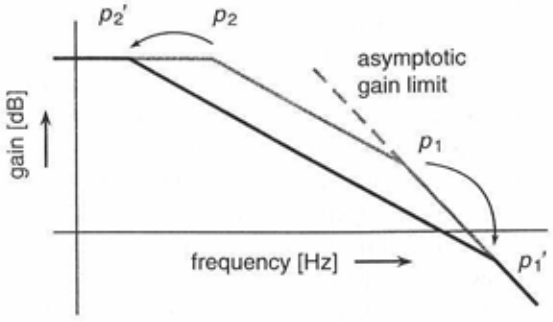


Fig. 5-6. *Bode plot of the two-stage Miller compensated amplifier.*

pole, the curve has a remarkable resemblance to the voltage-to-voltage transfer of a single transistor. The two-stage Miller compensated amplifier behaves as a single composite transistor with a capacitive load.

We can put this observation to practice by introducing the total two-stage circuit of Fig. 5-5 as the output transistor of an identical amplifier [13] [14]. Figure 5-7 shows the result. The circuit contains three gain stages and we can expect it to be eligible for compensation by choosing the proper values for the capacitors. The circuit of Fig. 5-7 contains two Miller capacitors, C_{m1} and C_{m2}, the feedback loop of C_{m1} enclosing the loop of C_{m1}. Because of this topology with local feedback loops at several levels, the structure is labelled Nested Miller compensation. We can easily extend the Nested Miller Compensation technique further, since the three-stage circuit again has a first order frequency response and we can insert it as the output stage of the two-stage circuit of Fig. 5-5. The result is a circuit with an additional Miller capacitor at a higher level for each added transistor stage.

The detailed pole positioning of the Nested Miller technique follows from inspection of the small-signal circuit diagram of Fig. 5-7*c*. The

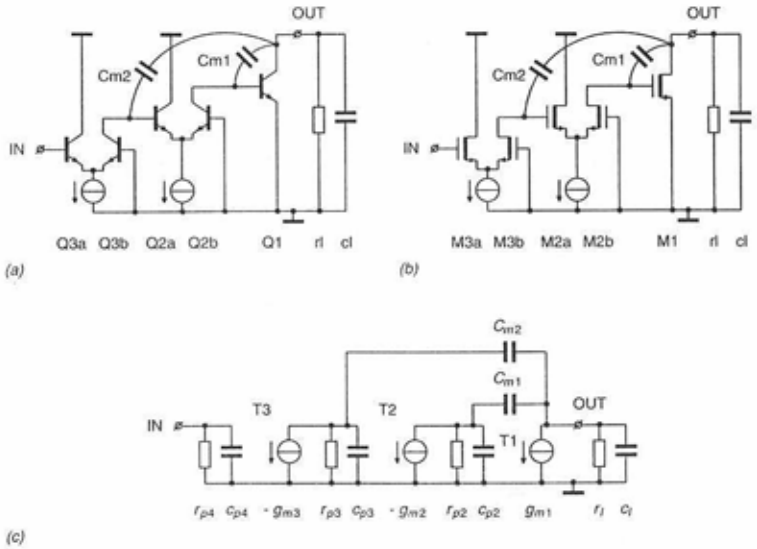


Fig. 5-7. *Three-stage Nested Miller compensation structure derived from two recursive two-stage circuits.*

Bode plot of Fig. 5-8*a* shows the effect of inserting C_{m1} in the circuit. The light line is the frequency response of the uncompensated amplifier, with the three dominating poles. The load impedance consisting of c_l and r_l provides the output pole p_1, c_{p2}/r_{p2} the pole p_2 and c_{p3}/r_{p3} the pole p_3. Inserting the Miller capacitor C_{m1} leads to the black curve. The Miller capacitor splits the poles p_1 and p_2 at both sides of the output transistor, moving them to the new positions p_1' and p_2'. This is the same effect as described in Section 4.3 for single Miller compensation. Miller capacitor C_{m2}, which closes the second loop, acts on the newly placed pole p_2' and the additional pole p_3, the latter being introduced by the input stage. Splitting these two poles results in a straight 20-dB/dec roll-off from the dominating pole frequency p_2'' up to the secondary poles at ω_c. Figure 5-8*b* presents the pole movements due to C_{m2}.

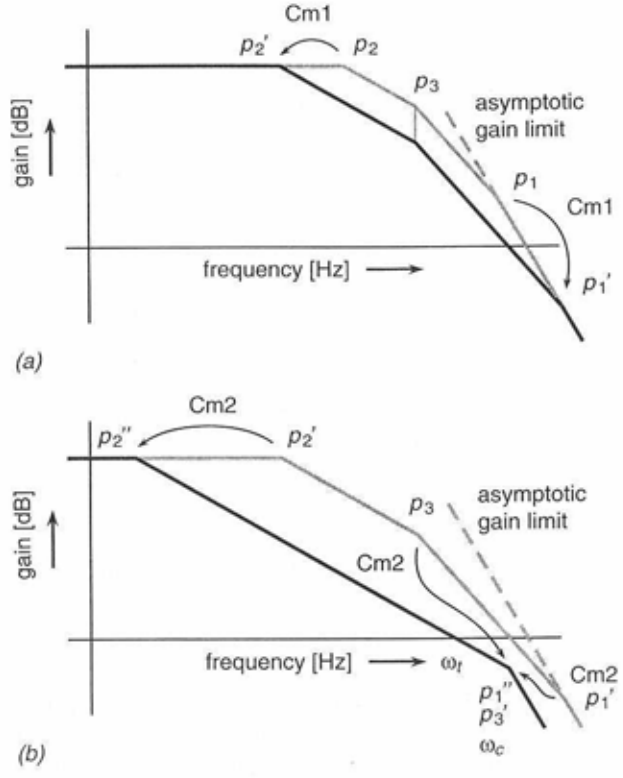


Fig. 5-8. *Bode plot showing the movement of the poles for Nested Miller compensation with* (a) *the effect of* C_{m1} *and* (b) *the effect of* C_{m2}.

To determine the values for the Miller capacitors, the best starting point is the transfer function of the amplifier with unity-gain feedback. Since under this condition the feedback loop returns the maximum signal from the output to the input, this is the worst-case situation for an absolutely stable amplifier. A well-accepted frequency response for general purpose amplifiers is the maximally flat or Butterworth function [6]. One

can find the pole positions of the amplifier by solving the nodal equations of the network. This method does not, however, give insight into the operation of the circuit. Since we will meet with other compensation methods later on that rely on the nesting of Miller feedback loops, it is a valuable investment to determine the pole locations by inspection of the circuit.

The amplifier with unity-gain feedback embodies three enclosing feedback loops. The Miller capacitor C_{m1} and the output transistor T_1 compose the loop at the lowest level. The loop consisting of the Miller capacitor C_{m2}, transistor T_2 and transistor T_1 is one level higher. The overall unity-gain feedback constitutes the outer feedback loop. We have analyzed the lowest level loop with C_{m1} and Q_1 extensively in Section 4.3. Together with the transconductance of the second stage T_2, it composes a voltage amplifier with the following transfer

$$\frac{v_{out}}{v_{in}} = \frac{K_1}{s\left(1 + s\frac{1}{p_1'}\right)} \tag{5-14}$$

where v_{in} is the voltage at the base (gate) of transistor T_2 and v_{out} is the voltage at the collector (drain) of T_1. The parameter K_1 in Eq. 5-14 is a gain constant. It equals the unity-gain frequency of the two-stage amplifier with T_1, T_2 and the Miller capacitor C_{m1}:

$$K_1 = \frac{g_{m2}}{C_{m1}} \tag{5-15}$$

The transfer of Eq. 5-14 represents an integrator, having its dominant pole p_2' in the origin and a secondary pole at p_1'. The transconductance of the output stage g_{m1} and the load capacitance c_l alone determine the secondary pole p_1'

$$p_1' = -\frac{g_{m1}}{c_l} \tag{5-16}$$

The pole in the origin suggests an infinite D.C. gain. It is an approximation that results from removing the resistances from the small-signal diagram of Fig. 5-7*c*. This assumption greatly simplifies the analysis. Assuming the

pole to lie in the origin does not affect the results, since the pole splitting secures that the actual position of the dominant pole p_2' is very low, far away from the high frequency poles we are interested in.

Closing the second loop of the amplifier by insertion of the outer Miller capacitor C_{m2} repositions the poles. We can find the new positions by considering the two-stage amplifier T_1 and T_2 inside the loop as the gain portion A of an elementary feedback amplifier. The Miller capacitor C_{m2} furnishes the feedback path β. For high frequencies this feedback path is a short circuit, since C_{m2} is generally much larger than the interstage capacitance c_{p3}. We can calculate the characteristic polynomial Δ, whose roots yield the desired poles, from $1 + A\beta$ as

$$\Delta(s) = \frac{1}{K_1 p_1'} s^2 + \frac{1}{K_1} s + 1 \tag{5-17}$$

The two poles that correspond to this second order polynomial are p_1'' and p_3', the high frequency poles of the black curve in the Bode plot of Fig. 5-8*b*.

The closing of the feedback loop with the Miller capacitor C_{m2} transforms the two-stages amplifier T_1 and T_2 into a current integrator. Adding the input stage T_3 results in an integrator with a voltage input, since the transconductance of this transistor converts the input voltage into a current. The unity-gain frequency of the integrator with T_3 and C_{m2} is

$$K_2 = \frac{g_{m3}}{C_{m2}} \tag{5-18}$$

The integrator displays three poles. The first is the dominant pole, approximately in the origin; the second and third follow from Eq. 5-17. Therefore the total open-loop transfer function is

$$\frac{v_{out}}{v_{in}} = \frac{K_1 K_2}{s\left(\frac{1}{p_1'} s^2 + s + K_1\right)} \tag{5-19}$$

The voltage v_{in} is now the voltage at the input of the amplifier, while v_{out} remains the voltage at the output. Notice that Eq. 5-19 gives the open-loop transfer of the total Nested Miller compensated amplifier. Therefore, K_2 denotes the unity-gain frequency ω_t of the open-loop response.

We can determine the closed-loop transfer function with unity-gain feedback using the open-loop gain of Eq. 5-19. Again denoting the open-loop gain as A and the feedback as β, we obtain

$$\frac{v_{out}}{v_{in}} = \frac{A}{1 + A\beta} = \frac{K_1 K_2 p_1'}{s^3 + p_1' s^2 + K_1 p_1' s + K_1 K_2 p_1'} \tag{5-20}$$

Since our aim is a maximally flat closed-loop frequency response, the denominator of Eq. 5-20 must equal a third order Butterworth polynomial. Letting the cut-off frequency be ω_0, the Butterworth polynomial is given by

$$B(s) = \left(\frac{s}{\omega_0}\right)^3 + 2\left(\frac{s}{\omega_0}\right)^2 + 2\frac{s}{\omega_0} + 1 \tag{5-21}$$

Equating Eq. 5-21 to the denominator of the right term from Eq. 5-20 yields the dimensioning data of the Nested Miller compensation structure

$$\begin{aligned} K_1 &= \frac{g_{m2}}{C_{m1}} = -\frac{1}{2} p_1' \\ K_2 &= \frac{g_{m3}}{C_{m2}} = -\frac{1}{4} p_1' \end{aligned} \tag{5-22}$$

The unity-gain frequency ω_t that corresponds to this solution is

$$\omega_t = K_2 = -\frac{1}{4} p_1' \tag{5-23}$$

Since single pole splitting results in a maximum unity-gain frequency of $p_1'/2$, Eq. 5-23 indicates a bandwidth reduction of a factor two compared to the basic two-stage circuit. This reduction also becomes clear from the Bode plot of Fig. 5-8. In part (*b*), showing the effect of the outer Miller capacitor C_{m2}, the black compensated amplifier curve steps back from the gain asymptote.

Finally, the cut-off frequency ω_0 of the Butterworth function follows from the calculations. This is the -3 dB point of the unity-gain closed-loop frequency response. It follows from

$$\omega_0 = -\frac{1}{2}p_1' \tag{5-24}$$

This result equals twice the unity-gain frequency of the open-loop frequency response. This confirms the well-known, but still surprising, fact that the bandwidth of a closed-loop amplifier can extend beyond the unity-gain frequency of the open-loop gain.

The root loci of Fig. 5-9 clarify the positioning of the poles of the Nested Miller compensation technique. Figure 5-9*a* shows the locus of the two stages T_1 and T_2 as a function of the Miller capacitor C_{m1}. Inserting the capacitor splits the poles into their new positions, marked by the squares. The zero in the origin of the root locus reflects the differentiating action that the Miller capacitor introduces in the feedback loop around T_1. The end positions of the root locus of Fig. 5-9*a* are the starting point of the root locus of the second Miller capacitor C_{m2}. Figure 5-9*b* shows this locus. Apart from the two split poles of the output and intermediate stage, the pole p_3, originating from the input stage T_3, appears in the loop gain of the outer Miller capacitor. The squares of plot *b* are at the open-loop pole positions of the total amplifier. The locus of Fig 5-10 presents the effect of the unity-gain feedback. The poles end up on the semi-sphere in the left half-plane that is characteristic to a Butterworth transfer function.

A good starting point for the design procedure of the Nested Miller compensation scheme is the value of the highest pole that results from connecting Miller capacitor C_{m1} across the output transistor T_1. Equation 5-16, which is the approximated form of the general expression in Eq. 4-13, gives this pole p_1'. From pole p_1', using Eq. 5-22, we can calculate the unity-gain frequency K_1 of the internal Miller feedback loop with C_{m2} and the unity-gain frequency K_2 of the total open-loop amplifier. We can then use the parameters K_1 and K_2, through Eq. 5-22, to determine the interstage transconductance g_{m2} and the outer Miller capacitor C_{m2}. In Section 4.3.1 we found that the Miller capacitor C_{m1}, connected across the output transistor, should be of the same order of magnitude as the intermediate capacitance c_{p2}. Complying with this rule, we can deduce the transconductance g_{m2} from K_1. The transconductance of the input stage T_1 and the outer Miller capacitor C_{m2} set the open-loop unity-gain frequency K_2. Often noise considerations dictate the transconductance of T_1. In these cases the value of the Miller capacitor C_{m2} unambiguously follows from K_2 in Eq. 5-22.

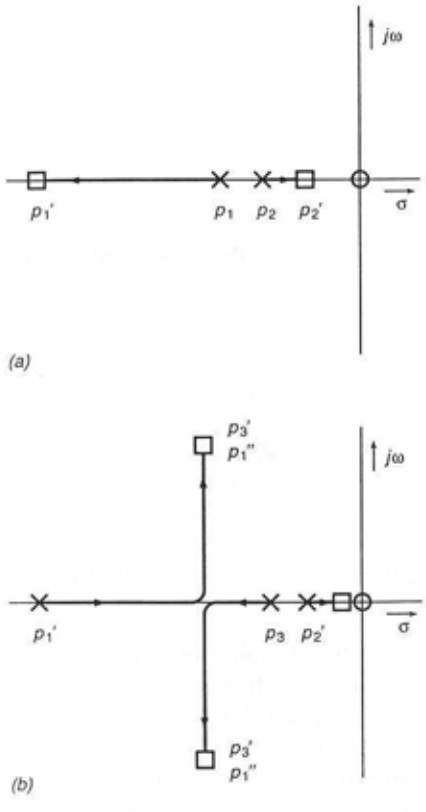


Fig. 5-9. *Root loci of the Nested Miller compensation structure with* (a) *the effect of* C_{m1} *and* (b) *the effect of* C_{m2}.

Consider the following example: g_{m1} = 10mA/V, g_{m3} = 10mA/V, $c_{p1} = c_{p2} = c_{p3}$ = 2pF, c_l = 10pF, C_{m1} = 5pF. Equation 5-16 returns $p_1' = -1 \cdot 10^9\ s^{-1}$ (-159MHz). Therefore $K_1 = 5 \cdot 10^8 s^{-1}$ (80MHz) and $K_2 = 2.5 \cdot 10^8 s^{-1}$ (40MHz), the latter also being the unity-gain frequency of the amplifier. Equation 5-22 now yields g_{m2} = 1.25mA/V and C_{m2} = 8pF.

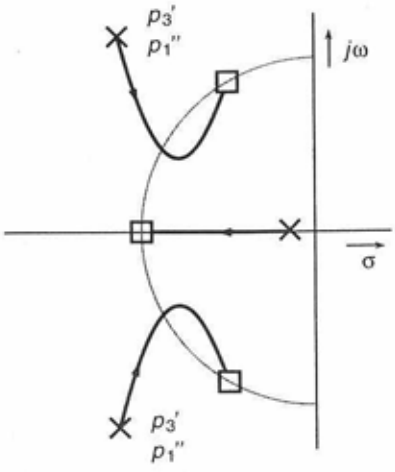


Fig. 5-10. *Root locus showing the result of closing the unity-gain feedback loop around the total amplifier.*

It is often advisable to substitute into Eq. 5-22 and Eq. 5-23 a lower value for p_1' than follows from Eq. 5-16 or, more precise, Eq. 4-13. Both expressions for p_1' are approximations that do not take into account, for instance, the phase lag of the higher order poles. A practical choice would be to use half the frequency of the pole p_1' in the calculations. In most cases this approach will lead to a well balanced circuit with a bandwidth close to the maximum attainable.

Fine tuning of the circuit parameters can be accomplished on the basis of the respective phase margins φ_1 and φ_2 that are associated with the unity-gain factors K_1 and K_2. We can examine the phase margin φ_1 of the internal Miller loop around the transistors T_1, T_2 and the Miller capacitor C_{m2} by opening that loop, for example between the Miller capacitor C_{m2} and the output terminal. The amplifier must have no overall feedback for this purpose. In cases with an exceptionally low load capacitance, the loading of the amplifier due to the outer Miller loop can be accounted for by adding the interstage capacitance c_{p3} to the total load capacitance. Calculating the internal phase margin for an optimally dimensioned circuit

requires examining the phase shift that the poles p_1' and p_2' in Fig. 5-9*a* cause at the internal unity-gain frequency K_1. The result is

$$\varphi_1 = 63° \tag{5-25}$$

The phase margin φ_2 that is associated with K_2 represents the phase margin of the total amplifier. Calculating the phase shift of the poles p_1'', p_2'' and p_3' in Fig. 5-9*b* at the unity-gain frequency K_2 leads to this figure. The calculations reveal that the overall phase margin of the Nested Miller compensated amplifier should equal

$$\varphi_2 = 60° \tag{5-26}$$

5.2.1. Dimensioning of NMC with more than three stages

We can apply the analysis of the previous section to Nested Miller compensated amplifiers with more than three stages. For each number of stages we would like to know the unity-gain factors K of all the internal loops and the unity-gain frequency of the amplifier itself. The general expression for the unity-gain factors is

$$K_{(k)} = \frac{g_{m(k+1)}}{C_{m(k)}} \tag{5-27}$$

where the index k ranges between one and the number of stages decrement by one N - 1. Figure 5-11 shows the corresponding structure. With straight forward calculations we can determine the values of the unity-gain factors. Table 5-1 collects the results for $N = 1$ to $N = 5$. The values are normalized with respect to the highest pole p_1'. Apart from the normalized gain factors $k_{(k)}$, the table also summarizes the normalized unity-gain frequency of the total amplifier ω_t, which equals k_{N-1}, and the normalized -3dB bandwidth ω_0 of the amplifier with unity-gain feedback. Setting the circuit parameters to the values in Table 5-1 results in a unity-gain amplifier with a maximally flat (Butterworth) amplitude response. The single-stage row is of course a special case, since one transistor does not require frequency compensation. The ratio of the transconductance of the transistor and the load capacitance limit the unity-gain frequency and the closed-loop bandwidth, however. This ratio exactly equal pole p_1', which explains the "1" entries of the table.

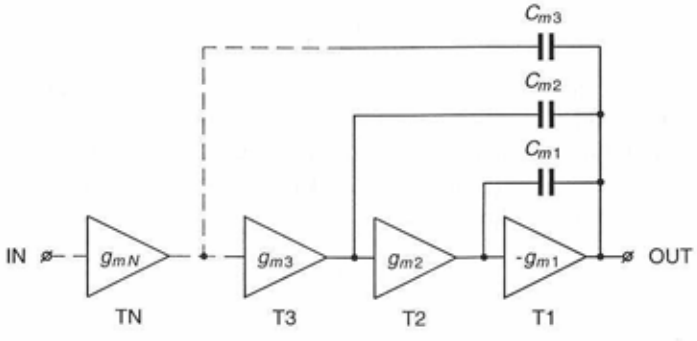


Fig. 5-11. *Nested Miller compensation structure for* N *stages.*

Table 5-1. *Normalized unity-gain factors of the NMC structure.*

# stages	k_1	k_2	k_3	k_4	ω_t	ω_0
1					1	1
2	0.5				0.5	0.71
3	0.5	0.25			0.25	0.5
4	0.5	0.29	0.15		0.15	0.38
5	0.5	0.31	0.19	0.1	0.1	0.31

Notice that in all cases the bandwidth of the closed-loop amplifier ω_0 is substantially higher than the unity-gain frequency of the open loop. The controlled peaking that the internal feedback loops introduces causes this bandwidth improvement. This is an advantage over parallel compensation, which would have to rely on inductors to introduce complex poles into the transfer function. Another advantage is that the Nested Miller Compensated circuits do not require cascodes to counteract the parasitic feedback capacitances between the collector (drain) and the base (gate) of the transistor stages. The Miller capacitors, being the feedback elements of

elementary integrators, bring down the impedances at the interstage nodes of the amplifier, thus eliminating the effect of the parasitic capacitances.

5.2.2. Current-mode Nested Miller compensation

In much the same way as with the voltage-mode Miller compensation, we can modify the current-mode Miller compensation structure to realize a Nested Miller compensated feedback amplifier with a current input and output [13]. The procedure requires recursively inserting the two-stage current-mode Miller compensated amplifier of Fig. 5-5 as the input transistor T_2 of the same circuit. Figure 5-12*a* presents the result. The figure

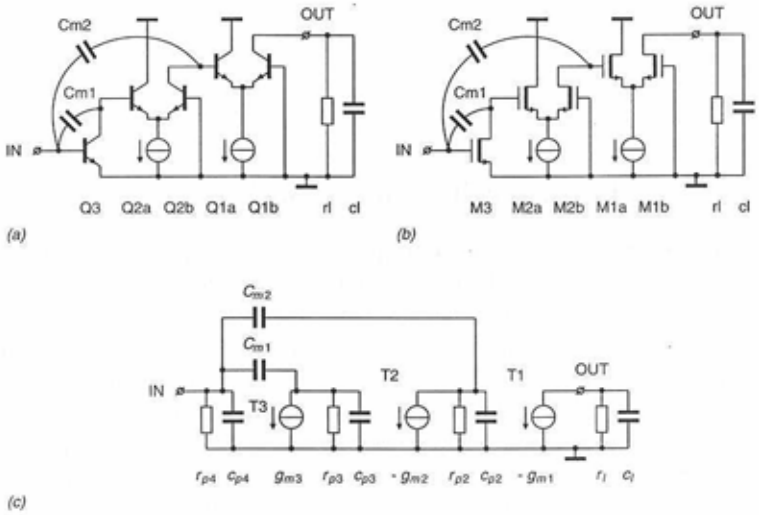


Fig. 5-12. *Three-stage current-mode Nested Miller compensated amplifier circuit.*

shows a three-stage current-mode Nested Miller compensated amplifier. The topological difference with the voltage-mode circuit of Fig. 5-7, is that the common node of the Miller capacitors now lies at the input instead of the output. This does not change the operation of the amplifier much, since through the unity-gain feedback the load capacitance is again con-

nected to this node. In fact, we can derive the dimensioning equations of the current-mode Nested Miller compensation structure from those of its voltage-mode counterpart by reversing the order of the transistors in Eq. 5-22. The resulting expressions are

$$K_1 = \frac{g_{m2}}{C_{m1}} = -\frac{1}{2}p_1' \qquad K_2 = \frac{g_{m1}}{C_{m2}} = -\frac{1}{4}p_1' \tag{5-28}$$

The position of the pole p_1' follows from Eq. 5-16. Reversing the transistor order in this equation results in

$$p_1' = -\frac{g_{m3}}{c_l} \tag{5-29}$$

5.3. Reversed Nested Miller compensation

Moving the output transistor of the current-mode Nested Miller compensation circuit of Fig. 5-12 to the input, we obtain the so-called Reversed Nested Miller compensation topology [13]. Figure 5-13 presents the structure. The Reversed Nested Miller technique has a considerable bandwidth advantage over Nested Miller compensation. This advantage is basically caused by the fact that the most inner feedback loop with the transistor T_2 and the Miller capacitor C_{m1} has no connection to the —relatively large— load capacitance. It is this lowest level loop that ultimately determines the bandwidth of the amplifier. To investigate the bandwidth, consider Fig. 5-14. The figure shows the transistors Q_2 and M_2, the inner Miller capacitor C_{m1} and the outer Miller capacitor C_{m2}. Further, the figure presents the same circuit with an idealized transconductor T_2 instead of the transistor.

The circuit of Fig. 5-14 behaves as an inverting voltage amplifier. The feedback around T_2 turns its base (gate) into a virtual ground. The capacitor C_{m2} converts the voltage at its input terminal into a current, which the basic integrator T_2/C_{m1} then reconverts into a voltage. The ratio of the impedances C_{m2}/C_{m1} sets the voltage gain. The voltage amplifier

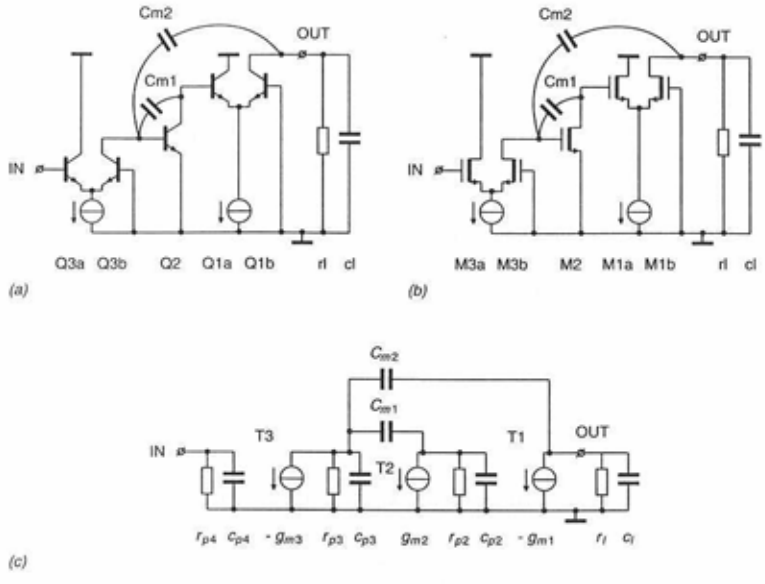


Fig. 5-13. *The Reversed Nested Miller compensation topology.*

multiplies the gain of the internal feedback loop through C_{m2}, T_2, and T_1 of the Reversed Nested Miller circuit of Fig. 5-13 by this ratio. By denoting the inevitable pole of the transfer of the voltage amplifier in Fig. 5-14 as p_1', we can reuse the analysis of the voltage-mode Nested Miller compensation of Section 5.2. The unity-gain frequency K_1 of the internal feedback loop that is closed by C_{m2} equals

$$K_1 = \frac{C_{m2}}{C_{m1}} \frac{g_{m1}}{c_l} \tag{5-30}$$

The first term in this equation represents the gain of the inverting voltage amplifier of Fig. 5-14. The second term is the unity-gain frequency of the intermediate transistor T_1 loaded by the load capacitance c_l. The gain of the inverting amplifier around T_2 lifts the first order gain magnitude plot of

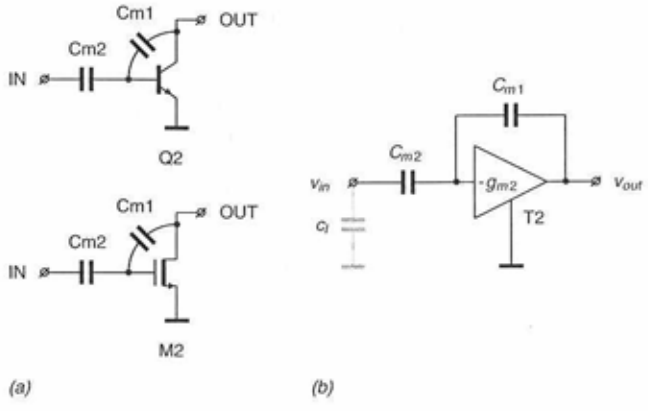


Fig. 5-14. *Basic voltage amplifier around transistor* T_2.

transistor T_1, thus proportionally increasing the unity-gain frequency K_1 of the Miller loop. The unity-gain frequency ω_t of the total amplifier equals K_2 and obeys

$$\omega_t = K_2 = \frac{g_{m3}}{C_{m2}} \tag{5-31}$$

According to Table 5-1, the unity-gain factor K_1 should equal $0.5p_1'$ and $K_2 = 0.25p_1'$ to obtain a maximally flat frequency response of the closed-loop amplifier.

The remaining question is the value of p_1'. The pole position follows from inspection of the simplified circuit from Fig. 5-14. Assuming that the load capacitance c_l of the total amplifier is relatively large, the combination C_{m1}/C_{m2} acts as a capacitive voltage divider that degenerates the local feedback loop through T_2 and C_{m1}. Further, the Miller compensated transistor T_2 is loaded by the interstage capacitance c_{p2}. Therefore the value of p_1' is

$$p_1' = \frac{C_{m1}}{C_{m2}} \frac{g_{m2}}{c_{p2}} \tag{5-32}$$

For this expression we assume that the Miller capacitor C_{m1} is considerably smaller than the interstage capacitance c_{p2}. This condition prevents loading of the circuit by the compensation capacitor. The required small value of the Miller capacitor C_{m1} is not in all cases convenient, although. Resistors in series with the capacitors may be necessary to reduce the loading by the capacitors. The explanation of the resistors follows later on in this section. Substituting Eq. 5-30, Eq. 5-31 and Eq. 5-32 into the expressions for K_1 and K_2 as a function of p_1' of Table 5-1 leads to the following dimensioning equations

$$\frac{C_{m2}^2}{C_{m1}^2} = \frac{c_l g_{m2}}{2 g_{m1} c_{p2}} \tag{5-33}$$

and

$$\omega_t = \sqrt{\frac{g_{m1} g_{m2}}{8 c_{p2} c_l}} \tag{5-34}$$

From Eq. 5-34 it follows that the asymptotic gain limit of the output stage T_1 and the intermediated stage T_2 together set the unity-gain frequency. Transistors T_1 and T_2 are exactly the devices that the outer Miller capacitor C_{m2} encloses. The result is better than that obtained from Nested Miller compensation, where the gain asymptote of the output transistor alone limits the unity-gain frequency. When comparing Nested Miller compensated circuits and Reverse Nested Miller compensated circuits of more than three stages, the difference becomes even more profound. The unity-gain frequency of the Nested Miller compensated circuit reduces below the gain asymptote of the output stage with every added stage, while Reversed Nested Miller compensation approaches the total gain asymptote of the N - 1 stages enclosed by the outer Miller capacitor.

Figure 5-15 draws the general Reversed Nested Miller compensated circuit with N stages. We can identify the unity-gain factors of the circuit as

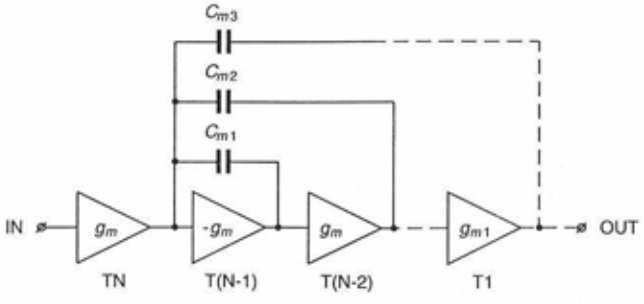


Fig. 5-15. *Reversed Nested Miller compensation extended to* N *stages.*

$$K_{(k)} = \frac{C_{m(k+1)} g_{m(N-k)}}{C_{m(k)} c_{p(N-k)}} \tag{5-35}$$

where k ranges between 1 and N - 2. We assume that c_{p1} denotes the load capacitance in Eq. 5-35. The expression represents the general case of Eq. 5-30. The overall unity-gain frequency of the Reversed Nested Miller circuit is given by

$$\omega_t = K_{(N-1)} = \frac{g_{m(N)}}{C_{m(N-1)}} \tag{5-36}$$

which corresponds to Eq. 5-31 of the three-stage circuit. The generalized form of Eq. 5-32 gives the pole p_1':

$$p_1' = \frac{g_{m(N-1)}}{c_{p(N-1)}} \frac{C_{m1}}{C_{m(N-1)}} \tag{5-37}$$

We can calculate the dimensioning expressions of the Reversed Nested Miller compensation technique by simultaneously setting the unity-gain factors of Eq. 5-35 equal to the corresponding values of Table 5-1. This leads to the following set of equations

$$K_{(k)} = k_{(k),N} \cdot p_1' \tag{5-38}$$

where $k_{(k),N}$ is the value of the normalized table entry.

The maximum unity-gain frequency ω_t of a Reversed Nested Miller compensated circuit containing N stages is expressed by

$$\omega_t = \sqrt[N-1]{\frac{k_{(N-1)}^{N-1}}{\prod_{i=1}^{N-2} k_{(i)}} \prod_{i=1}^{N-1} \frac{g_{m(i)}}{c_{p(i)}}} \tag{5-39}$$

where c_{p1} again represents the load capacitance c_l. Equation 5-40 reveals the structure of Eq. 5-39 by giving the example for $N = 4$

$$\omega_t = \sqrt[3]{\frac{k_3^3}{k_1 k_2} \frac{g_{m1} g_{m2} g_{m3}}{c_{p1} c_{p2} c_{p3}}} = 0.29 \sqrt[3]{\frac{g_{m1} g_{m2} g_{m3}}{c_l c_{p2} c_{p3}}} \tag{5-40}$$

The gain asymptote of the last N - 1 stages clearly limits the unity-gain frequency.

A drawback of the Reversed Nested Miller compensation technique is its dependency on the load capacitance. This dependency is inevitable for any compensation technique that approaches the theoretically maximum bandwidth, since the theoretical optimum itself depends on the load capacitance. We can also understand this from the structure itself. Figure 5-15 shows that the internal feedback is concentrated around the second stage of the amplifier. Only a small amount is available around the output stage to improve the robustness against parameter variations, including variations of the load capacitance.

5.3.1. Eliminating the loading caused by the Miller capacitors

In the previous section we found that for the best results the Miller capacitors of the Reversed Nested Miller compensation technique must be smaller than the respective interstage capacitances. In practice this condition can be difficult to fulfill, since the interstage capacitances are the result of parasitic effects and are therefore supposed to be small them-

selves. Also noise requirements can pose demands on the value of the outer Miller capacitor. Low noise generally requires a high transconductance of the input stage, which implies a large outer Miller capacitor.

In the cases where the Miller capacitors are relatively large, we can eliminate the loading of the circuit by inserting resistors in series with the capacitors. This is shown by Fig. 5-16. Since only the ratios of the imped-

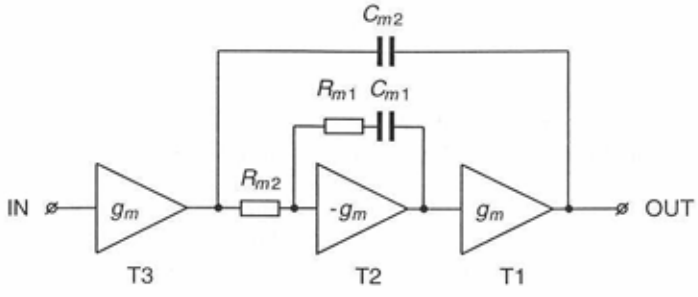


Fig. 5-16. *Reversed Nested Miller compensation with series resistors to reduce the loading of the circuit by the Miller capacitors.*

ances in the capacitor branches enter Eq. 5-33 of the Reversed Nested Miller compensation technique, the basic behavior of the circuit does not change when the series resistors have the same mutual ratio as the impedances of the Miller capacitors. In other words

$$\frac{R_{m1}}{R_{m2}} = \frac{C_{m2}}{C_{m1}} \tag{5-41}$$

The input stage of Fig. 5-16 supplies its output current to the common node of the Miller capacitor C_{m2} and the resistor R_{m2}. This assures that the overall frequency response remains dominated by the integrating action of the outer Miller capacitor C_{m2}. This capacitor is located in the feedback path of the circuit.

If the circuit does not meet Eq. 5-41 exactly, the presence of the resistors results in pole-zero doublets. To avoid slow-settling components

due to these doublets, their frequency must by higher than the unity-gain frequency of the total amplifier. The position of the pole-zero doublets follows from $(C_{m1}R_{m1})^{-1} = (C_{m2}R_{m2})^{-1}$. The condition to prevent slow-settling components therefore becomes

$$\frac{1}{C_{m1}R_{m1}} = \frac{1}{C_{m2}R_{m2}} > \omega_t \tag{5-42}$$

It may be necessary to provide an efficient path for high frequencies around the compensation resistors, especially when these are large. We can realize the path by insertion of small capacitors in parallel to the combination of the Miller capacitors and the associated resistors. The components C_{m1}' and C_{m2}' introduce these capacitors in Fig. 5-17. The ratio of the

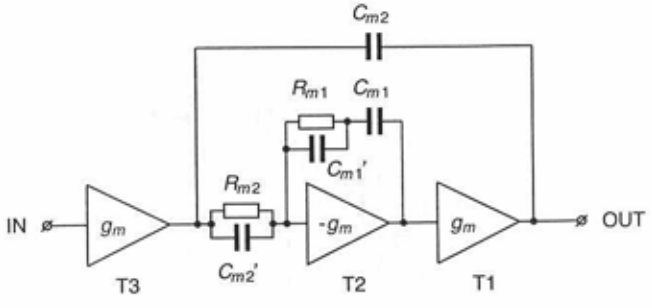


Fig. 5-17. *Reversed Nested Miller compensation with additional capacitors to bypass the decoupling resistors for high frequencies.*

capacitors must still obey Eq. 5-41. Nodal analysis unveils that there is an optimum value for the product of the bypass capacitors, which is given by

$$C_{m1}'C_{m2}' = c_{p2}c_{p3} \tag{5-43}$$

When Eq. 5-41 and Eq. 5-43 are combined, they yield the absolute value of the capacitors. The values are generally small, in the order of magnitude of the parasitic interstage capacitances. Therefore, the occupied die area is

negligible. When the boundary conditions of the amplifier do not require large compensation capacitors, we can remove the series connected resistors and capacitors from the circuit of Fig. 5-17, leaving only the bypass capacitors C_{m1}' and C_{m2}'.

5.3.2. Cascode based Reversed Nested Miller compensation

The operation of the Reversed Nested Miller compensation structure of the previous sections bases itself on the straight voltage-to-voltage gain of the transistor M_2 and the surrounding Miller capacitors C_{m1} and C_{m2} from Fig. 5-14. Any solution that provides an identical gain function can substitute this circuit part, without altering the characteristics of the amplifier, nor the dimensioning equations. An important example of such a modification is the circuit that we obtain when we replace transistor M_2 by a cascode device. Figure 5-18 shows the resulting circuit. The circuit is

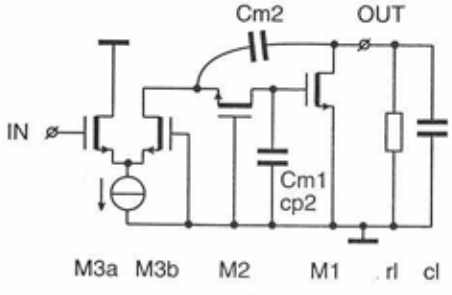


Fig. 5-18. *Cascode based Reversed Nested Miller compensation with a shunted compensation capacitor.*

depicted using MOS devices, since in this technology cascode topologies predominate. This choice imposes no restriction on the generality of the circuit, however. Bipolar transistors are equally applicable. The common-gate transistor M_2 represents the cascode. Its function is to increase the output impedance of the input transistor M_3, thus enhancing the gain of the total amplifier.

In order to frequency compensate the cascode circuit, the compensation capacitor C_{m1} has moved away from its original Miller position in Fig. 5-13*b*. It is now shunt connected between the drain of M_2 and the ground. Surprisingly, the behavior of the transistor M_2 and the capacitors C_{m1} and C_{m2} is very similar in the two situations of Fig. 5-13*b* and Fig. 5-18. In both cases these components constitute a voltage amplifier with a gain of C_{m2}/C_{m1}. For simplicity we assume that C_{m1} in the circuit of Fig. 5-18, apart from the inserted compensation capacitor, also accounts for the already present interstage capacitance c_{p2}.

We can understand the voltage gain of the cascode based circuit in Fig. 5-18 from the following reasoning. The source of M_2 presents a virtual ground to the left-hand terminal of capacitor C_{m2}. Therefore, the impedance of the capacitor converts an input voltage at the other terminal into a current through the capacitor. The current flows into the source of M_2, reappearing at its drain. At that point it is forced into the capacitor C_{m1}, which converts the current into a voltage again. The net voltage gain between the output terminal of the amplifier and the gate of the output transistor M_1 is

$$A_v = \frac{sC_{m2}}{sC_{m1}} = \frac{C_{m2}}{C_{m1}} \tag{5-44}$$

which is identical to the relation obtained for the Reversed Nested Miller compensation scheme of Fig. 5-14. In many practical situations, the contribution of the gate-source capacitance of the output transistor M_1 to the total compensation capacitance C_{m1} suffices for compensation. This eliminates the need for an additional compensation capacitor. The voltage gain is in this case given by the ratio C_{m2}/c_{p2}.

The pole that is caused by the Miller capacitor C_{m2} and the non-zero input impedance of the cascode M_2 ultimately limit the unity-gain frequency of the cascode based Reversed Nested Miller compensation circuit. The pole occurs inside the second Miller loop with capacitor C_{m2} and therefore takes the place of the bandwidth limiting pole p_1' of ordinary Reversed Nested Miller compensation. Up to the corner frequency of the cascode's input impedance and the Miller capacitor C_{m1}, the capacitor governs the relation between the voltage at the output terminal and the current flowing into the cascode. Above this frequency the (resistive) input impedance of the cascode takes over. Since the admittance at the source of

M_2 approximately equals the transconductance g_{m2}, the pole p_1' is given by

$$p_1' = \frac{g_{m2}}{C_{m2}} \tag{5-45}$$

This expression replaces Eq. 5-32, which gave the position of the pole p_1' for the original Reversed Nested Miller case. We can further use the unaltered expressions Eq. 5-30 and Eq. 5-31 for the dimensioning of the cascode based Reversed Nested Miller circuit.

Without changing the basic functionality of the circuit, the capacitor C_{m1} in Fig. 5-18 can be moved to a location across the output transistor M_1. Figure 5-19 shows the modified circuit. In its new position, the capac-

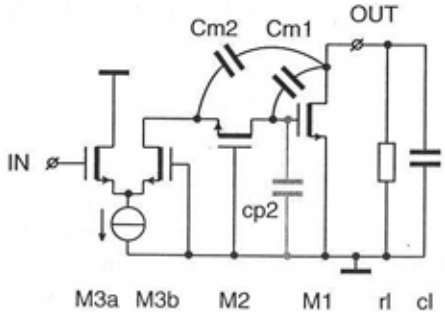


Fig. 5-19. *Modified cascode based Reversed Nested Miller amplifier.*

itor again converts the current supplied by the cascode M_2 into a voltage. The dimensioning equations of the new circuit express the similarity between the circuit of Fig. 5-18 and Fig. 5-19. The only difference is caused by the capacitive voltage division in the circuit of Fig. 5-19 due to the Miller capacitor C_{m1} and the interstage capacitance c_{p2}. The attenuation of the internal loop gain leads to a reduced bandwidth when C_{m1} is small.

The secondary pole frequency of the outer Miller capacitor loop p_1' is not affected by the voltage division and remains unalteredly given by

$$p_1' = \frac{g_{m2}}{C_{m2}} \tag{5-46}$$

For the outer Miller loop with C_{m2} we can find

$$\frac{C_{m1}}{C_{m1} + c_{p2}} \frac{C_{m2} g_{m1}}{C_{m1} c_l} = \frac{1}{2} p_1' \tag{5-47}$$

The left hand term of Eq. 5-47 expresses the voltage division due to the Miller capacitor C_{m1} and the interstage capacitance c_{p2}. Apart from this term, the equation is identical to Eq. 5-30 for ordinary Reversed Nested Miller compensation, and hence the expression for the shunt capacitor circuit of Fig. 5-18.

Finally, as before, the unity-gain frequency of the total amplifier should obey

$$\omega_t = \frac{g_{m3}}{C_{m2}} = \frac{1}{4} p_1' \tag{5-48}$$

An advantage of the modified cascode circuit of Fig. 5-19 over its shunt capacitor predecessor of Fig. 5-18 is the feedback that the Miller capacitor C_{m1} supplies across the output stage. This feedback helps to reduce the non-linear distortion. There is a comparable distinction in this respect between parallel and Miller compensation. Therefore, from a distortion point of view, the circuit of Fig. 5-19 is preferable.

From a bandwidth point of view, however, the shunt capacitor circuit of Fig. 5-18 holds the best cards. The position of the pole p_1' for ordinary Reversed Nested Miller compensation of Eq. 5-32 becomes identical to the pole position of the shunt based circuit, which was given by Eq. 5-45, and the pole position of the modified circuit, expressed by Eq. 5-46, when $C_{m1} = c_{p2}$. For the shunt capacitor based circuit of Fig. 5-18, this condition corresponds to the situation where only the interstage capacitance c_{p2} —and no added compensation capacitors— contributes to C_{m1}. This is also the optimal situation in terms of the obtainable bandwidth for this circuit. If we choose C_{m1} any larger than c_{p2}, the posi-

tion of the pole p_1' given by Eq. 5-45 will drop below the position indicated by Eq. 5-32, reducing the bandwidth of the amplifier.

Also in the modified circuit of Fig 5-19, the lower limit on C_{m1} is set by the interstage capacitance c_{p2}. In this case the capacitive voltage division in the Miller loop closed by C_{m1} is responsible for the limitation. If we choose C_{m1} too small, the gain around the internal loop will vanish. Choosing C_{m1} too high also works counter productive, because the capacitor will load the circuit in that case, reducing the obtainable bandwidth. These two opposite demands imply a trade-off that is very similar to the one discussed in Section 4.3.1 for two-stage Miller compensation. Resolving the trade-off leads to the requirement that C_{m1} should equal c_{p2}. In that case, however, the obtainable bandwidth of the circuit —all other parameters remaining the same— is a factor of two lower than that of the shunt capacitor circuit of Fig 5-18.

Although for both circuits a large capacitance C_{m1} is undesirable from a bandwidth view point, in practice the interstage capacitance c_{p2} may be too small to be able to realize this situation. Given that according to Eq. 5-33 and Eq. 5-47 the voltage gain C_{m2}/C_{m1} is fixed, a small value of C_{m1} directly results in a small overall compensation capacitor C_{m2}. Considerations other than the bandwidth, most notably the noise performance (see Section 4.4.2), can dictate a relatively high value of this overall Miller capacitor, however. This in its turn forces an increase of C_{m1}. In such situations the resistor schemes shown in Fig. 5-20 to prevent the loading of the amplifier by the Miller capacitors are advisable. We met a similar approach earlier in Fig. 5-17.

Fig. 5-20*a* shows the changes to the shunt capacitor circuit of Fig 5-18. Although capacitor C_{m1}' is inserted slightly differently, the dimensioning expressions Eq. 5-41-Eq. 5-43 of the original Reversed Nested Miller circuit from Fig. 5-17 remain valid. Note that the interstage capacitance c_{p2} ideally furnishes the bypass capacitor C_{m1}'.

The equations Eq. 5-41-Eq. 5-43, with a slight modification, also hold for the circuit of Fig. 5-20*b*. This is the loading reduction scheme for the circuit of Fig. 5-19. A modification is needed to account for the capacitive voltage divider C_{m1}'/c_{p2}. For correct operation, we must dimension the primed capacitors (C_{m1}' and C_{m2}') according to Eq. 5-47. Since in the optimal case Cm_1' equals c_{p2}, the attenuation term on the left of this equation equals exactly two. For the remainder of the dimensioning equations we can revert to the basic Reversed Nested Miller compensation discussion in the beginning of Section 5.3.

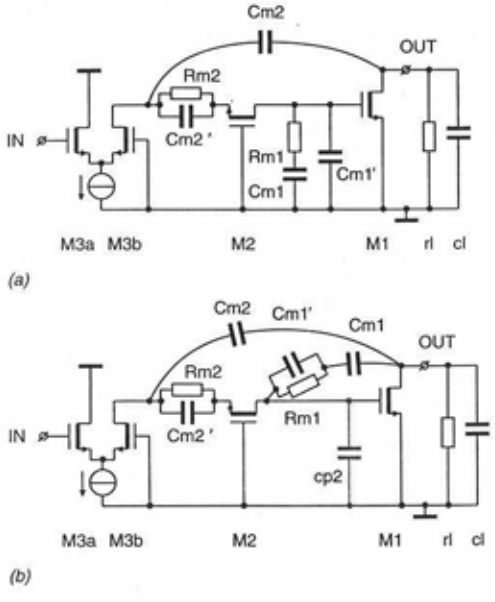


Fig. 5-20. *Cascode based Reversed Nested Miller compensation schemes with resistors to reduce the loading by the Miller capacitors.*

5.4. Hybrid Nested Miller compensation

The Nested Miller compensation technique of Section 5.2 and Reversed Nested Miller compensation discussed in Section 5.3 either depend on differential pairs to realize the correct feedback sign for the internal Miller loops or on cascodes. This dependency can be an important restriction for low-voltage amplifier circuits. Both the tail current sources of the differential pairs and the voltage drop across the cascodes require an increased supply voltage. This increase can be especially significant for CMOS amplifiers.

A compensation technique that overcomes the disadvantages of the other approaches for low-voltage applications is Hybrid Nested Miller compensation [15] [16]. This technique combines the topological features of Nested Miller and Reversed Nested Miller compensation to arrive at a circuit that, apart from the input stage, comprises of inverting amplifier stages only. Figure 5-21 shows the basic circuit. It consists of four transis-

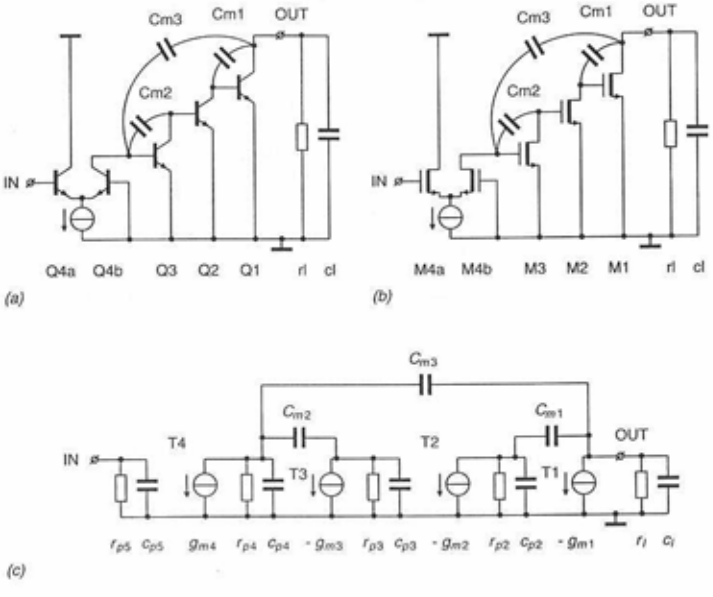


Fig. 5-21. *The Hybrid Nested Miller compensation technique.*

tor stages and three compensation capacitors. The fact that there are two capacitors, C_{m1} and C_{m2}, at the lowest nesting level instead of only one, reveals the joined influence of Nested and Reversed Nested Miller compensation. With Nested Miller compensation, the lowest level capacitor was connected across the output device, while the lowest nesting level of Reversed Nested Miller compensation was situated around the second stage. The Hybrid Nested Miller compensation scheme of Fig. 5-21 com-

bines both capacitor locations. The capacitor C_{m3} closes the outer Miller loop.

Apart from the low supply voltage, the advantages of Hybrid Nested Miller compensation comprise those of Nested and Reversed Nested Miller compensation combined. Because of the Miller loops at the same nesting level, two nesting levels suffice to frequency compensate the four-stage amplifier. The other techniques would require three levels for the same number of stages. The difference brings forth a bandwidth improvement of Hybrid Nested Miller compensation over Nested Miller compensation. The Miller loop with C_{m1} around the output transistor, on the other hand, results in an improved robustness of the circuit to load and output current variations as compared to Reversed Nested Miller compensation. We can easily confirm this as follows: to the first order, the output device and the Miller capacitor together constitute an integrator. The capacitor alone dictates the transfer function of this circuit part, which is therefore independent from the exact value of the load capacitance or the transconductance of the transistor.

For the purpose of analysis, Fig. 5-21*c* depicts the small-signal diagram of the Hybrid Nested Miller compensation circuit. The Bode plot of Fig. 5-22*a* shows the effect of the two Miller capacitors at the lowest nesting level C_{m1} and C_{m2}. The light line in the figure plots the frequency response of the circuit without the feedback capacitors. The four dominating poles are clearly visible. Inserting the two miller capacitors C_{m1} and C_{m2} causes the two poles at the input and the output of each of the transistor stages to split apart. Capacitor C_{m1} moves pole p_1 up to p_1' and p_2 down to p_2', while capacitor C_{m2} moves p_3 to the higher frequency p_3' and p_4 to the lower frequency p_4'. The shifting of the poles results in the black curve. The transfer contains the two dominating poles p_2' and p_4' and is the starting point for the plot of Fig. 5-22*b*. This figure shows the effect of the Miller capacitor C_{m3}. The insertion of the capacitor splits the poles p_2' and p_4' apart, as shown again by the black line. The result is the much desired 20dB/dec slope, up to the third order secondary pole p_1'', p_2'' and p_3''. The dominating pole frequency of the overall frequency response is p_4''.

We can derive the expressions for the dimensioning of the four-stage Hybrid Nested Miller compensation structure from the analysis of the Nested Miller compensation technique for three stages. The correspondence between the two compensation techniques is based on the identification of the voltage amplifier that is build up around transistor T_3. As was

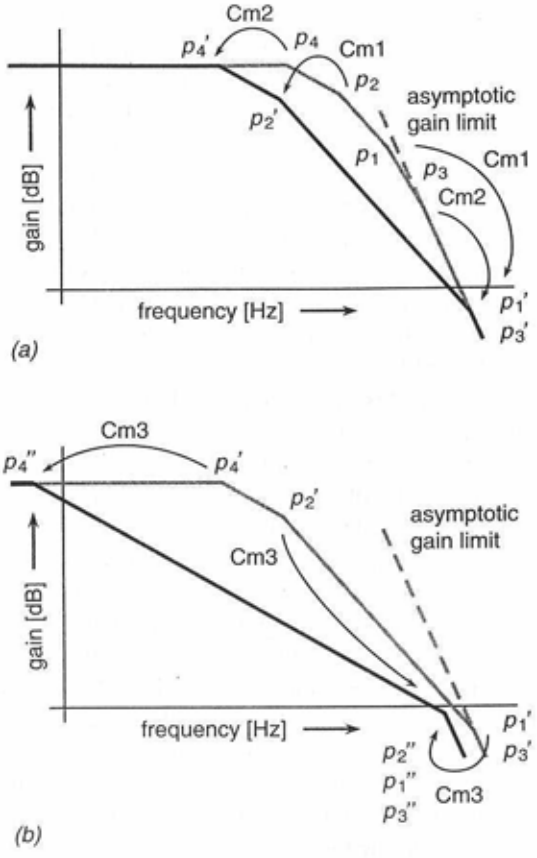


Fig. 5-22. *Bode plots of the Hybrid Nested Miller compensation structure with* (a) *the effect of* C_{m1} *and* C_{m2} *and* (b) *the effect of* C_{m3}.

the case with Reversed Nested Miller compensation, two Miller capacitors C_{m2} and C_{m3} and the transistor realize an inverting voltage gain function with a straight frequency response up to some limiting pole frequency (see Fig. 5-14). The capacitors C_{m2} and C_{m3} are the feedback elements of a

basic inverting amplifier. The ratio of the impedances of the feedback elements sets the voltage gain of such an amplifier. The gain equals C_{m3}/C_{m2}. Figure 5-23 symbolizes the gain of the basic amplifier by the triangle-

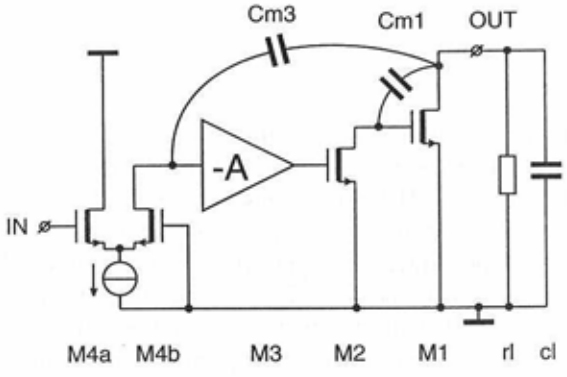


Fig. 5-23. *Hybrid Nested Miller structure reduced to a Nested Miller compensation topology.*

shaped amplifier block. This amplifier takes the place of transistor M_3. The resulting circuit is very similar to the Nested Miller compensation amplifier of Fig. 5-7. We can analyze it using the same calculations. Multiplying the transconductance g_{m2} of transistor T_2 by the proper amount can account for the internal gain that the additional transistor introduces. Rewriting the relevant relations of Section 5.2 leads to

$$K_1 = \frac{C_{m3}g_{m2}}{C_{m2}C_{m1}} = -\frac{1}{2}p_1' \qquad (5\text{-}49)$$

$$K_2 = \frac{g_{m4}}{C_{m3}} = -\frac{1}{4}p_1'$$

The unity-gain frequency ω_t associated with this solution is

$$\omega_t = K_2 = -\frac{1}{4}p_1' \tag{5-50}$$

The pole p_1', which is the highest due to the pole splitting around the output transistor T_1, again limits the bandwidth. The position of this pole was given earlier in Eq. 5-16:

$$p_1' = -\frac{g_{m1}}{c_l} \tag{5-51}$$

An additional design equation that has no equivalent in the Nested Miller compensation technique stems from the finite bandwidth of the voltage amplifier around T_3. The expressions Eq. 5-23 and Eq. 5-16 base themselves on the assumption that the frequency response of the internal amplifier has no notable effect on the performance of the circuit. To realize that condition, the parasitic pole that is associated with the amplifier must lie well above the bandwidth limiting pole p_1'. The finite admittance at the gate of T_3 causes the parasitic pole. It appears at approximately g_{m3}/C_{m3}. Eliminating the effect of this pole requires

$$\frac{g_{m3}}{C_{m3}} > p_1' \tag{5-52}$$

As already indicated, an advantage of Hybrid Nested Miller compensation is that all stages but the first can be of the inverting type. Since this property corresponds to the nature of the active components available, it helps to keep the realizing circuitry simple. One can easily see this from the simplified circuit in Fig. 5-21, where the last three stages are single common emitter/source transistors. In contrast, differential pairs implement the corresponding gain stages of Nested Miller compensation and Reversed Nested Miller compensation in Fig. 5-7 and Fig. 5-13, increasing the circuit complexity and power consumption.

Hybrid Nested Miller compensation is particularly suited for extremely low-voltage CMOS amplifiers. The low supply voltage dictates circuits based on common-source (inverting) transistor stages only, which matches the compensation technique. As an example, Section 7.3 presents an CMOS operational amplifier with Hybrid Nested Miller compensation which runs from a supply of less than 1.5 V.

5.4.1. Hybrid Nested Miller compensation for more than four stages

We can extend the Hybrid Nested Miller technique to any even number of stages more than four. The approach is similar to the pursuit made in Section 5.2 to transform single Miller compensation into Nested Miller compensation. Considering the three stages T_1, T_2, T_3 plus the Miller capacitor C_{m3} in Fig. 5-21 as a single Miller compensated 'super transistor', we can recursively substitute each of the Miller compensated transistor stages of the four-stage Hybrid Nested Miller compensation circuit by the new composite device. Figure 5-24 shows the two variations that fol-

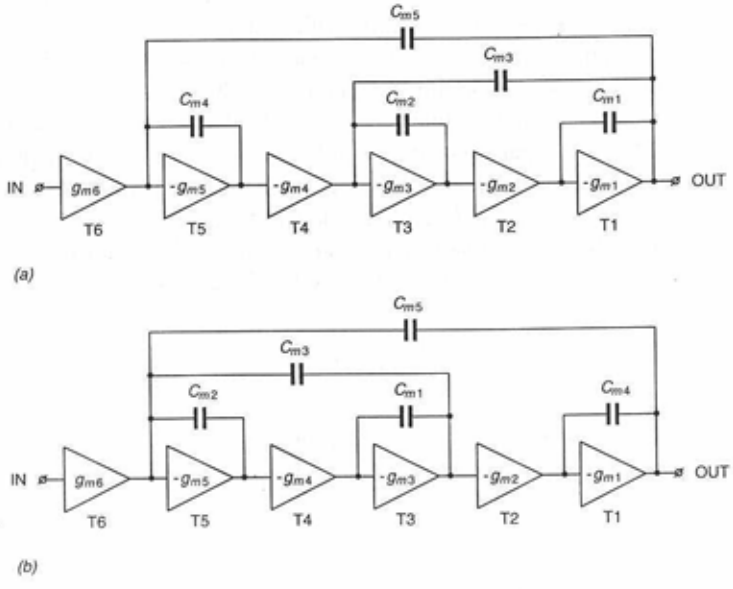


Fig. 5-24. *The two possible six-stage Hybrid Nested Miller compensation topologies.*

low when we subsequently replace T_1 and T_3. Both amplifiers contain six stages. We can easily derive the design equations from the expression that

we found for the simple case of four-stage Hybrid Nested Miller compensation. For the circuit of Fig. 5-24*a*, we have to take into account the reduction of the limiting pole frequency p_1' due to the additional nest in the output part. Referring to the analysis in the beginning of Section 5.4 of Hybrid Nested Miller compensation, the new value for p_1' that we must introduce into the calculations is half the original value given by Eq. 5-16. In other words

$$p_1' = -\frac{1}{2}\frac{g_{m1}}{c_l} \tag{5-53}$$

Starting from this pole position, we can reuse the earlier calculations for Hybrid Nested Miller compensation with four stages.

The additional nest of the (composite) second stage of Fig. 5-24*b* also requires a slight modification of the mathematics involved. In this case it is the condition introduced in Eq. 5-52 that needs revision. Assuming that the load capacitance c_{p3} of T_3 is sufficiently small, the Miller loop around T_5 fully determines the bandwidth of the Miller compensated 'super transistor' T_3, T_4, T_5 and the capacitor C_{m4}. The two nested Miller loops of the 'super transistor' again cause a factor of two down shift of the relevant pole. Therefore we may rewrite Eq. 5-52 into

$$\frac{g_{m5}}{C_{m5}} > 2p_1' \tag{5-54}$$

With this modification, we can directly apply the four-stage Hybrid Nested Miller compensation analysis to the six-stage circuit of Fig. 5-24*b*.

5.5. Merging of compensation techniques

A virtually unlimited number of new compensation methods results when we combine and permutate the compensation techniques of this chapter. The cyclic permutation of the order of the stages is generally an effective method to change the signal domain of the amplifier. One can readily change a voltage-mode circuit into a current amplifier and vice verse. Combining compensation techniques typically leads to amplifiers with more stages. The previous section demonstrated an example of a com-

bined scheme. Repeating the structure of Hybrid Nested Miller compensation in itself increased the number of compensatable stages. A similar approach in Section 5.2.1 leads to Nested Miller compensation starting from two-stage single Miller compensation. Applying the approach of cyclic permutation of the order of the stages to the Nested Miller compensation technique resulted in the Current Mode Nested Miller compensation scheme of Section 5.2.2.

The compounding of techniques is not restricted to Miller compensation alone. Figure 5-25 shows a topology where the output stage of the

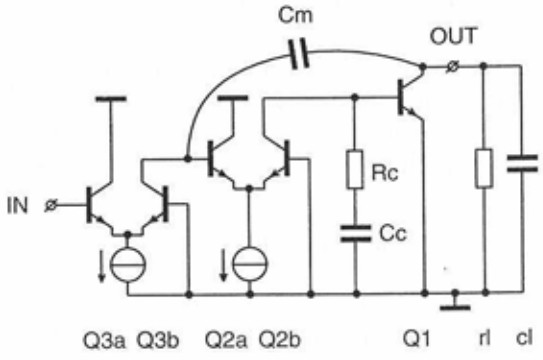


Fig. 5-25. *Three-stage combined parallel and Miller compensation.*

Miller compensated circuit of Fig. 4-7 incorporates the parallel compensated amplifier of Fig. 4-5. A theoretical advantage of the 'parallel-Miller' circuit is the absence of nested Miller loops, suggesting a higher bandwidth than that which we could obtain with Nested Miller compensation, for example. The disadvantages are numerous, however. The dependency on parallel compensation is mainly responsible for these drawbacks, which range from sensitivity to parameter variations to pole-zero doublets in the passband. Nevertheless, literature reports techniques based on the combination of parallel and feedback compensation. Since many of these amplifiers employ discrete components, the parallel compensation section

is allowed to depend on the use of inductors to optimally position the poles and the zeros [17].

5.6. Conclusions

We have seen that low-power operational amplifiers require compensation techniques that are suited for a low supply voltage. Traditional methods of compensating amplifiers with three stages or more do not comply with this demand. Other, more complex, approaches are necessary to achieve the maximum bandwidth-to-power ratio. The multistage techniques discussed in this chapter are: parallel compensation, Nested Miller compensation (NMC), Reversed Nested Miller compensation (RNMC) and Hybrid Nested Miller compensation (HNMC). The advantages that we arrived at in the previous chapter of Miller compensation over parallel compensation remain valid and in most cases becomes more apparent when the number of stages increases. Table 5-2 summarizes the properties of each of the dis-

Table 5-2. *Summarized properties of the various compensation techniques.*

	parallel	**NMC**	**RNMC**	**HNMC**
bandwidth/power	++	o	+	+
capacitor values	--	++	++	++
ruggedness	--	++	-	++
pole-zero doublets	--	++	++	++
need for cascodes	-	++	++	++
distortion	-	++	-	+
low supply voltage	++	+	+	++

++=excellent, +=good, o=moderate, -=poor, --=very poor

cussed techniques. Clearly the suitability of the Nested Miller techniques for monolithic integration pop into the eye. Most notably are the low required capacitor values and the ruggedness against parameter variations.

Another interesting point is that a high bandwidth-to-power ratio automatically seems to imply deterioration of the robustness against parameter variations. We see that for instance with parallel and Reversed Nested Miller compensation. The reason is that a compensation technique that approaches the theoretical limit set by the parameters of the circuit simply must be sensitive to any change of these parameters. Varying one of the parameters, for example the load capacitance, leads to a new optimum for the bandwidth and therefore requires redimensioning of the circuit. Compensation methods that do not quite reach the maximum obtainable bandwidth have a inherent reserve that is available for an improved ruggedness against parameter variations. This becomes manifest for Nested Miller compensation for instance, which couples a moderate bandwidth potential to an excellent robustness.

6 Multipath Compensation Techniques

In many of the compensation techniques described in the previous chapter, the addition of stages to the circuit leads to a reduction of the obtainable bandwidth. Going from simple two-stage Miller compensation to Nested Miller compensation with three gain stages, for instance, results in a fifty percent smaller bandwidth. Also Hybrid Nested Miller compensation suffers from this drawback. Although the bandwidth reduction effect is absent in Reversed Nested Miller compensation, the dependency of this compensation technique on the load capacitance makes it fundamentally unsuited for general purpose applications. Therefore we would like to eliminate the bandwidth reduction in Nested and Hybrid Nested Miller compensation, while retaining the high level of robustness inherent to these methods.

This chapter presents the Multipath technique to avoid the step back in bandwidth of certain nested compensation schemes. The technique is based on a parallel signal path that bypasses one or more transistors of the main path. This bypass realizes a short high-frequency signal path, while the circuit maintains the gain advantage for low frequencies of a longer path with more amplifier stages. A key aspect of the Multipath technique is that we can configure the parallel signal path independently from the main path. This avoids the appearance of pole-zero doublets in the passband. Earlier feedforward circuits [18] mostly base themselves on the insertion of capacitors to create a high frequency bypass for critical components, such as lateral PNPs. These bypass capacitors result in strong

and inevitable pole-zero doublets, however, disqualifying the circuits for a large number of applications where fast settling is of importance. Pole-zero doublets are pairs of a closely spaced pole and zero. Their presence deteriorates the settling performance of an amplifier by introducing signal components into the step response with a time constant much larger than the one associated with the bandwidth of the circuit. Section 6.2 will address the relation between the occurrence of pole-zero doublets and these so-called slow-settling components.

6.1. Multipath Nested Miller compensation

The first candidate for bandwidth improvement is the Nested Miller compensation technique described in Section 5.2. This section showed that three-stage Nested Miller compensation has a factor two bandwidth penalty as compared to ordinary, two-stage, Miller compensation. This is related to the down shift of the bandwidth limiting pole p_1' caused by the outer Miller loop. The root locus of Fig. 5-9*b*, which is repeated in Fig. 6-1, illustrated the down shift. Closing of the outer Miller loop shifts down the bandwidth limiting pole p_1', until it collides with pole p_3. The collision occurs at approximately half the frequency of p_1'. The addition of more stages, and consequently more Miller capacitor nests, even further reduces the obtainable bandwidth.

We can overcome the bandwidth reduction caused by the nested Miller loops by adding a Multipath input stage that provides an independent path directly driving the output stage [20]. The result is the Multipath Nested Miller compensation topology depicted in Fig. 6-2. In this figure, transistors T_1 through T_3, together with the two capacitors C_{m1} and C_{m2}, build up the conventional Nested Miller compensation structure. The Multipath input stage is transistor pair T_4. This differential pair directly drives the output transistor, bypassing T_2 for high frequencies.

In Figure 6-3 the Bode plot of the Multipath Nested Miller compensated amplifier is shown. The gray line represents the high-gain, low-frequency part established by the three-stage Nested Miller compensated amplifier. The black line shows the high frequency part created by the two transistors T_4 and T_1 and Miller capacitor C_{m1}. Adding the two plots yields a straight 20dB/dec frequency response of which only the low frequency part is provided by the three-stage amplifier. The high frequency portion is

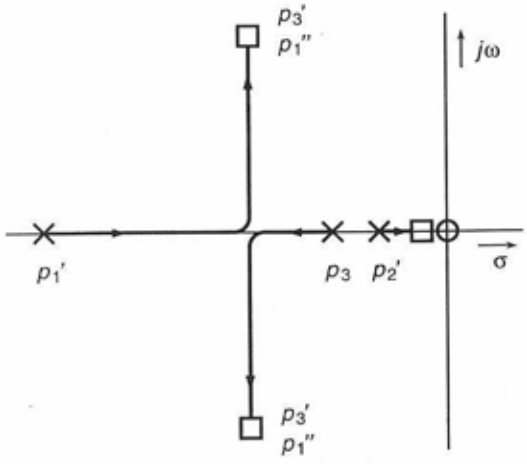


Fig. 6-1. *Root locus showing the bandwidth reduction caused by the closing of the outer Miller loop of Nested Miller compensation.*

determined by two stages. Therefore, the obtainable bandwidth approaches the theoretical maximum of a two-stage amplifier with single Miller compensation. A major advantage of the Multipath technique is that the matching of the high and low frequency part of the total frequency response is duly controllable. This is important, since even a slight mismatch can deteriorate the settling behavior of the amplifier, as demonstrated in Section 6.2. The following analysis confirms the accurate matching properties.

The Multipath input stage has no effect on the position of the poles of the original circuit. This is caused by the fact that the current generated by the input stage only depends on the input voltage. Therefore, we can regard this current as coming from an independent input signal source. We can use this observation fruitfully in Eq. 2-11, which expresses the transfer between any two nodes of a network as the ratio Δ_{ij}/Δ. Since any input source enters the cofactor in the numerator of this equation, the addition of a branch that carries a current proportional to the input signal only affects

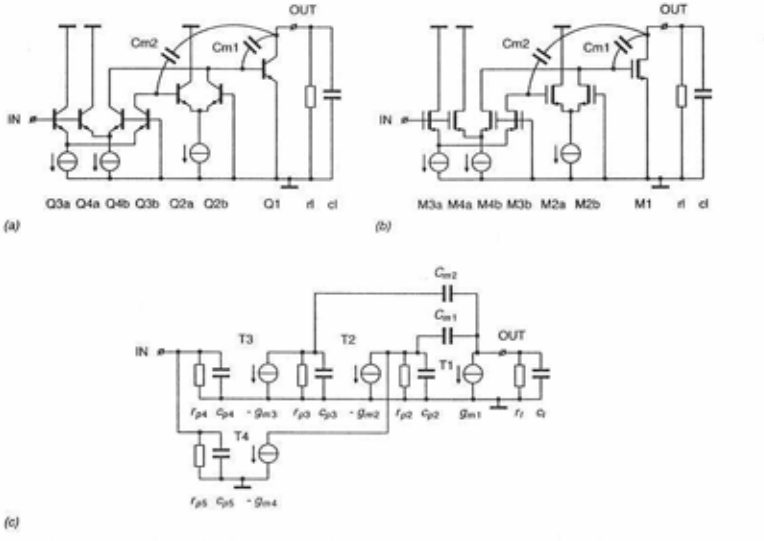


Fig. 6-2. *Three-stage amplifier with Multipath Nested Miller compensation.*

the number and position of the zeros. The poles, on the other hand, remain fixed at their original locations.

In the case of Multipath Nested Miller compensation the pole locations are identical to those of the Nested Miller compensation technique. These follow from solving the denominator of Eq. 5-19 and are given by

$$p_1'' = p_1'\left(\frac{1}{2} + \frac{1}{2}\sqrt{1 - \frac{4g_{m2}}{p_1' C_{m1}}}\right) \tag{6-1}$$

$$p_3' = p_1'\left(\frac{1}{2} - \frac{1}{2}\sqrt{1 - \frac{4g_{m2}}{p_1' C_{m1}}}\right) \tag{6-2}$$

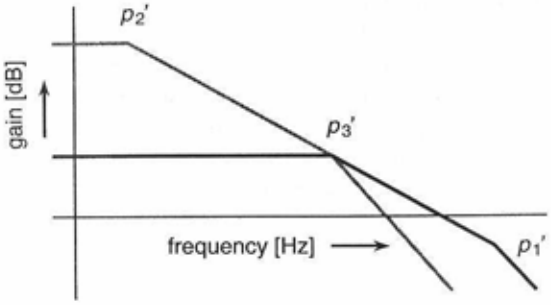


Fig. 6-3. *Bode plot of the Multipath Nested Miller compensated amplifier.*

under the condition that $p_2'' \ll p_1''$. Nodal analysis reveals that the Multipath input stage adds a single zero to the transfer function of the Nested Miller circuit. The position of this zero z_m is given by

$$z_m = -\frac{g_{m2}g_{m3}}{g_{m4}C_{m2}} \tag{6-3}$$

To obtain a straight 20dB/dec frequency response, the zero z_m must exactly cancel out the pole p_3'. This leads to the condition

$$\frac{g_{m4}}{C_{m1}} = \frac{p_1''}{p_1'}\frac{g_{m3}}{C_{m2}} \tag{6-4}$$

The ratio p_1''/p_1' denotes the relative movement of the bandwidth limiting pole p_1'. We can derive the ratio from Eq. 6-1. It equals

$$\frac{p_1''}{p_1'} = \frac{1}{2} + \frac{1}{2}\sqrt{1 - \frac{4g_{m2}}{p_1'C_{m1}}} \tag{6-5}$$

To realize the maximum bandwidth, the down shift of the pole p_1 must be as small as possible, which corresponds to maximizing Eq. 6-5. The optimum value is attained when the transconductance of the middle stage g_{m2} is set to zero. In that case no bandwidth reduction takes place, since in the absence of T_2 the circuit of Fig. 6-2 reduces to a two-stage amplifier consisting of the Multipath input stage T_4 and the output stage T_1. Although under these conditions the bandwidth is at its theoretical maximum, clearly this situation is not preferable from the gain point of view. Fortunately a compromise is possible with high gain and only a fractional reduced bandwidth. The compromise is given by

$$\frac{g_{m2}}{C_{m1}} \approx \frac{p_1'}{10} \tag{6-6}$$

With this ratio, the bandwidth reduction following from Eq. 6-5 only amounts to about 10%, and still three stages contribute to the low frequency gain. Since the relative bandwidth reduction is close to unity, we can simplify Eq. 6-4, giving the condition for pole-zero cancellation, into

$$\frac{g_{m4}}{C_{m1}} = \frac{g_{m3}}{C_{m2}} \tag{6-7}$$

In other words, to obtain a straight first-order frequency response, the unity-gain frequency determined by the original input stage T_3 and the outer Miller capacitor C_{m2} must equal the unity-gain frequency of the Multipath input stage T_4 and the inner Miller capacitor C_{m1}. One can also understand this condition intuitively. The left-hand part of Eq. 6-7 defines the slope of the frequency response of the two-stage single Miller compensated amplifier comprising T_1 and T_4, while the part on the right gives the slope of the three-stage Nested Miller compensated amplifier T_1, T_3 and T_4. Matching of the two slopes results in an unbroken curve.

Since the dimensioning of Multipath Nested Miller compensation depends on the ratios of capacitors and transconductances only —both being among the best controlled parameters of an I.C. process— matching can be very close. In a typical bipolar process the mismatch can be as low as 0.1%. CMOS technology yields slightly less controllable ratios, resulting in a mismatch of about 5%. We can find the position of the pole-zero doublet as

$$\omega_d = \frac{g_{m2}}{C_{m1}} \tag{6-8}$$

or, with Eq. 6-6

$$\omega_d \approx \frac{p_1'}{10} \approx \frac{\omega_t}{5} \tag{6-9}$$

The second approximation of this equation is based on the fact that, for a phase margin of 60°, the unity-gain frequency ω_t of the opamp should be half p_1'' ($\approx p_1'$). The unity-gain frequency is therefore given by

$$\omega_t = \frac{g_{m4}}{C_{m1}} = \frac{g_{m3}}{C_{m2}} = \frac{1}{2}p_1' \tag{6-10}$$

which is indeed a factor of two higher than the unity-gain frequency of Nested Miller compensation given by Eq. 5-23. Equation 6-6, Eq. 6-7 and Eq. 6-10 are the dimensioning expressions for Multipath Nested Miller compensation. The pole p_1' is given as before by $p_1' = g_{m1}/c_l$.

The introduction of the factor 1/10 in Eq. 6-6 indicates that the bias currents of the Multipath Nested Miller compensated circuit can be considerably lower than those of its Nested Miller compensated counterpart. In practical situations this will about balance the extra current required for the Multipath input stage. Therefore, the bandwidth improvement of the Multipath technique goes without a current penalty and the bandwidth-to-power ratio is improved. Also the die area of the chip does not increase when using Multipath Hybrid Nested Miller compensation, since the higher bandwidth allows for smaller compensation capacitors.

The root locus in Fig. 6-4 shows the movement of the poles in the Multipath Nested Miller compensation structure as a result of the introduction of capacitor C_{m2}. In contrast to Nested Miller compensation, closing the outer Miller loop only moves the poles slightly because of the low value g_{m2}/C_{m1}. Pole p_3' is eliminated by the Multipath zero z_m. Since pole p_1' remains virtually motionless, no bandwidth reduction occurs as compared to the single Miller compensation technique.

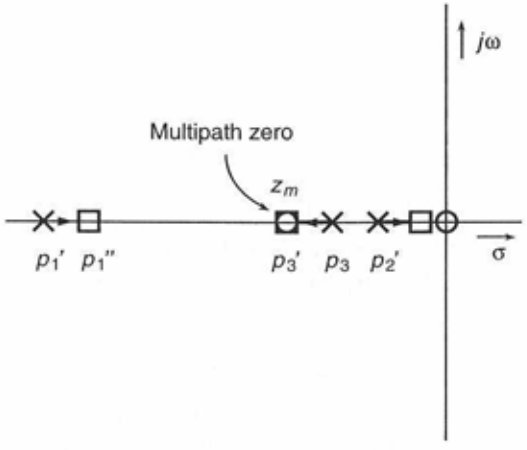


Fig. 6-4. *Multipath Nested Miller compensation's root locus showing the effect of the outer Miller capacitor* C_{m2}.

6.1.1. Multipath Darlington

The low transconductance of the intermediate stage of Multipath Nested Miller compensation makes a special type of circuit possible where the intermediate inverting amplifier stage is substituted by a voltage follower. The resulting circuit presents a wide bandwidth alternative to the Darlington configuration.

The Multipath Darlington is shown in Fig. 6-5. It consists of the output stage Q_1 and the emitter follower intermediate stage Q_2. The signal current sources on the left model the normal input stage Q_3 and the Multipath input stage Q_4. The current of each of the sources equals the input voltage of the amplifier times the respective transconductance g_{m3} and g_{m4}. The current that the Multipath input stage directly provides to the output transistor results in a bandwidth improvement over the normal Darlington composite. In the traditional configuration the output impedance of

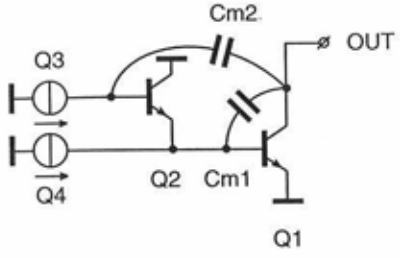


Fig. 6-5. *Improved Darlington composite based on a Multipath Nested Miller compensation circuit with an emitter follower as the intermediate stage.*

the emitter follower Q_2 and the base resistance of Q_1 together with base-emitter capacitance of Q_1 constitute a parasitic pole that deteriorates the high frequency behavior of an amplifier. If the circuit is not designed properly, the pole will lead to the so-called 'output bump' [11]. Reducing the effect of the parasitic interstage pole involves increasing the current through the emitter follower Q_2 and up-scaling of the output transistor Q_1. It is evident that such measures are very undesirable from the standpoint of power consumption and occupied die area.

The dimensioning equations of Multipath Nested Miller compensation given in the previous section directly apply to the circuit of Fig. 6-5. Complying with Eq. 6-6, Eq. 6-7 and Eq. 6-10 yields a Butterworth frequency response for unity-gain frequency feedback. The low transconductance of Q_2 that follows from Eq. 6-6 is the reason that we can use an emitter follower as the intermediate stage without deteriorating the performance. If the transconductance were too high, the low output impedance of the emitter follower would suppress the gain of the Miller feedback loop around the output transistor Q_1 and the frequency compensation would fail. We do not expect any deterioration of the compensation when the output impedance of Q_2 is smaller than or equal to the base-emitter impedance of the output transistor Q_1. In this situation the Miller loop with C_{m1} remains active, since most of the current flowing from the output

through C_{m1} is sunk in the base of Q_1, instead of the emitter of Q_2. The output impedance of Q_2 is given by

$$z_{o2} = \frac{1}{g_{m2}} \tag{6-11}$$

while the input impedance of Q_1 (without the Miller capacitor) equals

$$z_{i1} = \frac{\beta_f}{g_{m1}} \tag{6-12}$$

where β_f is the current gain of the output transistor. We can rewrite the condition to prevent the emitter follower from affecting the Miller loop around the output stage, $z_{o2} \geq z_{i1}$, with Eq. 6-11 and Eq. 6-12 as

$$\frac{g_{m1}}{\beta_f} \geq g_{m2} \tag{6-13}$$

We can also state this expression in terms of the quiescent currents I_{c1} and I_{c2} of the respective transistors, since the transconductance is proportional to the collector current

$$\frac{I_{c1}}{\beta_f} \geq I_{c2} \tag{6-14}$$

Noting that the current sources on the left of Fig. 6-5 only supply the input signals and no bias currents, the condition of Eq. 6-14 is exactly fulfilled for the equal sign by the circuit of Fig. 6-5. The emitter current of Q_2 is equal to the base current of Q_1. This base current in its turn equals the collector current I_{c1} divided by the current gain β_f.

We cannot apply the Darlington configuration —be it conventional or Multipath— in amplifiers operating at an extremely low supply voltage anymore. At voltages in the order of 1 V the stacked base-emitter voltages of the Darlington circuit exceed the supply voltage itself. The Darlington output stage therefore needs replacement under these circumstances. Figure 6-6 shows a low-voltage Darlington substitute based on the Multipath Nested Miller compensation structure [21]. In this circuit the inverting gain stage Q_2 provides the intermediate stage. To correct the sign of

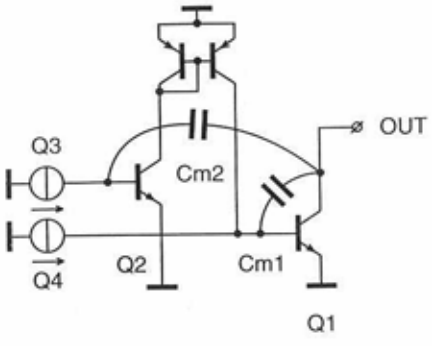


Fig. 6-6. *Low-voltage Darlington substitute.*

the feedback loop closed by C_{m2}, the drain current of Q_2 is fed through the —inverting— current mirror. In this fashion Q_1, Q_2 and the two Miller capacitors C_{m1} and C_{m2} constitute a Nested Miller structure. The Multipath signal path is provided by the input stage Q_4, again modelled by its collector current source alone. Since the Multipath input stage provides a path for high frequencies, the small-bandwidth PNP current mirror is only faced with low frequency signals. The low-voltage aspect becomes clear from the absence of stacked emitter voltages in the circuit of Fig. 6-6. A minimum supply voltage of one saturation voltage plus an emitter voltage, amounting to approximately 800 mV, suffices.

6.2. Pole-zero doublets and settling time

The Multipath compensation technique introduces a combination of a closely spaced pole and zero in the passband of the amplifier. Although we found in the previous section that the matching of this pole-zero doublet is well-controllable, we would still like to know what its effect is on the settling times of the amplifier. For this purpose, consider the amplifier A of Fig. 6-7, which displays a frequency response governed by the dominant

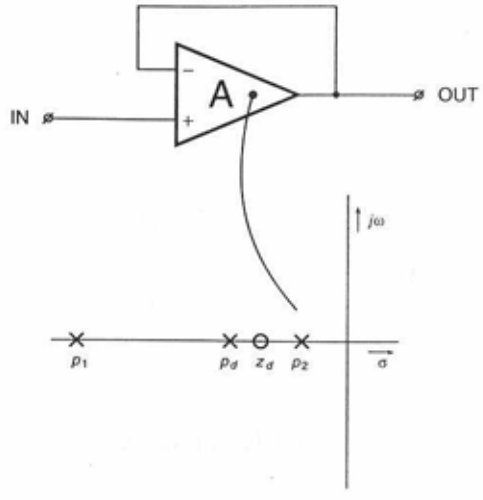


Fig. 6-7. *Amplifier with pole-zero doublet in the passband.*

pole p_2 and the secondary pole p_1. In the passband, at the frequencies ω_p and ω_z, the pole-zero doublet p_d/z_d is present. The pole and the zero are closely spaced so that their effects on the total frequency response almost cancel out. The position of the poles and the zero is presented in the complex s-plane. Applying unity-gain frequency feedback around the amplifier, the doublet is compressed, making it practically impossible to detect in the Bode plot of the buffer. This is illustrated by the root locus in Fig. 6-8. The poles p_1 and p_d end up in the Butterworth position p_1' and p_d', while the pole p_1 and the zero z_d constitute a new pole-zero doublet at approximately the same frequency, but with closer spacing than the original doublet p_d/z_d.

To determine the effect of the pole-zero doublet on the settling time, we need to consider the settling of the Butterworth poles, which set the bandwidth of the amplifier and the doublet separately. The settling time of a second order Butterworth transfer to an accuracy a is given by [3]

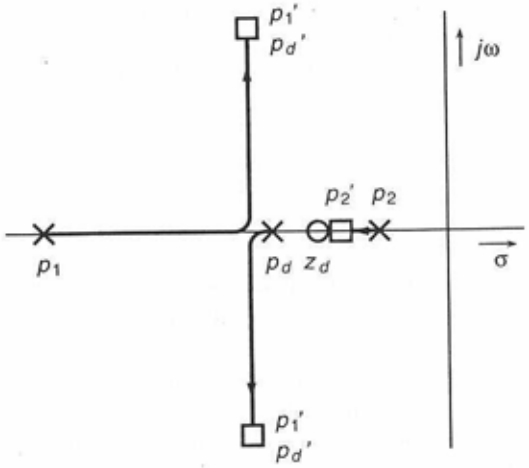


Fig. 6-8. *Root locus of the unity-gain frequency feedback loop. The squares indicate the pole positions after closing the loop.*

$$T_{s,b} = \frac{-\ln a}{\omega_c \frac{1}{2}\sqrt{2}} \tag{6-15}$$

where ω_c is the corner frequency of the Butterworth response. Since this corner frequency ω_c is related to the open-loop unity-gain frequency ω_t of the amplifier by $\omega_c = \omega_t\sqrt{2}$, Eq. 6-15 reduces to

$$T_{s,b} = \frac{-\ln a}{\omega_t} \tag{6-16}$$

The pole-zero doublet causes an exponentially decaying component in the response to a unity step input signal. The time constant of the exponential signal corresponding to the doublet's frequency, while the magnitude is given by $(\omega_p - \omega_z)/\omega_t$. Clearly, when the magnitude of the slow-settling

component is smaller than the required accuracy of the settling itself, we do not have to expect degradation due to the pole-zero doublet. When the required accuracy becomes higher, however, the time constant of the doublet instead of the amplifier's limited bandwidth starts dominating the settling time. The settling time of the pole-zero doublet is given by [8], [3]

$$T_{s,d} = \frac{\ln \frac{S_d \omega_d}{a \omega_t}}{\omega_d} \tag{6-17}$$

where ω_d is the doublet's frequency and S_d denotes the positive number value of the relative spacing of the pole-zero doublet $\left|(\omega_p - \omega_z)/\omega_d\right|$. In the case of the Multipath Nested Miller compensation scheme of Section 6.1, the spacing S_d corresponds to the relative mismatch of the condition Eq. 6-7 for the matching of the high and low frequency part of the total frequency response. Notice that, because of the close spacing of the pole and the zero, we may choose the doublet frequency ω_d arbitrarily, either the frequency of the pole ω_p or the zero ω_z.

An example, demonstrating how the slow-settling components can dominate the settling time of an amplifier, is provided by the following parameters: $\omega_t = 2\pi \cdot 100\,\text{MHz}$, $S_d = 10\%$. In this case the settling time to $a = 0.1\%$ due to the bandwidth of the amplifier is, according to Eq. 6-16, $T_{s,b} = 11\,\text{ns}$. The settling time of the pole-zero component results from Eq. 6-17 and equals $T_{s,d} = 29\,\text{ns}$. Clearly, the slow-settling component will deteriorate the overall settling behavior of the amplifier. As we can expect, the situation becomes more severe when the settling to $a = 0.01\%$ is considered. In that case, the time constant of the pole-zero doublet almost totally determines the settling time. Settling due to the bandwidth takes $T_{s,b} = 15\,\text{ns}$, while the doublet leads to $T_{s,d} = 53\,\text{ns}$. Figure 6-9 shows a typical step response of an amplifier with a pole-zero doublet in its transfer characteristic. The gray curve indicates the response without the doublet, while the black plots include the effect of the slow-settling component. The latter response is shown for two different pole-zero positions. The upper trace represents the case that the doublet pole is at a lower frequency than the zero. The lower trace depicts the opposite case.

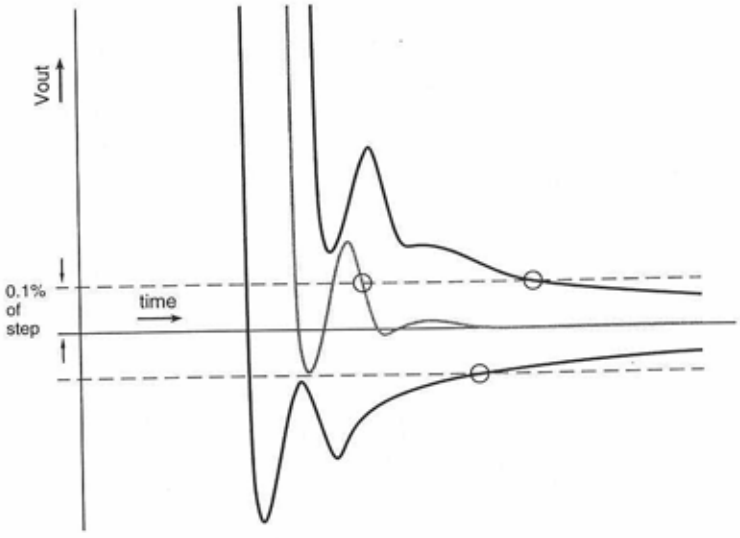


Fig. 6-9. *Typical step response showing the effect of slow-settling components on the step response. The gray curve is the response without slow-settling effects, the black curves give the slow-settling step responses for various pole-zero positions.*

6.3. Multipath Hybrid Nested Miller compensation

Although the bandwidth reduction of Hybrid Nested Miller compensation is small compared to Nested Miller compensation, in some cases even this small reduction is not acceptable. In this situation the Multipath technique is an attractive option resulting in Multipath Hybrid Nested Miller Compensation [15] [16]. Figure 6-10 shows a Multipath Hybrid Nested Miller compensated amplifier. The Multipath input stage T_5 that directly drives the output transistor T_1 is added to the Hybrid Nested Miller compensated circuit of Fig. 5-21. As with Nested Miller compensation, the extra input stage and the output stage together constitute a two-stage amplifier with

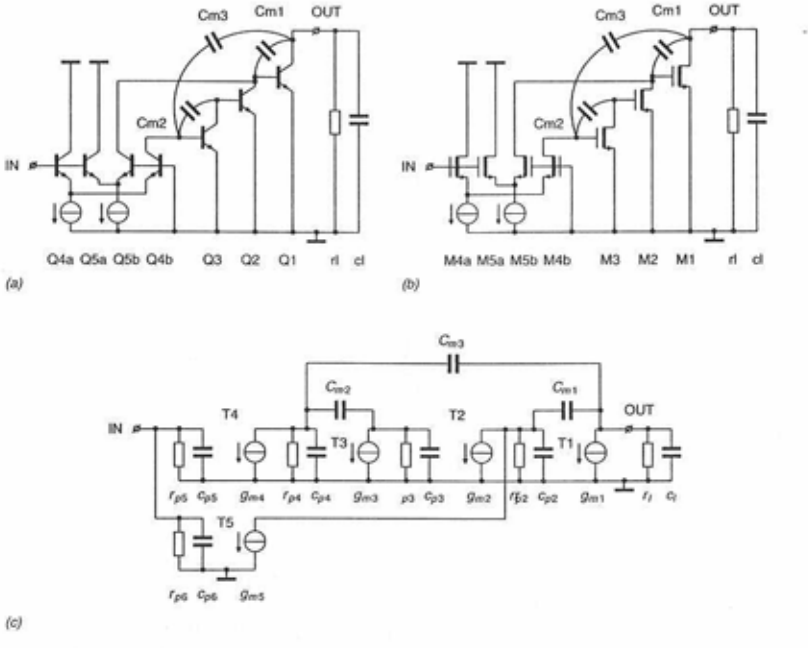


Fig. 6-10. *Amplifier with Multipath Hybrid Nested Miller compensation.*

relatively low gain, but a high bandwidth. The remainder of the circuit, a four-stage amplifier with Hybrid Nested Miller compensation, pairs a low bandwidth and a high gain. When the two circuit parts properly match, the result is a 20dB/dec roll-off with both a high unity-gain frequency and a high gain.

The analysis of Multipath Hybrid Nested Miller compensation is almost identical to that of Multipath Nested Miller compensation when the voltage gain around the second stage is taken into account according to the scheme of Fig. 5-23. Equation 6-6 therefore becomes

$$\frac{g_{m2}C_{m3}}{C_{m1}C_{m2}} \approx \frac{p_1'}{10} \tag{6-18}$$

Equation 6-7, which denotes the condition for pole-zero matching, for Multipath Hybrid Nested Miller compensation changes into

$$\frac{g_{m5}}{C_{m1}} = \frac{g_{m4}}{C_{m3}} \tag{6-19}$$

For the frequency of the pole-zero doublet, Eq. 6-9 remains valid:

$$\omega_d \approx \frac{p_1'}{10} \approx \frac{\omega_t}{5} \tag{6-20}$$

The unity-gain frequency of the Multipath Hybrid Nested Miller compensated circuit equals

$$\omega_t = \frac{g_{m5}}{C_{m1}} = \frac{g_{m4}}{C_{m3}} = \frac{1}{2}p_1' \tag{6-21}$$

To assure that the limited bandwidth of the voltage amplifier around T_3 does not deteriorate the overall performance of the circuit, Multipath Hybrid Nested Miller compensation puts an additional demand on the dimensioning. This condition is comparable to the condition found in Eq. 5-52 for Hybrid Nested Miller compensation without the Multipath input stage. Due to the easing of the bandwidth requirement of the four-stage amplifier part (it merely has to supply the gain) the condition for Multipath Hybrid Nested Miller compensation relaxes into

$$\frac{g_{m3}}{C_{m3}} > \frac{p_1'}{10} \tag{6-22}$$

6.4. Multipath Miller Zero Cancellation

In the Multipath techniques presented in the previous sections, the second signal path drives the output transistors. The bypassing of amplifier stages is therefore limited to all transistors but the last. We can obtain the Multipath Miller Zero Cancellation topology by extending the second path to the output terminal [22] [23]. Instead of introducing a zero in the transfer function, the Multipath Miller Zero Cancellation technique improves the bandwidth by removing the Right Half Plane (RHP) zero that arises in Miller compensated amplifiers. An advantages of the technique compared to traditional methods for removing this zero is that it is independent of the transconductance of the output stage. Especially in class AB output stages with a large quiescent-to-maximum current ratio this is of interest. The large variations of the currents through the output transistors of such circuits on their turn lead to large variations of the transconductances. Under such circumstances removal techniques that depend on the transconductance become unsuited.

Figure 6-11 shows a two-stage Miller compensated amplifier, con-

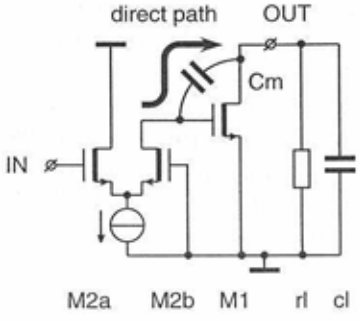


Fig. 6-11. *Miller compensated amplifier with direct current path through the Miller capacitor, degrading the phase margin at high frequencies.*

sisting of the output stage M_1 and the input differential pair M_{2a}/M_{2b}. A

zero in the right half of the complex s-plane occurs due to the direct path the Miller capacitor creates from input to output. A current that follows this path is in the opposite phase to the drain current of the transistor. The opposite phase current reduces the bandwidth in a feedback system since, for high frequencies, it changes the sign of the feedback from negative to positive. This phase reversal leads to a severely degraded phase margin of the amplifier. The position of the RHP zero is given by

$$z = \frac{g_{m1}}{C_m} \tag{6-23}$$

Figure 6-12 shows how the RHP zero changes the frequency characteristic

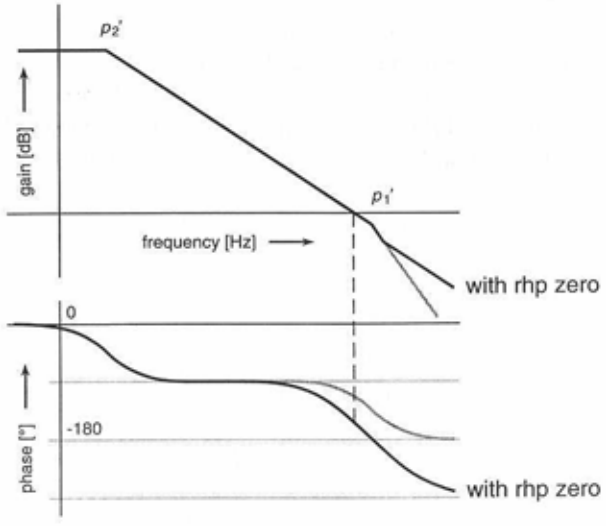


Fig. 6-12. *Effect of the Right Half Plane zero on the frequency response of the amplifier.*

of the amplifier. The light line in the figure shows the frequency response without the zero, the black curve shows the one with the RHP zero.

Although the frequency response is lifted due to the zero, its position in the right half of the *s*-plane increases the phase shift at higher frequencies. This on its turn leads to a diminishing of the phase margin.

The effect of the RHP zero is inversely proportional to the transconductance of the transistor. Both bipolar and MOS circuits suffer from the bandwidth degradation of the zero, but because MOS devices normally have a lower transconductance than bipolar transistors, the RHP zero most strongly affects these types of circuits. In many cases the RHP zero determines the maximum bandwidth of a MOS amplifier. But also in bipolar designs, the RHP zero can be a bandwidth limiting factor.

A conventional solution for the RHP zero is the insertion of a small resistor in series with the Miller capacitor [19]. This resistor must closely match the transconductance of the output stage. In circuits with largely varying output currents, e.g. low-power amplifiers with class AB output stages, the transconductance will not be constant, however. The value will vary along with the output current, while the resistor value remains fixed. This makes exact zero cancellation impossible. Other classical techniques also lose their usefulness when the output current varies. It is for these types of circuits that Multipath Miller Zero Cancellation presents a robust solution.

Figure 6-13 shows the principle of operation of the Multipath

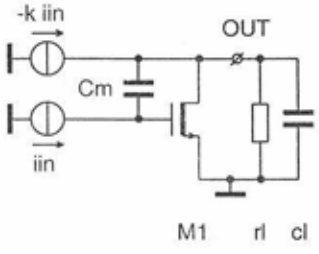


Fig. 6-13. *Principle of operation of the Multipath Miller Zero Cancellation technique.*

Miller Zero Cancellation technique. The circuit consists of transistor M_1

with Miller capacitor C_m. The input current source i_{in} drives the transistor. The current source $-k \cdot i_{in}$ provides the parallel path. Because this current is in the opposite phase to i_{in}, the current through the parallel path counteracts the current feeding through the Miller capacitor. We can write the transfer function of the circuit in Fig. 6-13 as

$$\frac{u_{out}}{i_{in}} = \frac{-g_{m1} + sC_m - ks\,(C_m + c_{gs})}{s\,(s\,(c_l C_m + c_l c_{gs} + C_m c_{gs}) + C_m g_{m1})} \tag{6-24}$$

The constant k in Eq. 6-24 determines to what extent the parallel path is active. The absence of k in the denominator demonstrates that the parallel path has no influence on the positions of the poles. Looking at the numerator, we can determine the effect of the parallel path on the zero. Setting $k = 1$ leads to the following zero position

$$z = -\frac{g_{m1}}{c_{gs}} \tag{6-25}$$

As the equation indicates, the zero changes position from the right to the left half plane. Since Eq. 6-25 exactly equals the *GB* product of the output transistor, in all practical applications the newly placed zero will lie outside the passband. Moreover, in the left half of the *s*-plane the zero increases the phase margin of the amplifier instead of worsening it.

When k equals one, a simple circuit suffices to generate the opposite phase current. Figure 6-14 presents the concept. The circuit bases itself on the two-stage Miller compensated amplifier. Transistor T_{2a} provides the Multipath input stage. The current in this parallel path and the drain current of transistor T_{2b} have the same magnitude but are in opposite phases. Since the drain current of T_{2b} flows into the Miller capacitor C_m for the most part, the current in the parallel path compensates for this feedthrough. The arrow indicates how the transistor T_{2b}, transistor T_{2a} and the Miller capacitor constitute a closed current loop. The loop cancels the Right Half Plane zero, since it effectively isolates the output terminal from the currents of the input stage. In Section 7.2 a CMOS implementation of the Multipath Miller Zero Cancellation technique to remove the RHP zero is demonstrated.

Setting k to a value higher than one opens the possibility to use the Multipath Miller Zero Cancellation scheme to eliminate a non-dominant pole in the transfer of the amplifier. This requires moving the Right Half

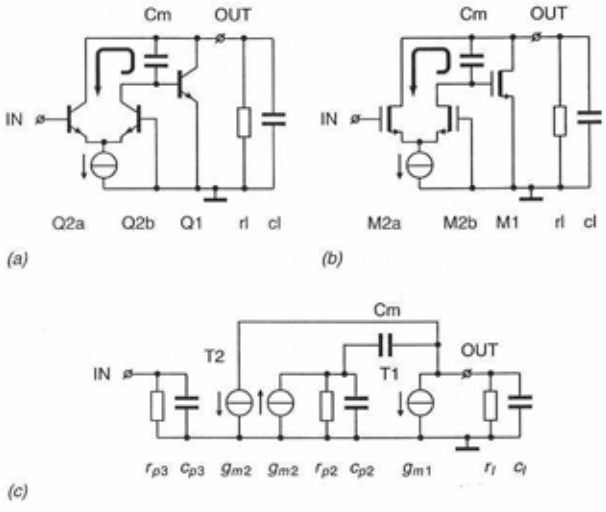


Fig. 6-14. *Implementation of Multipath Miller Zero Cancellation. The arrows indicate the current loops closed by the Multipath input transistors* M_{2a} *and* Q_{2a}.

Plane zero to the left half plane, exactly on top of the pole p_1' in Fig. 6-12. In other words, the constant k must have such a value that the zero originating from the numerator of Eq. 6-24 cancels the pole p_1', which was given before by Eq. 4-13. The resulting expressing for k, assuming that $c_{gs} \ll C_m$, is

$$k = 1 + \frac{c_l}{C_m} \tag{6-26}$$

According to Eq. 6-26, the ratio of the signal currents in the normal and the direct path depends on the ratio of the load capacitance c_l and the Miller capacitor C_m. The dependency on the load capacitance is a drawback in general purpose applications where this capacitance is not known in advance. For on-chip amplifiers with a fixed load, however, this limitation does not apply and Multipath Miller Cancellation is an effective measure to remove the second order pole. The result is a two-stage amplifier

with a single-pole frequency response similar to that of the one-stage amplifier of Fig. 4-2. The absence of a secondary pole is an important advantage in high-Q active filters [24]. The bandwidth of such circuits is very much affected by the presence of the secondary pole. In many cases removing the pole will lead to a bandwidth improvement of one or more orders of magnitude.

The non-unity factor k that is necessary to remove the secondary pole prohibits the use of the opposite branch of the input stage to generate the Multipath signal current. Instead, a second input stage is required. This approach was followed earlier to obtain the Multipath Nested Miller compensation circuits of Section 6.1. Scaling the transconductances of the two input stages enables setting of the factor k. We can achieve scaling by adjusting the tail currents of the input stages and, for CMOS, the W/L ratio of the input stage transistors. Figure 6-15 shows the bipolar and CMOS

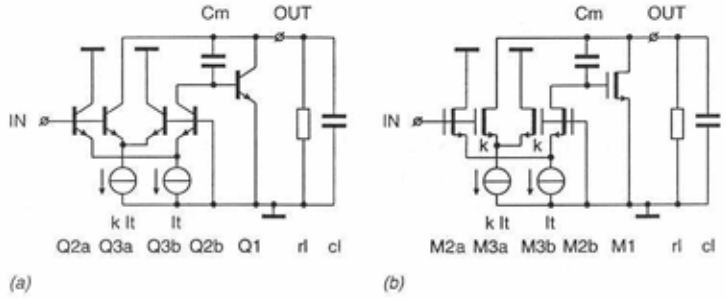


Fig. 6-15. *Bipolar and CMOS implementation of the Multipath Miller Zero Cancellation technique used for removal of the secondary pole.*

circuits. In this figure, the transistors M_{3a} and M_{3b} have a k times greater W/L ratio than the transistors M_{2a} and M_{2b}.

6.4.1. Multipath Miller Zero Cancellation for more than two stages

We can also apply the Multipath Miller Zero Cancellation technique to any of the Nested Miller compensation strategies of the previous chapter. We can add, for instance, Multipath signal paths that directly drive the output

terminal to a Nested Miller compensated circuit. Such an addition can lead to a significant bandwidth improvement over the original circuit. Figure 6-16 shows the topology. For simplicity, the idealized transconduc-

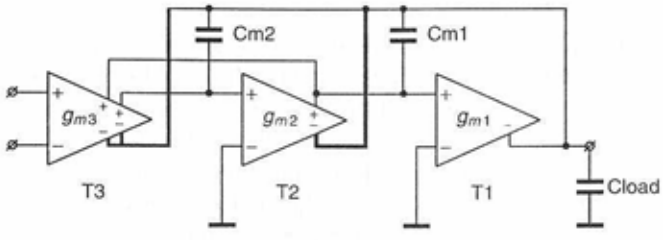


Fig. 6-16. *Multipath input stage added to a three-stage Nested Miller compensated amplifier. The bold lines indicate the Multipath current paths.*

tors represent the amplifier stages in this figure. The input stage T_3 and the intermediate stage T_2 are of the differential type and therefore generate opposite phase output currents. The implementation of Multipath Miller Zero cancellation follows the rule that an opposite phase current supplied to the output must compensate for each current driving a Miller capacitor. The bold lines indicate the signal paths that accomplish this.

6.5. Multipath Conditionally Stable compensation

All of the Multipath techniques of this chapter were based on the requirement that the realized amplifiers should be absolutely stable. Although this is of prime importance for general purpose building blocks such as opamps, we can obtain better results for some applications when the condition of absolute stability is left. Chapter 3.2.3 showed that conditionally stable amplifiers fundamentally have better properties with respect to the loop gain over a limited frequency band. A specific area where this is an advantage is that of the audio amplifiers. These types of circuits require a

bandwidth of only 20kHz, paired to a high accuracy over that bandwidth. Adhering to absolute stability, the combination of bandwidth and accuracy is difficult to fulfill. An 20dB/dec amplifier, having a unity-gain frequency of 2MHz, can sustain a flat open-loop gain characteristic of 100dB only up to about 20Hz! A conditionally stable frequency compensation method eases the requirements. In the case of a third order dominant pole, the width of the useful band increases to approximately 40kHz. The Bode plot of Fig. 6-17 illustrates this. At frequency ω_0 = 40kHz three poles are

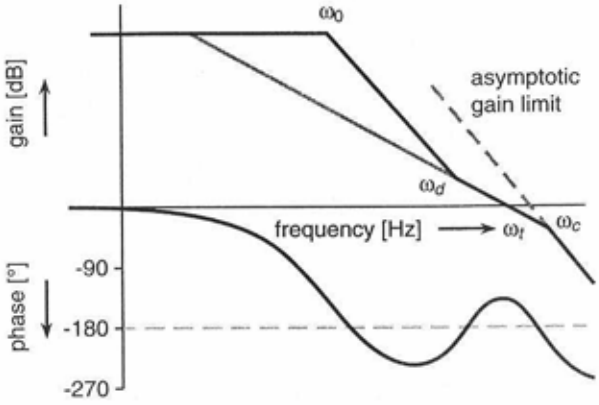


Fig. 6-17. *Bode plot of a conditionally stable amplifier.*

present, resulting in a 60dB/dec roll-off. Two zeros are located at the frequency ω_d, just below the unity-gain frequency. The slope reduces because of the zeros and the curve intersects with the 0dB axis at a rate of 20dB/dec. The steep slope beyond the useful band (inversely proportional to the third power of the frequency) is what causes the drastically improved useful bandwidth.

The Multipath technique is well suited for realizing a conditionally stable amplifier. The independent path that the Multipath input stage provides is liable for controlling the position of the 20dB/dec section of the frequency response. Since this part of the curve is decisive for the stability of the amplifier, conditionally stable amplifiers that are robust against

parameter variations become feasible with this technique. The demands on the low frequency gain part of a Multipath Conditionally Stable amplifier are relatively mild. Apart from sufficient gain over the useful band, the only restriction is that above the takeover frequency, the gain of this part of the circuit must drop below that of the Multipath section. The gain difference guarantees a smooth takeover by the 20dB/dec section.

Figure 6-18 presents the circuit of a Multipath Conditionally Stable

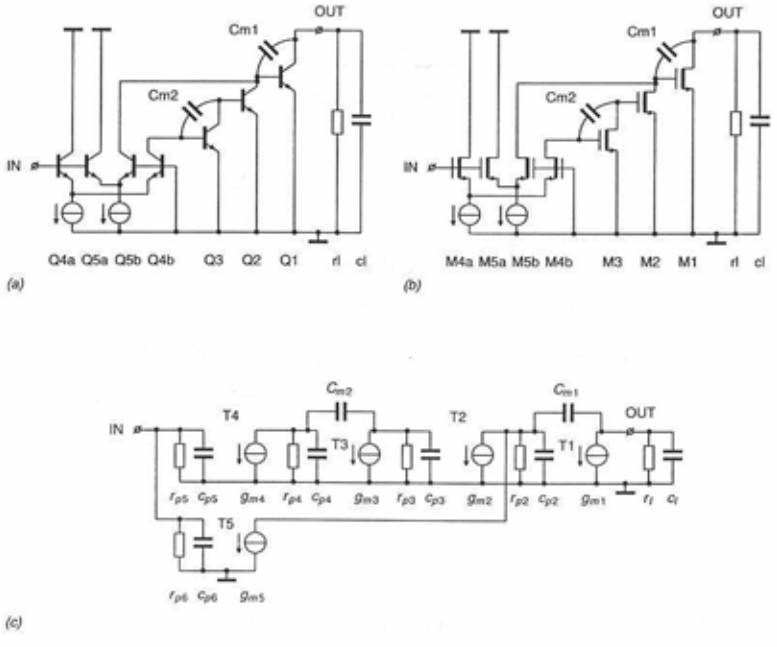


Fig. 6-18. *Four-stage Multipath Conditionally Stable amplifier.*

amplifier. It consists of a four-stage gain part comprising T_1 through T_4, and the Multipath input stage T_5 that directly drives the output stage. The Multipath input stage T_5 and the output stage T_1, as before, set the high frequency part of the frequency response. These constitute a two-stage Miller compensated amplifier with the appropriate 20dB/dec roll-off. The

transistors T_1 through T_4 realize the low frequency part. The two Miller capacitors C_{m1} and C_{m2} restrict the gain at high frequencies in order to allow the Multipath section to take over. We can think of the transistor chain as a cascade of two-stage Miller compensated amplifiers. The unity-gain frequencies of these parts are g_{m2}/C_{m1} and g_{m4}/C_{m2}. If both unity-gain frequencies are set to equal the takeover frequency ω_d, the gain of the transistor chain at this frequency will be exactly 0dB. This is below the gain of the Multipath section and therefore prevents the low frequency part to interfere with the high frequency part of the amplifier. The corresponding dimensioning equation is

$$\frac{g_{m2}}{C_{m1}} = \frac{g_{m4}}{C_{m2}} = \omega_d \tag{6-27}$$

The Multipath input stage T_5 and the output stage T_1 determine the unity-gain frequency of the amplifier. The unity-gain frequency therefore obeys

$$\omega_t = \frac{g_{m5}}{C_{m1}} \tag{6-28}$$

For a sufficient phase margin, the unity-gain frequency must equal half the frequency of the bandwidth limiting pole p_1':

$$\omega_t = \frac{1}{2} p_1' \tag{6-29}$$

The bandwidth limiting pole p_1' is given as before by g_{m1}/c_l. The takeover frequency ω_d must lie well below the unity-gain frequency ω_t. A proper choice is

$$\omega_d \approx \frac{\omega_t}{5} \tag{6-30}$$

6.5.1. Conditionally stable compensation for more than four stages

The two cascaded simple Miller amplifiers in the gain path of the circuit of Fig. 6-18 result in a 40dB/dec roll-off up to the takeover frequency ω_d. Adding more stages to the gain path not only improves the DC gain, but

also increases the rate at which the gain plot descents. The steeper slope further relaxes the bandwidth requirement of the circuit compared to an absolutely stable solution. Amplifiers with any even number of stages are feasible in this fashion. Figure 6-19 shows the simplified block diagram of

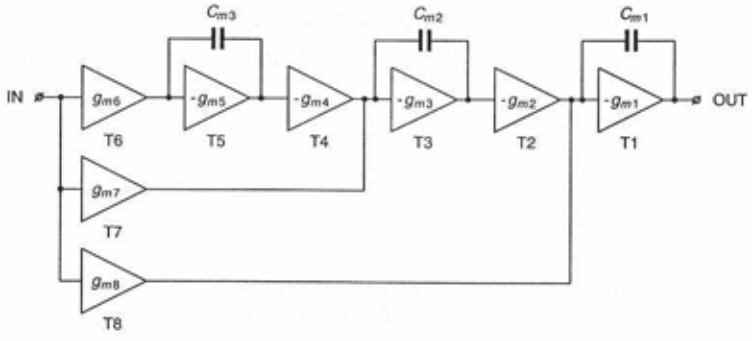


Fig. 6-19. *Representation with idealized transconductors of a six-stage conditionally stable amplifier.*

a six-stage conditionally stable amplifier. The idealized transconductors model the transistors. The gain part consists of the transistors T_1 through T_6, while T_8, directly driving the output stage, furnishes the Multipath input stage. T_7 is an additional Multipath input stage, the function of which we will investigate shortly.

The dimensioning equations follow from similar investigations as those into the four-stage circuit of Fig. 6-18. The expression for the dimensioning of the gain path is

$$\frac{g_{m2}}{C_{m1}} = \frac{g_{m4}}{C_{m2}} = \frac{g_{m6}}{C_{m3}} = \omega_d \tag{6-31}$$

while for the Multipath input stage T_8 holds

$$\omega_t = \frac{g_{m8}}{C_{m1}} = \frac{1}{2} p_1' \tag{6-32}$$

Furthermore, the takeover frequency ω_d must obey

$$\omega_d = \frac{\omega_t}{5} \tag{6-33}$$

The additional Multipath input stage T_7 in the circuit of Fig. 6-19 prevents destructive interference at the takeover frequency ω_d. Without this stage, at some point close to the takeover frequency ω_d, the phase difference between the signals going through the high frequency and the gain paths will amount to exactly 180°. Since the magnitudes of the signals are nearly equal at this point, the signals will cancel out. The result is a sharp dip in the frequency response at ω_d. The introduction of T_7 eliminates this dip by keeping the phase shift of the six-stage gain section within bounds. The second Multipath input stage is correctly dimensioned when

$$\frac{g_{m7}}{C_{m2}} = \omega_d \tag{6-34}$$

6.6. Conclusions

This chapter presented the Multipath technique to overcome the bandwidth reduction of Nested and Hybrid Nested Miller compensation. The technique theoretically doubles the attainable bandwidth of the Miller compensated amplifiers without a power penalty. We can independently configure the Multipath input stage from the remainder of the amplifier. This is an important advantage over the more traditional capacitive bypassing of PNPs or other bottleneck components. This classical technique generally results in strong and inevitable pole-zero doublets. The doublets of the Multipath technique can be well controlled and closely spaced, however. This is due to the favorable matching, which depends on transconductance and capacitor ratios only. Table 6-1 gives an overview of

Table 6-1. *Summarized properties of the Multipath and non-multipath compensation techniques.*

	parallel	NMC	RNMC	HNMC	MNMC	MHNMC
bandwidth/power	++	o	+	+	+	++
capacitor values	--	++	++	++	++	++
ruggedness	--	++	-	++	++	++
pole-zero doublets	--	++	++	++	o	o
need for cascodes	-	++	++	++	++	++
distortion	-	++	-	+	+	+
low supply voltage	++	+	+	++	+	++

++=excellent, +=good, o=moderate, -=poor, --=very poor

the properties of the Multipath Nested Miller compensation technique and the Multipath Hybrid Nested Miller technique. For comparison, the table also repeats the results of the compensation methods of Chapter 5.

A special type of Multipath circuit appears when the Multipath signal is fed directly into the output terminal of the amplifier. The resulting Multipath Miller Zero Cancellation technique is capable of counteracting the Right Half Plane (RHP) zero that occurs in Miller compensated amplifiers. Especially in CMOS circuits the zero in the right half of the complex *s*-plane can be a major factor that limits the bandwidth. The Multipath Miller Zero Cancellation technique removes the RHP zero by exact compensation of the direct current through the Miller capacitor that causes the zero. This fundamental approach again leads to the inherent virtue of the Multipath technique, being robustness against parameter variations. Further, no pole-zero doublets are introduced in the frequency response. Even in class AB amplifiers with largely varying output transistor currents we can obtain accurate cancellation of the bandwidth impairing zero, whereas traditional techniques would fail under these circumstances.

A final application of the Multipath technique discussed in this chapter is the Multipath Conditionally Stable amplifier. Using a Multipath

input stage to bypass a cascade of basic two-stage Miller compensated amplifiers, we obtain a circuit with a well-controlled conditionally stable frequency response. Since the 20dB/dec section that determines the stability of a conditionally stable amplifier is set by the Multipath input stage and the Miller compensated output stage alone, stability is guaranteed for a wide range of process and load variations. A conditionally stable frequency response drastically increases the useful bandwidth of such specialized applications as audio amplifiers and filters. With the Multipath Conditionally Stable technique this improvement goes without an increase of the consumed power or the die area occupied by the amplifier.

7 *Realizations*

To evaluate their specific properties, this chapter gives several illustrative implementations of the various compensation methods presented in the preceding chapters. We will investigate ten amplifier realizations, distributed over a total of five chips. Four of the circuits are fabricated in a bipolar technology, while six of them are CMOS based.

The first circuits addressed, in Section 7.1, are two bipolar operational amplifiers with Nested (NMC) and Multipath Nested Miller compensation (MNMC) respectively [14]. The first opamp pairs a unity-gain frequency of 60MHz and a D.C. gain of 100dB. The second, employing the Multipath technique to overcome the bandwidth reduction of Nested Miller compensation, attains a unity-gain bandwidth of 100MHz with the same level of gain. Both opamps embody an all-NPN signal path for the maximum obtainable bandwidth-to-power ratio.

Section 7.2 presents a Miller compensated CMOS opamp with Multipath Miller Zero Cancellation to remove the zero in the right half of the s-plane [23]. The section compares this circuit with three others: a reference circuit without removal of the Right Half plane zero, a circuit with a traditional series resistor to remove the RHP zero and finally an amplifier which has the Miller capacitor connected to the input of a cascode stage instead of the gate of the output transistor, which is another classical approach for counteracting the RHP zero.

Section 7.3 describes two programmable CMOS class AB operational amplifiers with Hybrid Nested Miller and Multipath Hybrid Nested Miller compensation, respectively [16]. The minimum supply voltage of these opamps is 1.5 V at a total supply current of 300 µA. An even lower supply voltage of 1.1 V is possible by programming down the total supply current. The section further addresses a bipolar opamp operating at a minimum supply voltage of 1 V, which uses the Multipath Hybrid Nested Miller compensation strategy.

Finally, Section 7.4 introduces a four-stage circuit with Multipath Conditionally Stable compensation. It is based on the bipolar Multipath Hybrid Nested Miller compensated opamp of Section 7.3. With a minor modification, this amplifier transforms into a circuit with a conditionally stable frequency characteristic, offering a sufficient phase margin only at a specific gain. This limitation results in a substantial improvement of the midband loop gain, without a power or die area penalty.

7.1. Precision operational amplifiers with NMC and MNMC

This section presents two bipolar operational amplifiers. The first uses Nested Miller compensation for obtaining a stable loop gain characteristic, the second Multipath Nested Miller compensation. The Multipath input stage of the latter increases the unity-gain frequency of the circuit from 60 MHz to 100 MHz. This closely corresponds to the theoretically expected doubling of the bandwidth. The phase margin of both amplifiers with a 100 pF load is 40°. The Nested Miller compensated opamp settles to 0.1% on a 1 V step in 70 ns, while the Multipath amplifier reduces this time to 60 ns.

7.1.1. Introduction

The Nested Miller and Multipath Nested Miller compensation techniques (Section 5.2 and Section 6.1) are very suited for realizing operational amplifiers with a wide unity-gain bandwidth and a high D.C. gain. These properties make the resulting circuits qualified for applications which require accurate signal processing. An example of such a demanding application is the filtering and buffering of the analog signal that appears at the D/A convertor's output of a digital audio system. The high resolution

of such systems, typically 16 bits, demands for a great precision and hence a high open-loop gain. Furthermore, minimizing the non-linearity calls for a fast amplifier to be able to follow the steep edges that characterize the output signal of the D/A convertor. Both the gain and speed demands are fulfilled by the Nested Miller compensation strategies. A third requirement that is satisfied by the compensation techniques is a low H.F. output impedance. Keeping the output impedance low is essential to reduce the occurrence of voltage spikes feeding directly through the feedback network to the output of the amplifier. If not counteracted, these spikes affect the overall linearity of the system. A low output impedance at high signal frequencies is guaranteed by the local feedback that the Miller capacitors of (Multipath) Nested Miller compensation provide across the output stage.

To take full benefit of the (Multipath) Nested Miller Compensation techniques, the operational amplifiers presented in this section are based on an all-NPN topology, allowing the maximum unity-gain bandwidth. Combined with an effective class AB control, the results represent leading edge high bandwidth precision operational amplifiers. The Nested Miller compensated circuit displays a unity-gain bandwidth of 60MHz with a 100pF load. The Multipath input stage of its counterpart raises the bandwidth to 100MHz under the same conditions. Both circuits provide a D.C. gain of 100dB and have a supply quiescent current of 9.5mA.

The next section addresses the circuits of the two realized opamps, including the all-NPN topology and class AB control. The experimental results are covered in Section 7.1.3. Finally, Section 7.1.4 gives the conclusions.

7.1.2. Circuit description

A. All-NPN Topology

Figure 7-1 is the simplified schematic of the opamp with Nested Miller compensation. The compensation capacitors are inserted according to the Nested Miller compensation scheme of Section 5.2. To assure a high bandwidth, only NPN transistors are present in the signal path. Consequently, in the push-pull output stage the use of an emitter follower for the push and an inverting amplifier for the pull transistor are mandatory. The emitter follower Q_{400} has the capacitor C_{p1} connected from its base to the

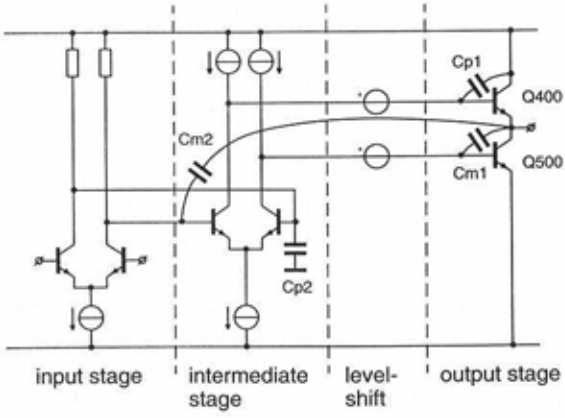


Fig. 7-1. *Simplified schematic of the Nested Miller compensated opamp.*

ground and the inverting amplifier Q_{500} the Miller capacitor C_{m1} from the base to the collector. Capacitors C_{m1} and C_{p1} have equal values. Surprisingly, when driven by a current signal both transistor configurations behave symmetrically [26]. Not only do they have the same transimpedance z_t, but also their output impedances z_{out} are equal. Because of the differential second stage in Fig. 7-1, the circuit requires the capacitor C_{p2} to balance out the outer Miller capacitor C_{m2}.

The levelshift circuit, depicted as a voltage source in Fig. 7-1, has the characteristics of an all-pass current network [26]. Figure 7-2 shows the circuit. For input currents there are two separate routes from input to output. Low frequency signals pass through the resistor and the PNP transistor, while for high frequencies the path is through the capacitor and the NPN current follower. The RC product of the resistor and the capacitor sets the cross-over frequency f_{RC}. As long as f_{RC} is lower than the f_{tp} of the PNP transistor, no pole-zero doublets occur, because no current disappears in the all-pass network. The bandwidth of the levelshift circuit in that case equals the NPN's f_{tn}. Noise considerations dictate the location of the levelshift in the circuit. Situating the levelshift directly following the input stage would have increased the noise level because of the noise contribu-

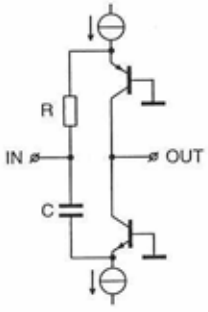


Fig. 7-2. *All-pass current network implementing the levelshift function.*

tion of the current sources in the all-pass current network. For this reason, the levelshift is located after the intermediate stage.

The Multipath Nested Miller compensation opamp is largely similar to the Nested Miller compensation opamp in Fig. 7-1. Added to the amplifier is the Multipath input stage, as Fig. 7-3 points out. This input stage is connected in parallel to the original input stage and directly drives the output transistors. Since the Multipath current path in this way bypasses the intermediated stage, we can theoretically expect a factor of two bandwidth improvement. We found this earlier in Section 6.1, giving the elaborated analysis of Multipath Nested Miller compensation.

B. Class AB Control

The feedback class AB circuit controls the quiescent current of the output stage and prevents cut-off of the output devices by assuring a minimum quiescent current [27]. The control circuitry is independent from the signal path, as demonstrated in Fig. 7-4. The differential intermediate stages in Fig 7-1 and Fig. 7-3 supply the amplified input signal to the output transistors as two currents that are of the opposite phase. The class AB circuit, on the other hand, controls the biasing by supplying two in-phase currents through the parallel transistors Q_{601} and Q_{602}. The operation of the class AB control does not interfere with the output signal, because the in-phase

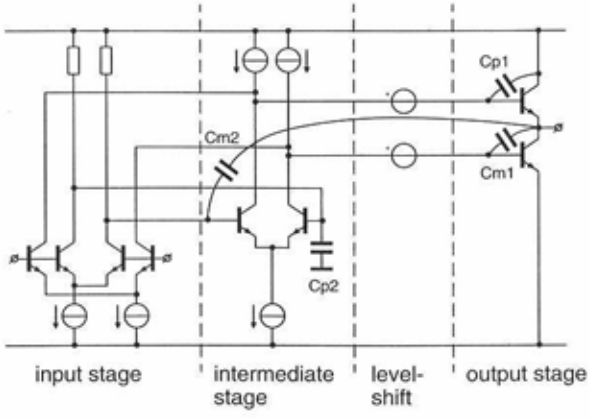


Fig. 7-3. *Simplified schematic of the Multipath Nested Miller compensated opamp.*

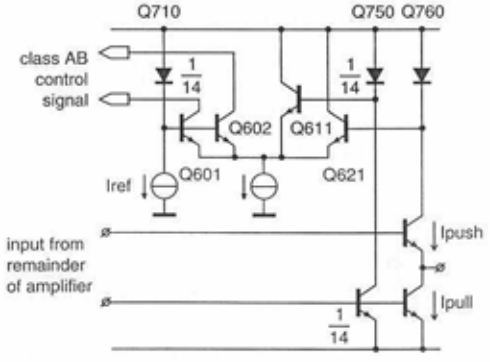


Fig. 7-4. *The class AB control circuit.*

currents cancel in the output stage, i.e. when one of the transistors is con-

trolled at its quiescent current due to a high output current condition, the class AB control doubles the driving of the other transistor.

The class AB circuit incorporates a combined error amplifier and decision gate. The decision gate, comprising the emitter followers Q_{611} and Q_{621}, selects the smaller of the two transistor currents in the output stage. The diodes Q_{750} and Q_{760} derive the input voltages of the decision gate from the push and pull transistor currents, respectively. The decision gate transistors Q_{611} and Q_{621} function as two emitter followers, but only one of the devices, corresponding to the lowest output current, becomes properly biased. This active emitter follower transfers its input voltage to the common emitter node of the error amplifier. The error amplifier consists of the decision gate together with the two transistors Q_{601} and Q_{602}. The reference voltage of the error amplifier is determined current I_{ref} across diode Q_{710}. Figure 7-5 shows the class AB characteristic. The current I_{ref} and the emitter ratios of the transistors set the quiescent current. The minimum value is limited to half the quiescent current.

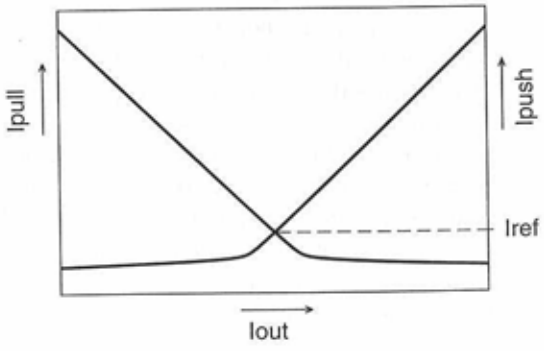


Fig. 7-5. *Class AB characteristic. Neither one of the output transistor currents ever becomes zero, preventing cut-off of the devices.*

C. Total Schematic

Figure 7-6 shows the total schematic of the Nested Miller compensated operational amplifier. The circuit diagram includes the biasing and level shift elements. The actual circuit uses Darlington transistors in the output stage to improve the gain. The PTAT current source consisting of Q_{800}-Q_{850} generates the bias current [28]. Resistor R_{801} initiates a small current in the right-hand branch. Because of the cross-coupled structure, the magnitude of the current is of no consequence to the PTAT output current and start-up is guaranteed. The PTAT current is 100µA at room temperature. The quiescent current of the output transistors is set to approximately 4.5mA.

The Multipath Nested Miller compensation opamp (Fig. 7-7) is, apart from the additional input stage with Q_{105} and Q_{115}, comparable to the operational amplifier without the Multipath technique. According to Section 6.1, which gives the dimensioning equations of Multipath Nested Miller compensation, limiting the bandwidth reduction due to the outer Miller loop demands the transconductance of the intermediate stage to be reduced. This is done in the circuit of Fig. 7-7 by lowering the tail current of Q_{200} and Q_{210} and insertion of the degeneration resistors R_{200} and R_{210}. The lower tail current of the intermediate stage ensures that, despite the extra input stage of the Multipath Nested Miller compensation opamp, the total supply currents of the two amplifiers are equal. The frequency of the pole-zero doublet ω_d, according to Eq. 6-9, is 15MHz. The expected pole-zero mismatch is approximately 1%, due to the inevitable mismatch of the transconductances of the input transistors Q_{100}/Q_{110} and Q_{105}/Q_{115} and the mismatch of the capacitors C_{m1}/C_{p1} and C_{m2}/C_{p2}.

7.1.3. Realizations and experimental results

The chips were fabricated in a 3GHz NPN f_t bipolar I.C. process. To be able to drive a 100pF load with a unity-gain bandwidth of 100MHz, two output pins and corresponding bond wires decouple the load from the feedback loop. This is shown by Fig. 7-8. The output pins act as current and voltage terminals and isolate the driving of the load from the feedback path. Without this measure the load capacitance and inductance of the bond wires would introduce a pair of complex poles in the feedback loop, resulting in instability of the circuit.

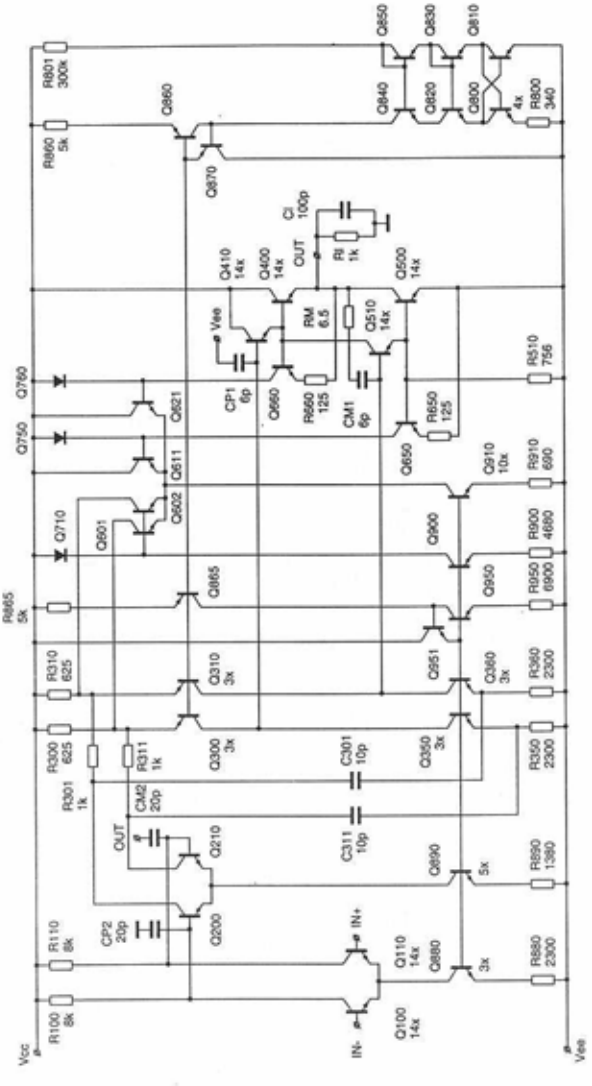


Fig. 7-6. *Total schematic of the Nested Miller compensated amplifier.*

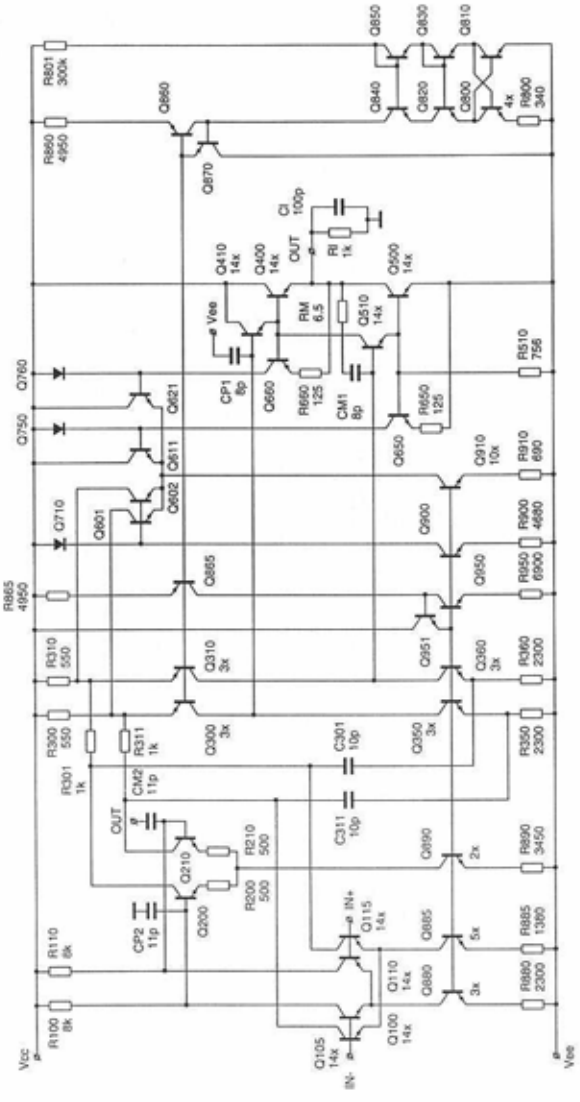


Fig. 7-7. *Total schematic of the Multipath Nested Miller compensated opamp.*

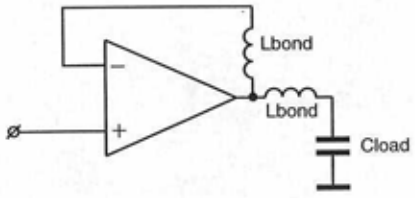


Fig. 7-8. *Voltage and current terminals at the output to separate the feedback loop from the load capacitance.*

Figure 7-9 clearly shows the two output bonding wires. Figure 7-9*a* is a photomicrograph of the Nested Miller compensation and Fig. 7-9*b* of the Multipath Nested Miller compensation opamp. The die area of both amplifiers is equal. The extra die area needed for the larger compensation capacitors of the Nested Miller compensated circuit of Fig. 7-9*a* about cancels out the area required for the Multipath input stage in Fig. 7-9*b*. The capacitors have higher values due to the lower bandwidth of the opamp. The chips both measure 1.2mm x 1.5mm^2.

Figure 7-10 shows the Bode plots of the opamps. The Nested Miller compensation opamp has a unity-gain bandwidth of 60MHz with a phase margin of 40°. The unity-gain bandwidth of the Multipath Nested Miller compensation opamp is 100MHz, with a phase margin of slightly less than 40°. Both opamps face a load of a 100pF capacitor parallel to a 1kΩ resistor, as is the case in the following measurements.

Figure 7-11 gives the slew response of the opamps on an input step of 1volt. Since the input stages are not degenerated by emitter resistors, the unity-gain bandwidths of the amplifiers determine the respective slew rates. The slew rate of the Nested Miller compensation opamp is 20V/μs; that of the Multipath Nested Miller compensation opamp is 35V/μs.

Figure 7-12 presents the small-signal settling of the amplifiers. The input step is 100mV. The 0.1% settling time corresponding to the top trace (NMC) is 40ns. The step response very much resembles the desired Butterworth curve. As the bottom plot of Fig. 7-12 indicates, a small slow-settling component is detectable in the step response of the Multipath Nested

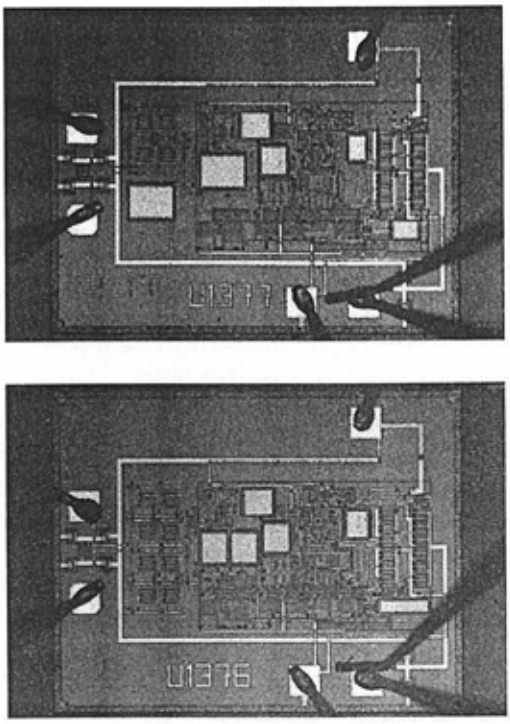
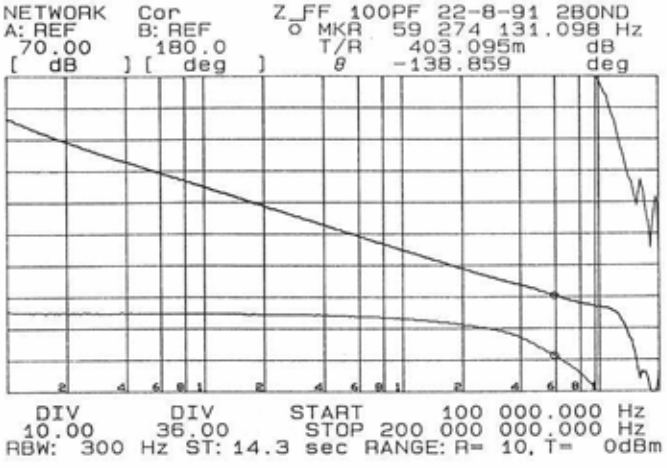


Fig. 7-9. *Photomicrographs of* (top) *the Nested Miller compensated opamp and* (bottom) *the Multipath Nested Miller compensated opamp.*

Miller compensation amplifier. The doublet spacing corresponding to the slow-settling component is approximately 5%. The 0.1% settling time is 50ns. The contribution of the slow-settling component to the total settling time becomes relatively less important for large input steps. Because slewing of the opamp governs most of the large signal step response, the Multipath Nested Miller compensation settles faster to 0.1% after a 1V input step than its Nested Miller compensation counterpart. The plots in Fig. 7-13 confirm this. Settling times are 70ns and 60ns, respectively. The



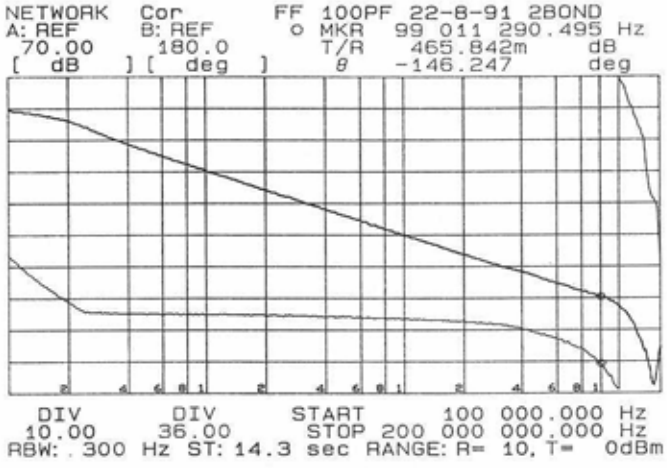


Fig. 7-10. *Bode plots of* (top) *the Nested Miller compensated opamp and* (bottom) *the Multipath Nested Miller compensated opamp.*

difference between the two circuits will be even more pronounced with

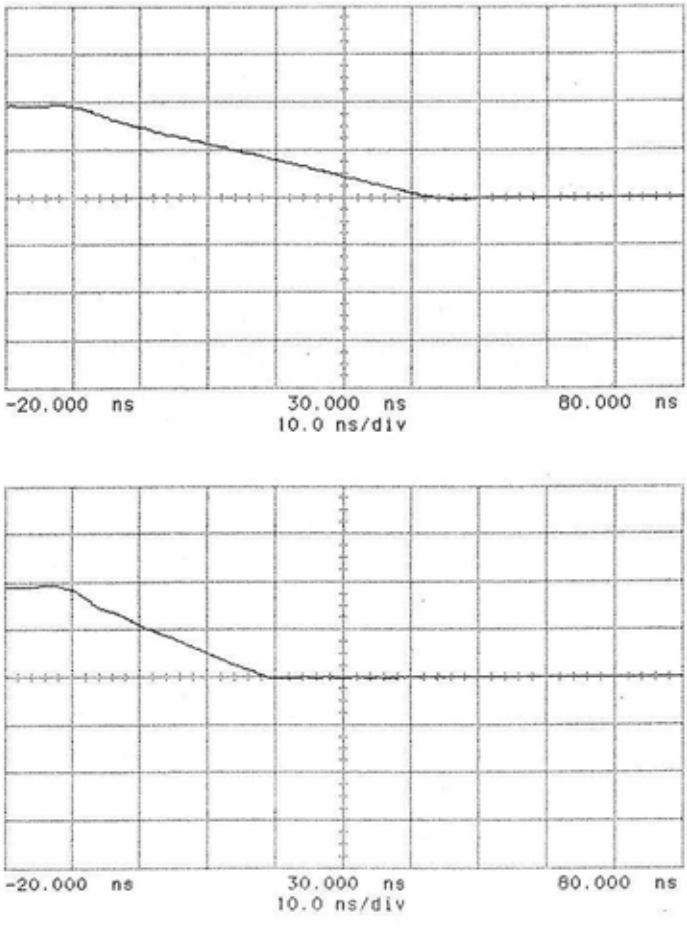


Fig. 7-11. *Slew response of* (top) *the Nested Miller compensated opamp and* (bottom) *the Multipath Nested Miller compensated opamp.*

higher step voltages.

The final plot concerning the two opamps is shown in Fig. 7-14, which represents the equivalent input referred voltage noise of the Nested

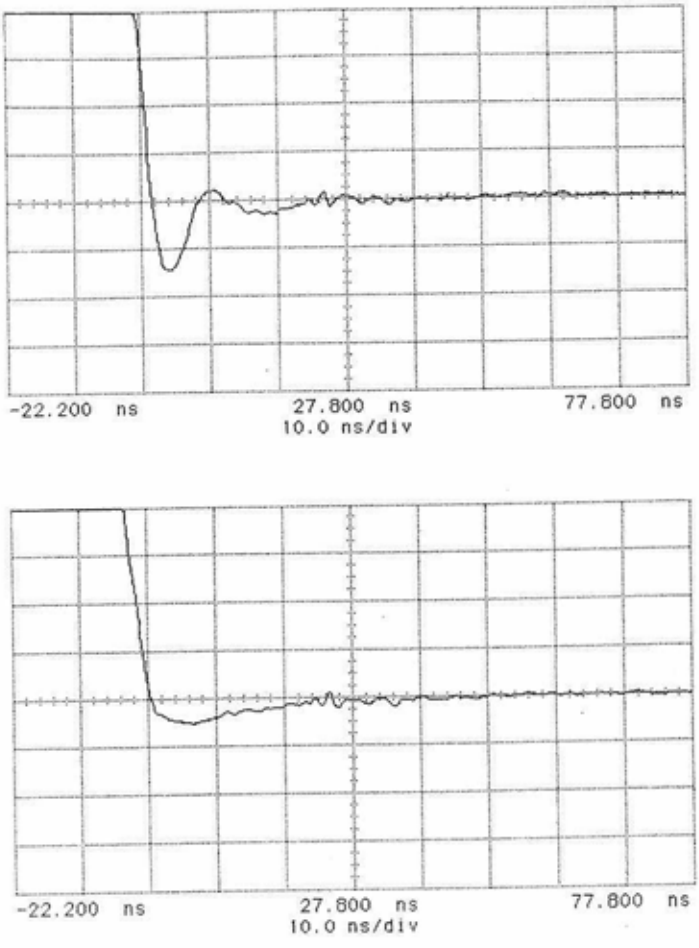


Fig. 7-12. *Small-signal settling of* (top) *the Nested Miller compensated opamp and* (bottom) *the Multipath Nested Miller compensated opamp.*

Miller compensation (*top*) and Multipath Nested Miller compensation (*bottom*) opamp. The voltage noise of the opamps is $2\text{nV}/\sqrt{\text{Hz}}$. For frequencies above 15MHz (the cross-over frequency of the Multipath input

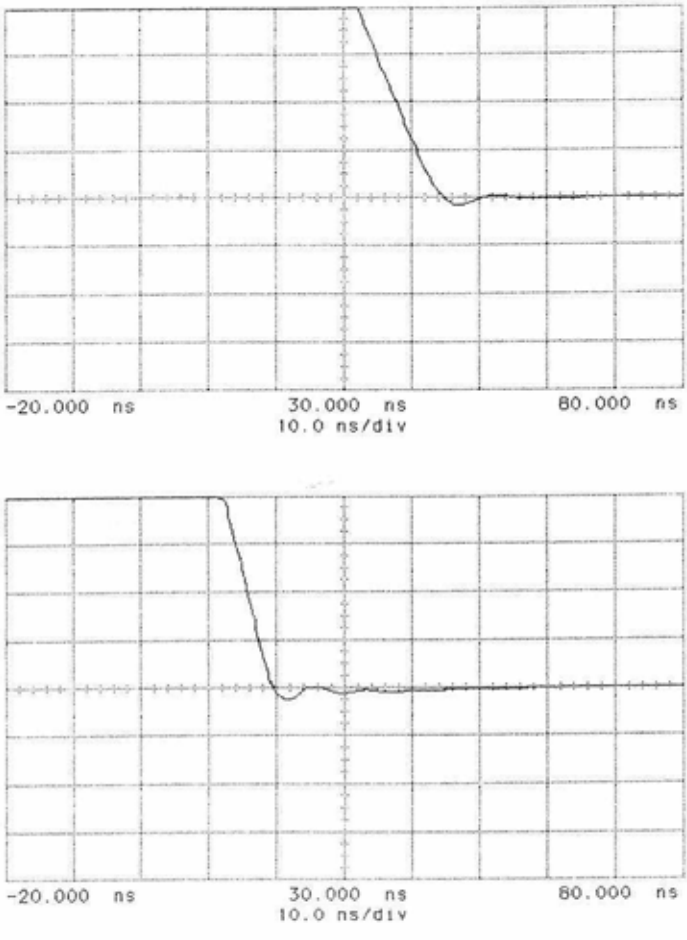


Fig. 7-13. *Large-signal settling of* (top) *the Nested Miller compensated opamp and* (bottom) *the Multipath Nested Miller compensated opamp.*

stage), the noise of the Multipath Nested Miller compensation opamps rises slightly. Because the intermediate stage is not active in this frequency

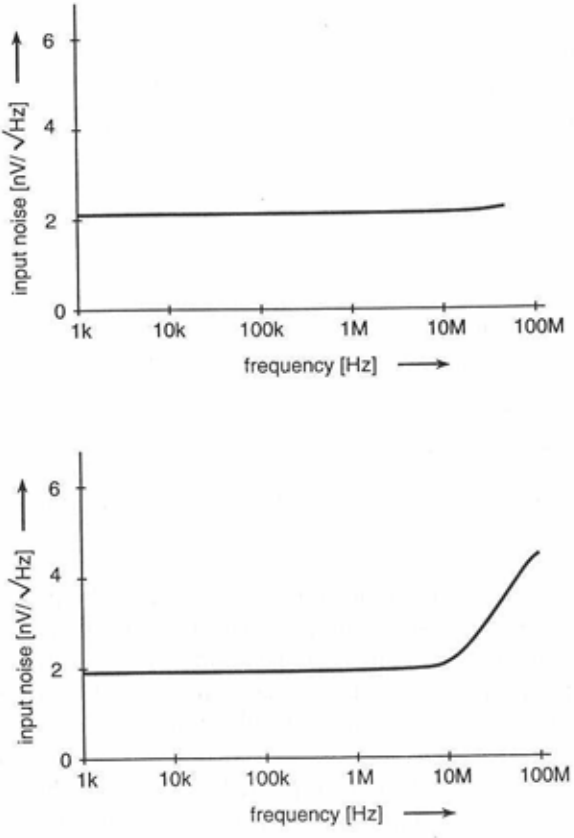


Fig. 7-14. *Equivalent input voltage noise of* (top) *the Nested Miller compensated opamp and* (bottom) *the Multipath Nested Miller compensated opamp.*

region, noise of the levelshift circuits contributes to the input noise through the Multipath input stage. The total input noise does not exceed $4nV/\sqrt{Hz}$ over the entire bandwidth of the Multipath Nested Miller compensation opamp, however. Table 7-1 summarizes the measurement results

Table 7-1. *Measurement results of the bipolar precision opamps.*

Parameter	NMC	MNMC
Unity-gain Frequency	60MHz	100MHz
D.C. Gain	100dB	100dB
Settling Time (0.1%) ($V_s = 0.1$ V)	40nS	50nS
Settling Time (0.1%) ($V_s = 1$ V)	70nS	60nS
Input Noise Voltage (1kHz)	2nV/√Hz	2nV/√Hz
(50MHz, 100MHz)	2nV/√Hz	4nV/√Hz
Output Impedance (50MHz)	10Ω	10Ω
Maximum Output Current	50 mA	50 mA
Supply Current	9.5 mA	9.5 mA
Bandwidth-to-Power Ratio	0.8GHz/W	1.3GHz/W

$T_A = 25°$ C, $V_{cc} = 8$ V, $C_l = 100$pF, $R_l = 1$kΩ except where indicated.

of the Nested Miller compensated (NMC) and Multipath Nested Miller compensated (MNMC) circuit. A salient feature of both Nested Miller compensation methods is the low output impedance at high frequencies which these can realize. This is due to the local feedback with the Miller capacitors across the output stage. The low output impedance —10Ω at 50MHz according to Table 7-1— is of prime importance in digital audio applications. When the opamp is the first to follow the digital-to-analog convertor in digital audio systems, a low impedance at high frequencies is necessary to handle the sharp pulse edges coming from the convertor. A high output impedance can become a major source of non-linear distortion in this situation [25].

7.1.4. Conclusions

Nested Miller Compensation and Multipath Nested Miller Compensation open the door to the next generation of fast and accurate operational amplifiers. The presented Multipath Nested Miller compensation amplifier drives a 100pF load with a unity-gain bandwidth of 100MHz. The Multipath technique introduces a well controlled pole-zero doublet, the match-

ing of which depends on capacitor and current ratios only. On a test chip the doublet spacing was 5%. The unity-gain bandwidth of the Nested Miller compensation opamp is 60MHz. The gain of both opamps is 100dB.

7.2. Opamp with Multipath Miller Zero Cancellation

This section presents a Miller compensated operational amplifier with the Multipath Miller Zero Cancellation (MMZC) technique to remove the Right Half Plane (RHP) zero. For comparison, four different CMOS amplifiers have been integrated: a reference circuit without RHP zero removal, two circuits with traditional resistor and cascode removal techniques and finally the circuit with Multipath Miller Zero Cancellation.

Measurements confirm that Multipath Miller Zero Cancellation increases the bandwidth of a micro-amp Miller compensated output stage from 3MHz to 7MHz, while maintaining the robustness necessary for class AB operation. The traditional methods for RHP zero removal fail under the same conditions and give rise to peaking (the 'output bump')

7.2.1. Introduction

The Miller technique, widely used for frequency compensation of opamps, is robust against parameter variations. Since only one transistor is present in the feedback loop with the Miller capacitor, the loop is guaranteed to be stable under any operating condition. This dynamic stability is crucial in class AB output stages with high quiescent-to-maximum current ratios, as encountered in low-power opamps.

A disadvantage of the Miller technique is the zero that appears in the right half of the complex s-plane. This Right Half Plane (RHP) zero severely degrades the phase margin of the amplifier and thus the obtainable bandwidth. The zero is caused by the direct path from input to output which the Miller capacitor creates for high frequencies. Since current that follows this path is in the opposite phase to the output current of the transistor, it reverses the sign of the open-loop gain at high frequencies. This sign reversal, in its turn, can lead to instability of the closed-loop circuit.

In the past, several measures to counteract the RHP zero have been proposed. A classic solution is inserting a resistor R_m in series with the

Miller capacitor [19], as shown in Fig. 7-15. For exact cancellation of the

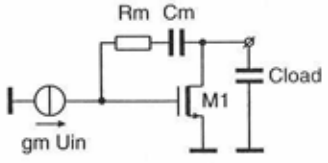


Fig. 7-15. *Tradition methods for RHP zero removal: series resistor.*

zero, the reciprocal of the series resistor R_m must track the transconductance g_m of the output transistor. This condition is difficult to meet when the current through the transistor varies. In this situation, which is typical for class AB output stages, we can only optimize the resistor for one value of the output current, e.g. the quiescent current. At higher output currents, the matching will become poorer and instability of the overall circuit is an actual danger.

Figure 7-16 shows other propositions to eliminate the RHP zero.

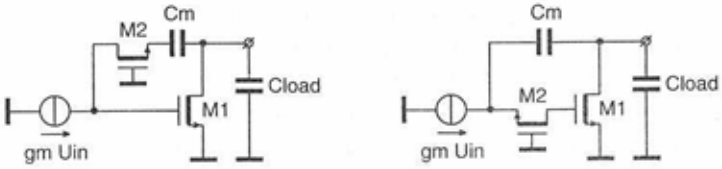


Fig. 7-16. *Tradition methods for RHP zero removal:* (left) *current buffer and* (right) *cascode stage.*

The circuit on the left employs a current buffer in the Miller capacitor branch [29], while the circuit on the right includes a cascode stage to which the Miller capacitor is connected [30]. Both added transistor stages prevent the input current from going directly through the Miller capacitor. Their applicability is also restricted to 'tame' output stages, however, since the dominant poles introduced by the additional active elements give rise

to peaking at high currents. This effect is known as the 'output bump' [11] and hampers the use of the current buffer and cascode circuit in class AB output stages with a large dynamic range.

In the next section we will encounter the Multipath Miller Zero cancellation technique to remover the RHP zero independently from the output current. Section 7.2.3 addresses the realizations.

7.2.2. Multipath Miller Zero Cancellation

The traditional solutions for RHP zero removal intend to *obstruct* the direct path through the Miller capacitor. Multipath Miller Zero Cancellation, conversely, counteracts the RHP zero by a parallel path that *compensates* for the direct feed-through. The simplified schematic of Fig. 7-17

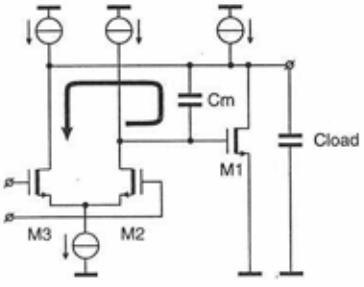


Fig. 7-17. *Simplified Schematic of the amplifier with Multipath Miller Zero Cancellation.*

clarifies the principle of operation of Multipath Miller Zero Cancellation. The circuit comprises differential pair M_2/M_3 and output transistor M_1 with Miller capacitor C_m. Transistor M_2, one half of the differential pair, drives the output transistor M_1. The other half, with transistor M_3, provides the parallel path. Because the drain current of M_3 has exactly the same magnitude as the drain current of M_2 but is of the opposite phase, the current through the parallel path counteracts exactly the current feeding through the Miller capacitor. The arrow indicates the closed current loop which the transistors M_2 and M_3 of the differential pair together with Miller capacitor C_m constitute. The current loop isolates the load from the

input transistors. The current loop is independent of the transconductance of output transistor M_1. Multipath Miller Zero Cancellation therefore does not degrade the robustness typical to Miller compensation.

7.2.3. Realizations

Figure 7-18 shows the circuit diagram of the operational amplifier which

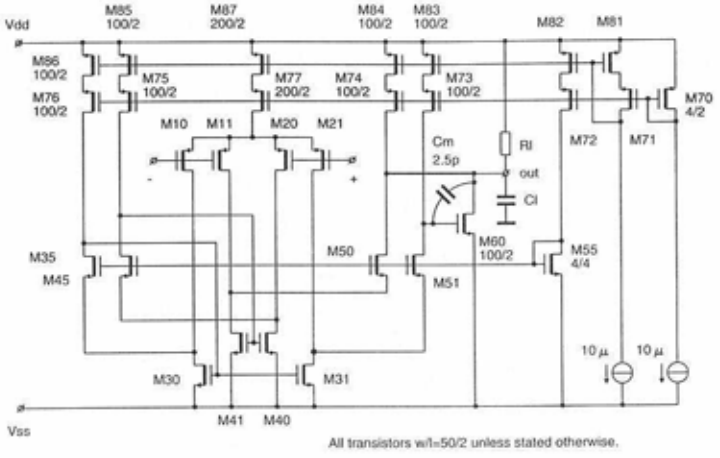


Fig. 7-18. *Circuit diagram of the opamp typifying Multipath Miller Zero Cancellation.*

employs the Multipath Miller Zero Cancellation technique. The amplifier consists of the input differential pair M_{10}/M_{21} and output stage M_{60}, separated by cascode transistor M_{51}. To include the negative supply rail in the common mode input range, the low voltage drop current mirror with M_{30}, M_{31} and M_{35} performs the differential-to-single conversion. The input of the current mirror is the drain of transistor M_{30}. The feedback loop through M_{35} forces the common gate voltage of M_{30} and M_{31} to such a value that the drain current of the input transistor M_{30} equals the input current of the mirror. Transistor M_{35} merely serves as a levelshift in this loop. Since the gate-source voltage of M_{31} equals that of M_{30}, the drain current of M_{31} —

the output current of the mirror— copies the input current accurately. The low voltage drop current mirror also helps reducing the input offset voltage by keeping the drains of the input differential pair transistors at equal potentials.

The additional differential pair M_{11}/M_{20}, current mirror M_{40}, M_{41} plus M_{45} and cascode stage M_{50} implement the Multipath Miller Zero Cancellation technique. The differential pair plus the current mirror generate the opposite phase current required to cancel the signal current feeding directly through the Miller capacitor. This current is passed through the cascode M_{50} for levelshift purposes. The bold line to the output terminal in Fig. 7-18 depicts how the Multipath current is supplied to the output terminal. The output stage of the opamp operates from the negative to the positive supply rail. Although the cascode transistor M_{50} enters the triode region of operation when the output voltage approaches the negative rail, this does not hamper the functionality of the Multipath Miller Zero Cancellation scheme. When in its triode region, the cascode device basically acts as a resistor. Since the differential pair M_{11}/M_{20} supplies a current with a high output impedance to this resistor, the Multipath compensation current will pass through virtually unobstructed.

As shown in Section 4.3, the maximum load capacitance of the opamp in Fig. 7-18 is proportional to the transconductance of the output transistor M_{60}. We can program the quiescent current, which sets the transconductance of the output transistor, by choosing a proper value for the load resistor R_l. Measurements have been performed with a fixed load capacitor of 3 pF and quiescent currents varying from 20 µA to 2 mA.

Apart from the Multipath Miller Zero Cancellation circuit in Fig. 7-18, two similar operational amplifiers with the traditional RHP zero removal methods and a reference circuit without any RHP removal have been integrated on one chip. The implemented traditional removal methods are those of Fig. 7-15, with a resistor of 3 kΩ in series with the Miller capacitor C_m, and Fig. 7-16, connecting C_{m1} to the source of the cascode M_{51}. Figure 7-19 shows the photomicrograph. To compare, all four opamps were evaluated. Table 7-2 summarizes the measurement results. The table clearly shows the bandwidth improvement due to the removal of the RHP zero: 2.8 MHz for the reference circuit against about twice that value for the other three opamps. Increasing the current through the output transistor M_{60} from 20 µA to 2 mA improves the phase margins of the simple Miller compensated opamp and the Multipath Miller Zero Cancellation

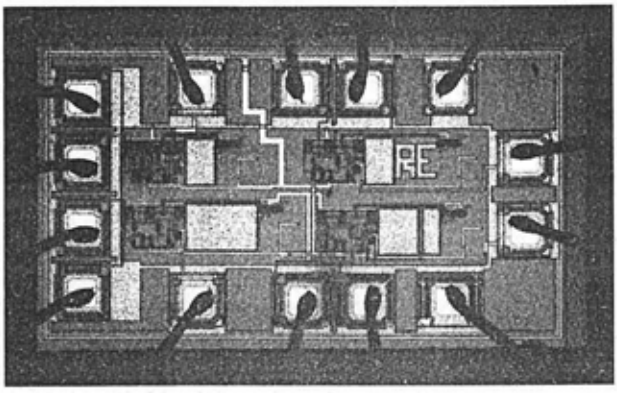


Fig. 7-19. *Photomicrographs of the chip containing four amplifiers, including the Multipath Miller Zero circuit.*

Table 7-2. *Measurement results of the four circuits.*

param.	I_q	Reference	Res.	Casc.	MMZC	unit
Unity-gain Freq.	20μA	2.8	5.5	5.4	6.5	MHz
	2mA	3.8	peaking	peaking	9.5	MHz
Phase Margin	20μA	42	50	42	50	°
	2mA	82	peaking	peaking	80	°
Supply Current	20μA	110	110	110	150	μA
B/P	20μA	7.7	15	15	13	GHz/W
			V_{dd}=3.3V, C_l=3pF, T=20°C			

circuit. The circuits with conventional RHP zero removal techniques (resistor and cascode) start peaking under the high current condition.

Figure 7-20 through Fig. 7-23 present the Bode plots of the four realized amplifiers. The plot of Fig. 7-20 is of the Miller compensated ref-

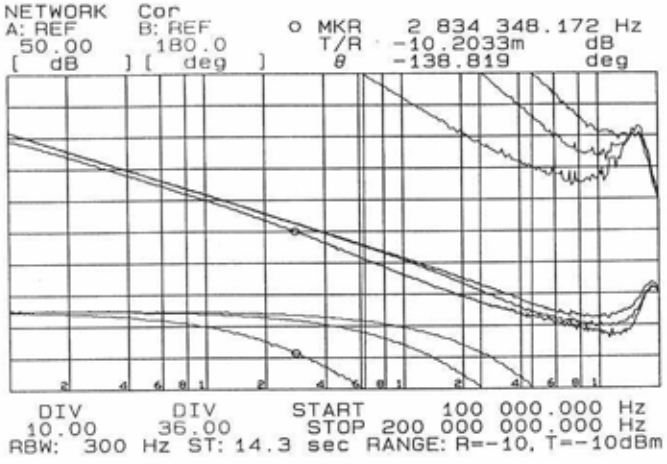


Fig. 7-20. *Bode plot of the reference opamp without RHP zero removal technique.*

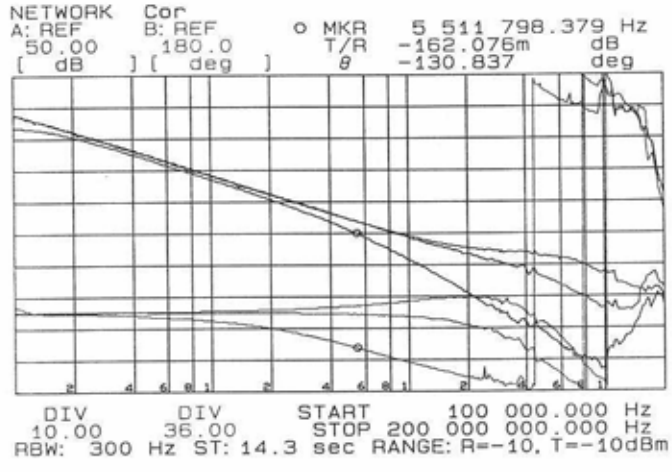


Fig. 7-21. *Bode plot of the opamp with resistive RHP zero removal.*

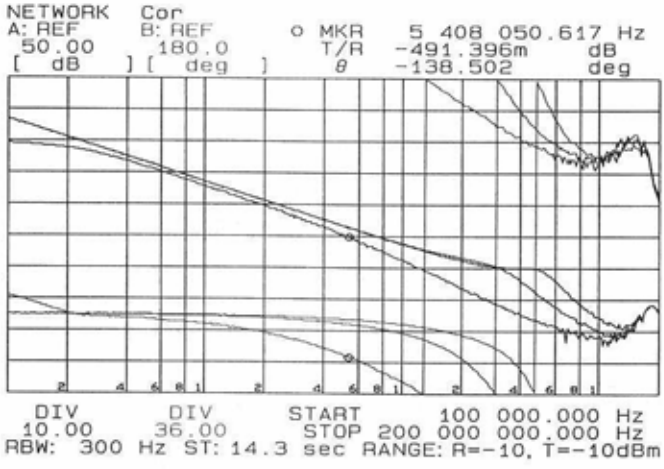


Fig. 7-22. *Bode plot of the opamp with cascode RHP zero removal.*

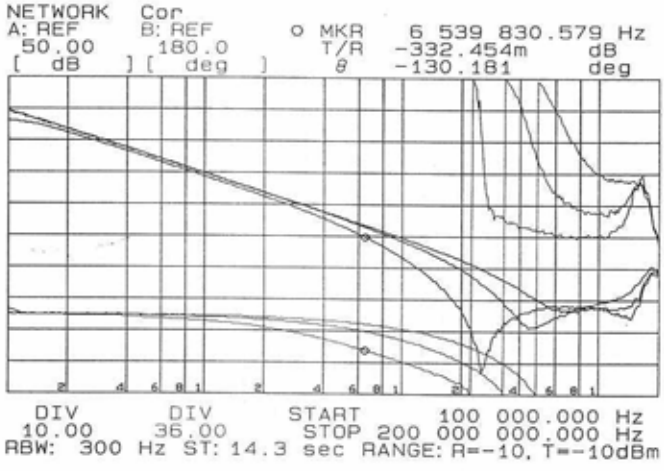


Fig. 7-23. *Bode plot of the opamp with the Multipath Miller Zero Cancellation technique for RHP zero removal.*

erence circuit without RHP zero removal. Figure 7-21 and Fig. 7-22 illustrate the peaking problems of the resistor and cascode RHP zero solutions. No peaking is detectable in the Bode plot Fig. 7-23, which gives the results of the Multipath Miller Zero Cancellation opamp. This verifies the robustness of the Multipath Miller Zero Cancellation technique.

7.3. Low-voltage opamps with HNMC and MHNMC

This section presents a rail-to-rail class AB operational amplifier in a standard V_{th}=0.6V CMOS technology operating at 1.5V. The amplifier incorporates the Hybrid Nested Miller Compensation technique to yield a unity-gain frequency of 6MHz at 300µA supply current, 120dB gain and programmability. Operation down to 1.1V at 15µA is possible with a unity-gain frequency of 400kHz. The realized chip measures 0.05mm^2. Further, a bipolar opamp operating at 1V has been realized using the Multipath Hybrid Nested Miller compensation technique. The bandwidth-to-power ratio of this circuit exceeds that of earlier reported bipolar low-voltage opamps by almost a factor of three.

7.3.1. Introduction

Traditionally CMOS operational amplifiers apply cascoding techniques to ensure an acceptable gain with a minimal number of gain stages. The use of cascodes, however, fundamentally limits the lowest supply voltage at which the amplifier circuit is able to operate. Furthermore, the apparent trend towards lower threshold voltages of CMOS processes endangers the applicability of cascoded circuits in the near future. The two restrictions on the use of cascoded opamps are illustrated by Fig. 7-24 and Fig. 7-25.

The most immediate drawback of the cascoded amplifier is its relatively high minimum supply voltage. Figure 7-24 presents the simplified schematic of a conventional operational amplifier with a push-pull output stage. It consists of the input stage M_{30}-M_{35} and the output stage M_{10}/M_{20} separated by cascode transistors M_{21} and M_{24}. The bold lines indicate the places where the stacked drain-source voltages of two cascode transistors are present on top of the gate-source voltage of an output transistor. The toted saturation voltages of the cascodes raise the lowest supply voltage one saturation voltage above the minimum obtainable. The minimum sup-

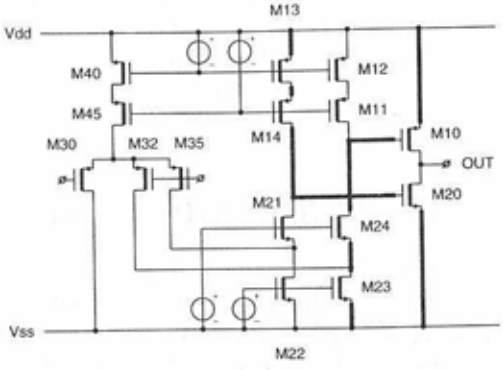


Fig. 7-24. *Limitation on the minimum supply voltage of the conventional cascoded opamp.*

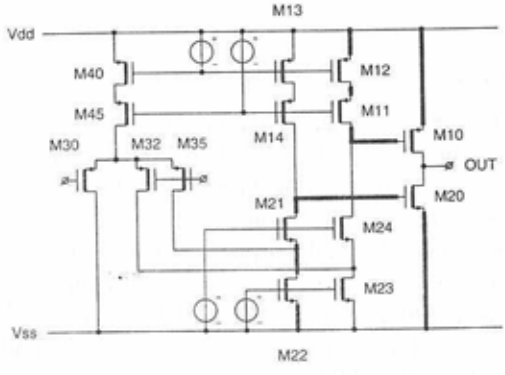


Fig. 7-25. *Limitation on the minimum threshold voltage of the cascoded opamp.*

ply voltage of a CMOS operational amplifier is mainly determined by the gate-source voltage of the output transistors under the maximum output current condition. When driving high output currents, the supply voltage

must suffice to build up the gate-source voltage of the output transistors plus the loss in the driving circuit. The two cascode devices account for a loss equal to their two summed saturation voltages V_{dsat}. The minimum supply voltage V_{min} is therefore given by

$$V_{min} = V_{gs,max} + 2V_{dsat} \tag{7-1}$$

where $V_{gs,max}$ is the gate-source voltage of the output transistors when driving the maximum output current. Considering that theoretically a single transistor can drive the output transistors, an optimal circuit requires a supply voltage of only

$$V_{min} = V_{gs,max} + V_{dsat} \tag{7-2}$$

The supply voltage that follows from Eq. 7-2 is the lowest minimum supply voltage for any CMOS amplifier. A circuit that operates at this limit will therefore be referred to as ultimate low-voltage.

A second drawback of the cascode circuit is that it is not suited for future I.C. processes with lower threshold voltages. The bold lines in Fig. 7-25 indicate the places where two stacked drain-source voltages are present *between* the gate and the source of an output transistor. These are the critical points with respect to the lowest acceptable threshold voltage of the manufacturing I.C. process. Under quiescent conditions, when the current through the output devices is low, the gate-source voltage V_{gs} will be close to the threshold voltage V_{th}. For correct operation of the amplifier, the stacked drain-source voltages V_{ds} of the cascode transistors M_{11}/M_{12} and M_{21}/M_{22} must fit into the gate-source voltages V_{gs} of the output transistors. Since the lowest drain-source voltage for normal operation is the saturation voltage V_{dsat}, the summed saturation voltages have to be smaller than the gate-source voltages of the output transistors. The gate-source voltages V_{gs} of the output transistors lower proportionally to the threshold voltage, however, while the saturation voltage remains fixed. Redimensioning the circuit for a lower threshold process therefore requires an increased *W/L* ratio of the cascode transistors, enlarging the occupied die area. The continuously diminishing threshold voltages will eventually result in excessively large cascode components.

The observation that the cascode transistors must be widened to reduce their saturation voltage in processes with a lower threshold voltage becomes more alarming when we consider a fundamental limit to the

reduction of the saturation voltage. This lower limit is set by the subthreshold saturation voltage of a MOS device. The cascode transistors enter the subthreshold mode of operation when they have been widened to such an extent that the current density drops below the limit for normal operation. Under the subthreshold regime, a MOS transistor essentially operates as a bipolar component, with a fixed saturation voltage. The value is given by [31]

$$V_{dsat,st} \approx \frac{4kT}{q} \tag{7-3}$$

which amounts to approximately 100mV at room temperature. The minimum threshold voltage of the process in which the opamp can still be realized will eventually be limited by this subthreshold saturation voltage.

We can find the absolute minimum threshold voltage for the cascoded opamp by considering that the two stacked saturation voltages of the cascodes not only have to fit into the nominal threshold voltage, but also have to enable an acceptable yield for the minimum threshold voltage. The maximum deviation of the threshold voltage is in practice given by the statistical 3σ limit. It lies in the order of 150mV and is not expected to improve appreciably in the future. Combining this value with the subthreshold saturation voltage, the minimum acceptable threshold voltage $V_{th,min}$ can be found as

$$V_{th,min} = 2V_{dsat,st} + V_{d,max} \tag{7-4}$$

where $V_{dsat,st}$ is the subthreshold saturation voltage of the cascode devices and $V_{d,max}$ is the 3σ deviation of the threshold voltage from its typical value.

Equation 7-4 becomes approximately 350mV at room temperature. Realized in a CMOS process with a threshold voltage lower than this value, the gain of the cascoded amplifier will rapidly decrease and the amplifier will become inoperative. This finding strongly suggests that the future applicability of cascoded opamps is uncertain. Figure 7-26 substantiates the prediction by showing the development of the threshold voltage as derived from a decade of *IEEE Journal of Solid State Circuits* volumes. The horizontal line in the figure represents the minimum threshold voltage of 400mV. The descending line, which plots the threshold voltage of a typical process, lies well above this limit for early and more recent processes.

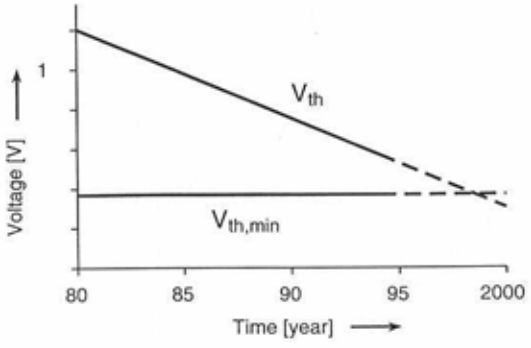


Fig. 7-26. *Development of the threshold voltage compared to the minimum value required for conventional topologies.*

Although the *W/L* ratio of the cascode transistors already may have to be large, which is a drawback in itself, we are not faced with insurmountable problems when realizing the cascoded opamp at this moment. By the turn of the century, however, the gate-source voltage of the output transistors might drop below the value where it is impossible to accommodate the two stacked drain-source voltages of the cascode transistors over a reasonable temperature range. At that time an important class of opamp circuits will lose its significance.

The traditional alternative to cascodes to obtain a high gain is the use of a cascade topology. Cascading or chaining amplifier stages is an effective approach to obtain a high gain. Conventional cascaded CMOS amplifiers suffer from the same restrictions as their cascode counterparts. These limitations are strongly related to the necessity for frequency compensation.

Methods to robustly frequency compensate operational amplifiers containing four stages are rare. An exception is the Nested Miller compensation structure. Figure 7-27 shows a four-stage Nested Miller compensated amplifier with a push-pull output stage. Capacitors C_{m1}, C_{m2} and C_{m3} are the respective nested capacitors. Differential pairs implement the stages M_{30} and M_{40} to accomplish a correct sign for the feedback loops

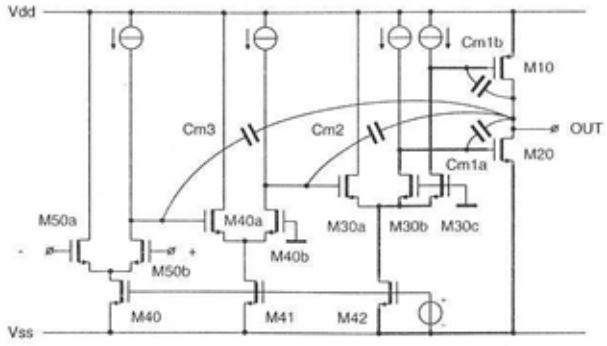


Fig. 7-27. *Opamp with four cascaded gain stages and Nested Miller Compensation.*

closed by the Miller capacitors. The differential pairs in the NMC circuit lead to the presence of two stacked drain-source voltages inside the gate-source voltages of the output transistor and two drain-source voltages on top of these gate-source voltages. The bold lines in Fig. 7-27 reveal the locations of these stacked voltages.

This section addresses two ultimate low-voltage CMOS amplifiers. Realized in a $V_{th} = 0.6\,V$ process and with a total supply current of 300µA, their minimum supply voltage is 1.5 V. Reducing the total supply current to 16µA, the minimum supply voltage becomes as low as 1.1 V. Furthermore, the opamps will remain functional in future processes with extremely low threshold voltages. The keys to these results are the techniques used for the frequency compensation. The first opamp employs the Hybrid Nested Miller compensation scheme, resulting in a unity-gain frequency of 2MHz. The Multipath Hybrid Nested Miller structure compensates the second amplifier. The unity-gain frequency of this circuit is 6MHz. The bandwidth improvement relies on the Multipath input stage that directly drives the output transistors, thus bypassing the intermediate stages. The pole-zero doublet that the Multipath technique introduces is found to match within 1%. Both opamps have a D.C. gain of 120dB.

To demonstrate the general applicability of the new compensation methods, a third, bipolar, low-voltage operational amplifier with Multipath

Hybrid Nested Miller compensation is presented. The opamp operates from a 1 V supply, consumes 250µA of current and displays a unity-gain frequency of 3.5MHz. The D.C. gain is again 120dB. Comparing these results to previously published 1 V bipolar opamps, the Multipath Hybrid Nested Miller compensated amplifier has a bandwidth-to-power ratio two times higher. Moreover, the circuitry is considerably simpler and the area occupied on the silicon die is smaller.

In section 7.3.2 the principles of operation of Hybrid Nested Miller compensation and Multipath Hybrid Nested Miller compensation are explained. Section 7.3.3 addresses two ultimate low-voltage CMOS opamps, the bipolar counterpart is discussed in section 7.3.4. Section 7.3.5 shows the experimental results, and finally section 7.3.6 gives the conclusions.

7.3.2. Principle of operation

In summary, the two demands on CMOS ultimate low-voltage operational amplifier circuits prepared for future lower threshold voltages are

- Only one stacked drain-source voltage V_{ds} in between the gate and the source of an output transistor.
- Only one drain-source voltage V_{ds} on top of the gate-source voltage of an output transistor.

Figure 7-28 suggests a basic topology of the opamp that meets the ultimate

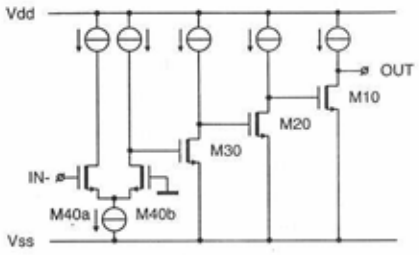


Fig. 7-28. *Basic ultimate low-voltage topology.*

low-voltage demands. It comprises of three cascaded common source

NMOS transistors M_{10}-M_{30} and a differential input stage M_{40}. Because of the inverting nature of the intermediate and output stages, traditional frequency compensation techniques fail. Consider Fig. 7-29 for instance,

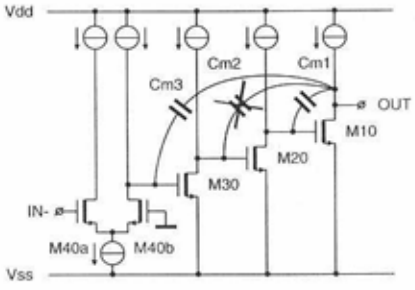


Fig. 7-29. *Ultimate low-voltage circuit with traditional Nested Miller compensation.*

which shows the circuit with three additional capacitors for Nested Miller compensation. The capacitors intend to separate the poles at each of the nesting levels, resulting in a straight 20dB/dec roll-off of the gain magnitude. The middle Miller capacitor C_{m2} closes a feedback loop with a positive sign, however. The positive feedback directs the poles into the right half of the complex s-plane and the circuit becomes unstable, even before overall feedback is applied.

The proposed solution for frequency compensation is the Hybrid Nested Miller Compensation (HNMC) structure, depicted in Fig. 7-30. The structure again employs feedback capacitors, comparable to the Nested Miller Compensation, but in this case the two Miller loops with C_{m1} and C_{m2} are at the same nesting level. We can easily verify that the feedback sign of all internal loops is correct. This is the case for the loops with C_{m1} and C_{m2}, since these are ordinary Miller capacitors, but also C_{m3} is part of a negative feedback loop, because three inverting stages are present between its terminals.

Although the bandwidth reduction of Hybrid Nested Miller compensation is small compared to other methods for frequency compensation, in some cases even this small reduction is not acceptable. When this occurs, the Multipath Hybrid Nested Miller compensation (MHNMC)

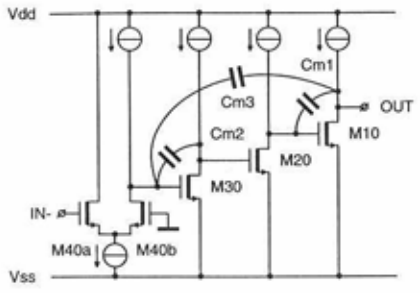


Fig. 7-30. *Ultimate low-voltage circuit with Hybrid Nested Miller compensation.*

technique is an attractive option. Figure 7-31 shows a Multipath Hybrid

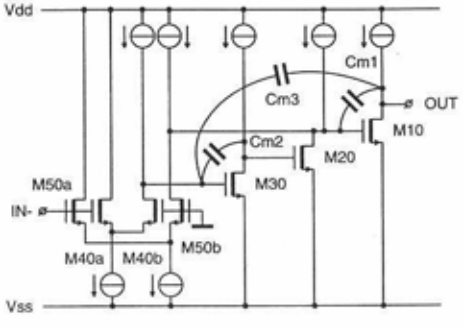


Fig. 7-31. *Ultimate low-voltage circuit with Multipath Hybrid Nested Miller compensation.*

Nested Miller compensated amplifier. Added to the previous circuit in Fig. 7-30 is an additional input stage M_{50} that directly drives the output transistor M_{10}. The extra input stage and the output stage together constitute a two-stage amplifier with relatively low gain, but a high bandwidth.

This is presented in the Bode plot of Fig. 7-32 by the dashed line. The rest

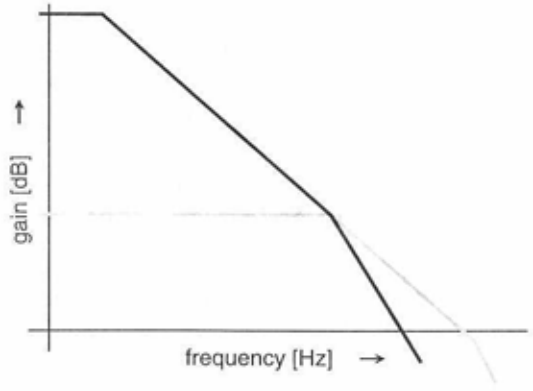


Fig. 7-32. *Bode plot showing the built-up of the frequency response of the Multipath Hybrid Nested Miller compensation structure.*

of the circuit, a four-stage amplifier with Hybrid Nested Miller compensation, couples a low bandwidth to a high gain. This leads to the solid line in Fig. 7-32. The two curves together yield a 20dB/dec roll-off with both a high unity-gain frequency and a high gain. Since the bandwidth is determined by the two-stage amplifier with single Miller compensation, it is close to the theoretical maximum that can be obtained with a given transconductance of the output stage and load capacitance.

Section 5.4 and Section 6.3 discuss the exact operation of Hybrid Nested Miller compensation and Multipath Hybrid Nested Miller compensation, respectively.

7.3.3. The ultimate low-voltage opamps

Figure 7-33 shows the simplified circuit of the ultimate low-voltage opamp with Hybrid Nested Miller compensation having a push-pull output stage for obtaining rail-to-rail operation. The *P*-channel input pair M_{110}/

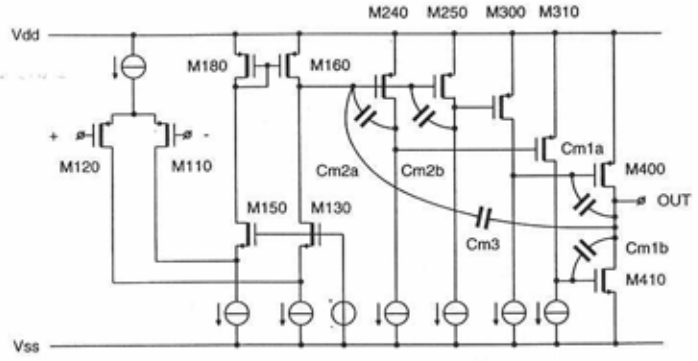


Fig. 7-33. *Simplified schematic of the ultimate low-voltage amplifier.*

M_{120}, followed by the folded cascodes M_{130}/M_{150}, ensures a common mode input range that includes the negative rail. Current mirror M_{160}/M_{180} performs differential-to-single conversion of the signal. We can recognize the four gain stages as the input stage M_{110}/M_{120}, the first intermediate stage M_{240}/M_{250}, the second intermediate stage M_{300}/M_{310} and finally the output stage M_{400}/M_{410}. The intermediate stages comprise two transistors, each to correctly generate the driving signals for the push-pull output stage. The capacitors in Fig. 7-33 are inserted according to the Hybrid Nested Miller compensation scheme. The lowest level capacitors are also doubled to deal with the separate signal paths for the push-pull output stage.

Figure 7-34 shows the total schematic of the realized low-voltage opamp circuit with Hybrid Nested Miller compensation. Added to the simplified circuit of Fig. 7-33 are biasing elements and a feedback class AB control which ensures a minimum quiescent current through the inactive output transistor to prevent shut-off. The four amplifier stages in Fig. 7-34 are: input stage M_{110}/M_{120}, second stage M_{240}/M_{250}, third stage M_{300}/M_{310} and output stage M_{400}/M_{410}. The Hybrid Nested Miller capacitors are inserted according to Fig. 7-33. The class AB control is constructed around differential amplifier M_{501}-M_{520}. The currents through output

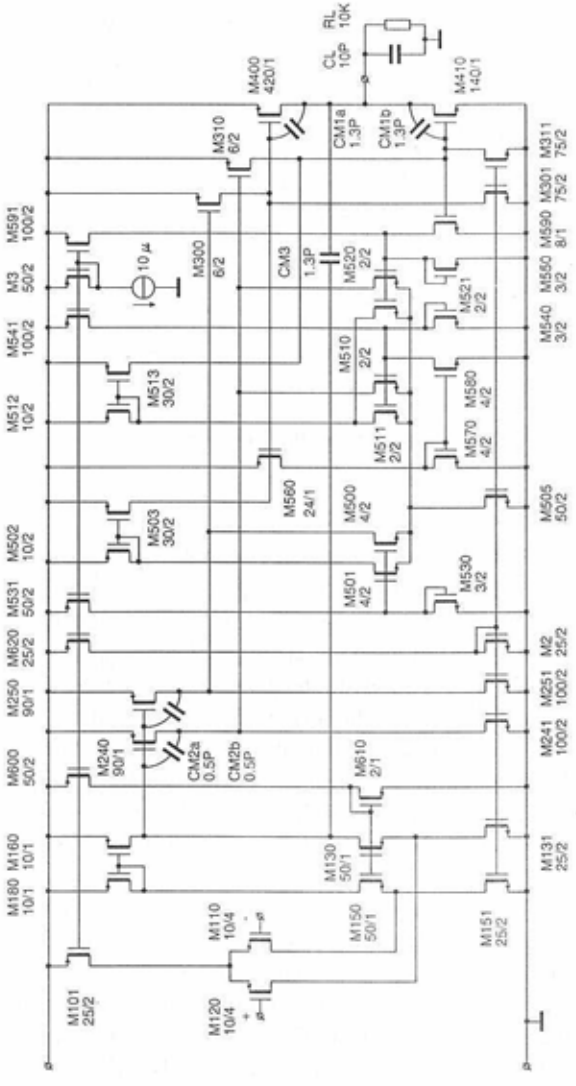


Fig. 7-34. *Total schematic of the CMOS ultimate low-voltage opamp with Hybrid Nested Miller compensation.*

devices M_{400} and M_{410} are measured by M_{560} and M_{590}. The drain currents of these measurement transistors are converted into voltages at the gates of M_{511}/M_{510} and M_{521}/M_{520}, two inputs of the class AB control amplifier. The voltage that represents the lowest of the two output transistor currents is compared to the reference voltage available at the third input of the control amplifier, the gates of transistors M_{501}/M_{500}. The selection of the input voltage that corresponds to the smallest transistor current is performed by the decision pair consisting of transistors M_{510} and M_{520}. The transistor that senses the highest voltage at its gate enters the active current conducting mode. It therefore acts as a source follower, transferring its gate voltage to the common source node of the class AB amplifier. The other device, with the lowest gate voltage, is shut off and does not contribute to the signals in the class AB loop. Depending on the difference between the actual lowest transistor current and the desired limiting value, a correction signal is fed to the input of the third gain stage (M_{300}/M_{310}). To guarantee stability of the class AB control loop, an identical signal is fed directly to the output transistors, bypassing the third stage. This is analogous to the use of a Multipath input stage for the overall frequency compensation of the circuit. The low gain-high frequency parallel path passes through the current mirrors M_{502}/M_{503} and M_{512}/M_{513}.

Figure 7-35 shows the total schematic of the ultimate low-voltage opamp with Multipath Hybrid Nested Miller compensation. Apart from the Multipath input stage, comprising transistor M_{100}, and the subsequent transistors in its signal path, the circuit is identical to that of the Hybrid Nested Miller compensated amplifier. Some of the biasing components have been redimensioned, however, to comply with the changed parameter conditions for compensation.

7.3.4. The bipolar MHNMC opamp

The (Multipath) Hybrid Nested Miller technique can also be used for bipolar low-voltage operational amplifiers. The exponential relation between driving voltage and output current of a bipolar transistor allows an even lower supply voltage than the weaker quadratic relation of MOS, since a small driving voltage already yields a relatively large output current. The exponential voltage-current relation also assures that the base-emitter voltages of the output transistors are virtually constant under varying current and process conditions. Therefore no limitations apply that are equivalent to those postulated for CMOS amplifiers in the previous section.

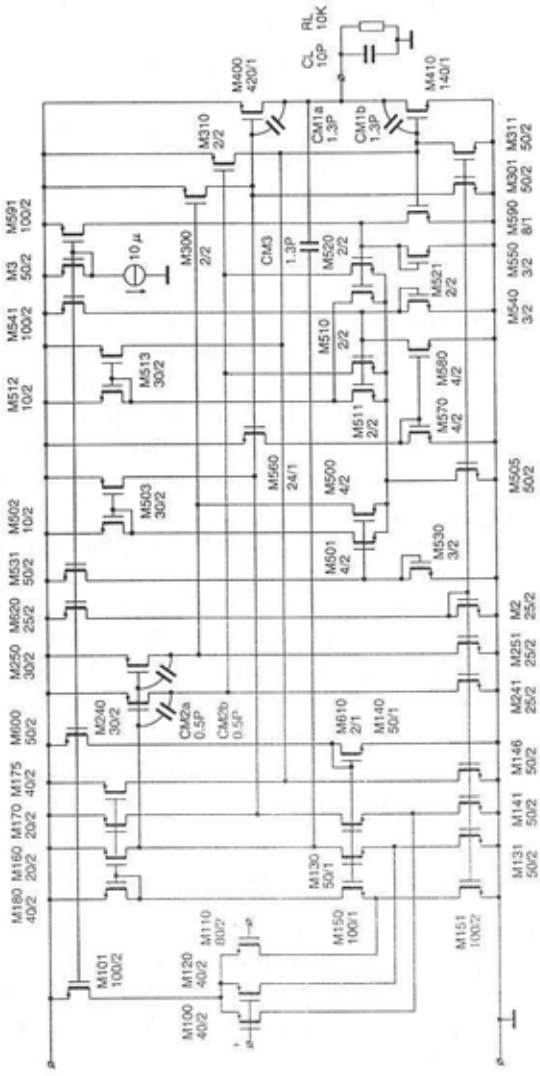


Fig. 7-35. *Total schematic of the CMOS ultimate low-voltage opamp with Multipath Hybrid Nested Miller compensation.*

Figure 7-36 shows the circuit diagram of a bipolar opamp with Multipath Hybrid Nested Miller compensation that can operate at supply voltages down to 1 V. Although this minimum value has been reported before for a four-stage bipolar circuit with similar functionality [21], the Multipath Hybrid Nested Miller compensated opamp is considerably simpler and the bandwidth-to-power ratio almost a factor of three higher. The improvement of the bandwidth-to-power ratio is due to the reduced number of nesting levels of Hybrid Nested Miller compensation as compared to the conventional Nested Miller compensation technique. A Nested Miller compensated circuit having four stages requires a total of three nesting levels, each inducing a bandwidth reduction of approximately a factor of two, whereas for Hybrid Nested Miller compensation only two nesting levels are needed.

The structure of the bipolar opamp is almost identical to that of the CMOS variety. The circuit has four stages. The input stage consists of transistors Q_{100} and Q_{112}. It is followed by the folded cascode Q_{200} and Q_{210} and the current mirror Q_{220}/Q_{230} that performs the differential to single conversion. The first intermediate stage is the pair Q_{300}/Q_{310} and the second intermediate stage are the transistors Q_{400}/Q_{410}. These devices drive the push-pull output stage, which comprises transistors Q_{500} and Q_{510}. The Hybrid Nested Miller compensation capacitors at the inner nesting level are C_{m1a}, C_{m1b}, C_{m2a} and C_{m2b}, while C_{m3} closes the outer Miller loop. Small resistors in series with the capacitors at the first level counteract the appearance of a zero in the right half of the complex s-plane, thus theoretically improving the bandwidth. Their effect is marginal, however, and they can be left out without dramatic changes to the performance. The Multipath input stage is implemented through the transistors Q_{110} and Q_{111}. These directly drive the output transistors through the folded cascodes.

The class AB control again assures a minimum rest current through the non-active output transistor. It does so by measuring the currents through the respective output transistors and converting these to voltages across R_{601} and R_{610}. Transistors Q_{620} and Q_{630} act as a decision pair, transferring the smallest of the two voltages to the common emitter node. This voltage, which corresponds to the smallest output transistor current, is compared to a reference value by the differential amplifier Q_{320}/Q_{330}. The result of the comparison is amplified and fed back to the input of the third stage, thus closing the control loop. The transistors Q_{321} and Q_{331}, which are connected parallel to the class AB differential amplifier, gener-

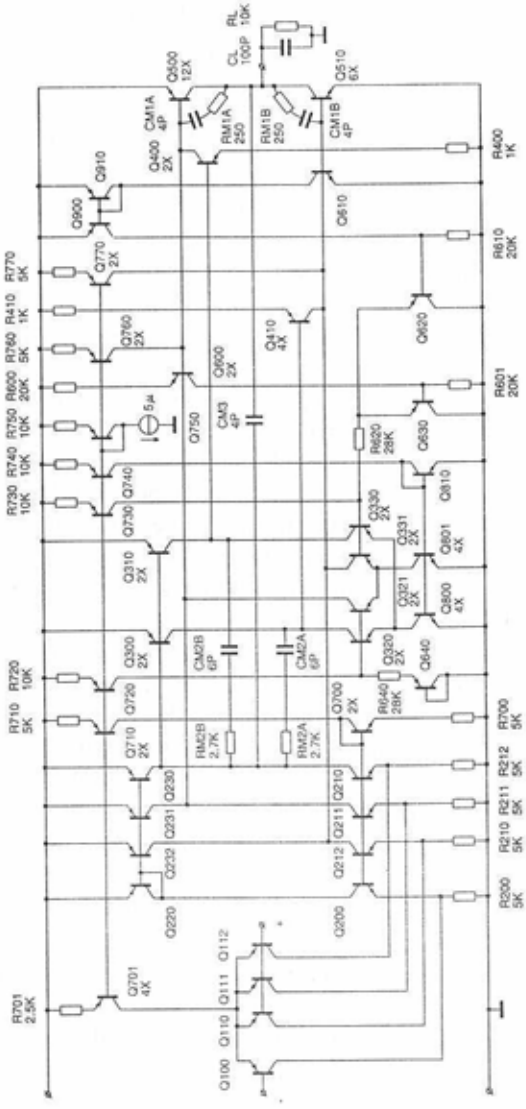


Fig. 7-36. *Total schematic of the bipolar low-voltage opamp with Multipath Hybrid Nested Miller compensation.*

ate a second pair of control currents that directly drive the output stage. They serve to improve the stability of the class AB loop by bypassing the third stage for high frequency signals.

7.3.5. Realizations and measurement results

The CMOS operational amplifier with Hybrid Nested Miller Compensation and the circuit with the Multipath technique to double the bandwidth have been realized in a standard V_{th}=0.6V process with a minimum feature size of 0.8µm. The photomicrographs of the chips are shown in Fig. 7-37 and Fig. 7-38, respectively. The micrographs demonstrate that,

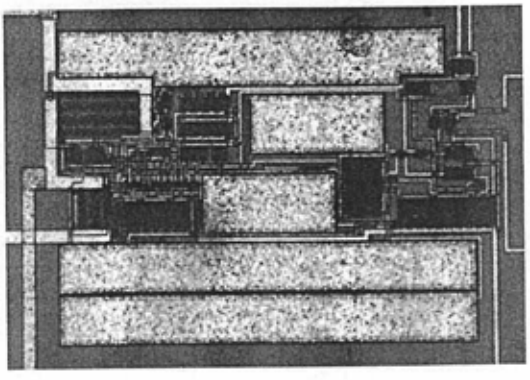


Fig. 7-37. *Die photomicrograph of the Hybrid Nested Miller compensated low-voltage CMOS opamp.*

despite the fact that the capacitors of both circuits were chosen to be the same values, there is hardly any die area penalty for the addition of a Multipath input stage. The area of both chips is 0.05mm^2 (70mil^2). Further, although the class AB control appears to be rather complex in the circuit diagram, the occupied die area of this circuit part is negligible. This is due to the small sizes of the devices in the class AB circuit. Figure 7-39 shows the open-loop Bode plot of the opamp with Hybrid Nested Miller Compensation in Fig. 7-33 at a total supply current of 300µA. The unity-gain frequency is 2MHz.

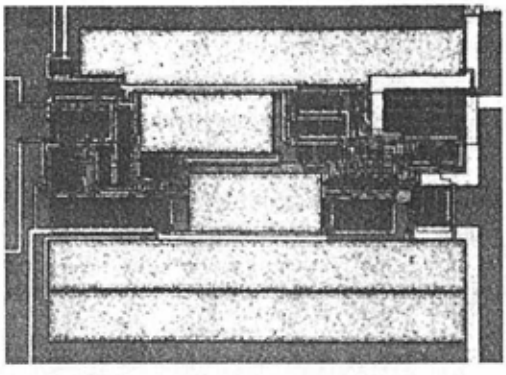


Fig. 7-38. *Die photomicrograph of the Multipath Hybrid Nested Miller compensated low-voltage CMOS opamp.*

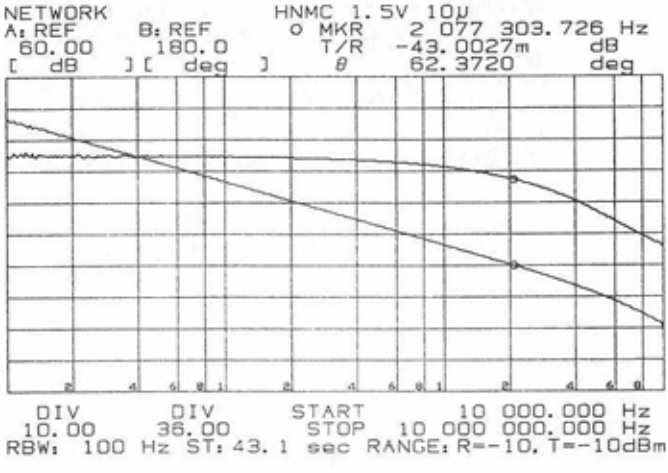


Fig. 7-39. *Measured frequency response of the CMOS ultimate low-voltage opamp with Hybrid Nested Miller compensation.*

The Bode plot of Fig. 7-39 demonstrates the effect of a second

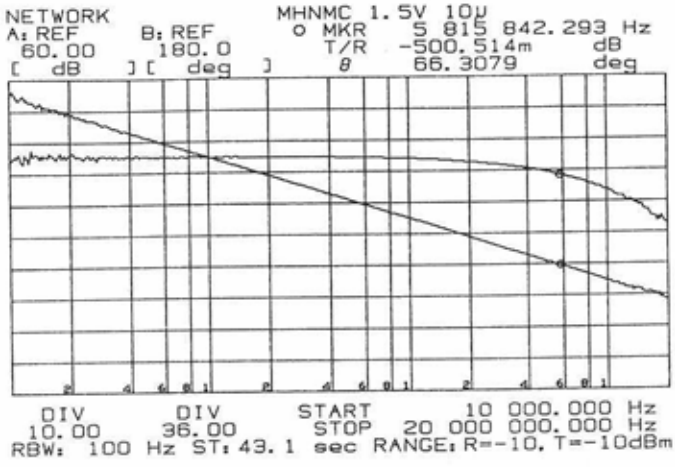


Fig. 7-40. *Measured frequency response of the CMOS ultimate low-voltage opamp with Multipath Hybrid Nested Miller compensation.*

input stage on the unity-gain frequency by showing the open-loop gain of the circuit with Multipath Hybrid Nested Miller Compensation. The unity-gain frequency now rises to 6MHz, slightly more than expected. In both cases the load of the opamp consisted of a 10kΩ resistor parallel to a capacitor of 10pF. Finally, the responses of the two amplifiers to a 100mV step input are shown in Fig. 7-41 and Fig. 7-42. The higher bandwidth of the Multipath Hybrid Nested Miller compensated amplifier is reflected in the more narrow peak in the step response. No slow-settling components are detectable in the plot, indicating that the matching of the pole-zero doublet is better than 1%. The settling times to 1% are 320ns and 140ns, respectively.

The experimental results of the CMOS opamps are gathered in two tables. To illustrate the robustness of the Hybrid Nested Miller Compensation structure, the measurements were performed at two quiescent supply currents: 300µA (Table 7-3) and 16µA and 15µA respectively (Table 7-4). The programming of the quiescent current was accomplished by varying the bias current source feeding M_3. In the 16µA/15µA case, the unity-gain bandwidth reduces to 400kHz for the Multipath Hybrid

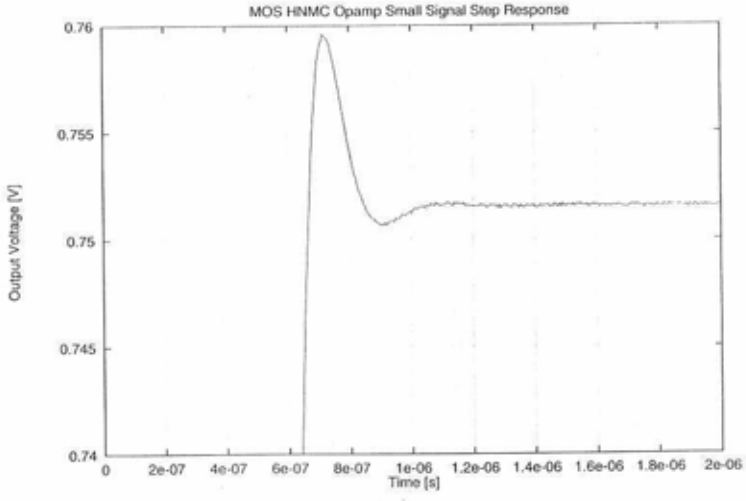


Fig. 7-41. *Step response of the CMOS ultimate low-voltage opamp with Hybrid Nested Miller compensation.*

Nested Miller compensated amplifier, but also the minimum allowable supply voltage lowers to 1.1 V, bringing the total power consumption down to 17 µW. The net bandwidth-to-power ratio improves from 14 GHz/W to 36 GHz/W. This is in correspondence to the analysis of Section 4.4.1, which states that for the best power-to-bandwidth ratio the MOS devices should be operated as close as possible to their subthreshold regions.

The bipolar opamp was fabricated in a process offering f_t = 3 GHz NPNs and f_t = 1 GHz vertical PNPs. The photomicrograph of the chip, which occupies 0.5 mm^2 (700 mil^2), is shown in Fig. 7-43. The circuit operates from a 1 V supply and the unity-gain frequency is 3.5 MHz. Figure 7-44 gives the measured frequency response with a load consisting of a 100 pF capacitor parallel to a 1 kΩ resistor. The same load was used for the measurement results that are summarized in Table 7-5. The realized bandwidth-to-power ratio of 14 GHz/W presents an improvement of almost a factor of three compared to similar circuits presented before [21].

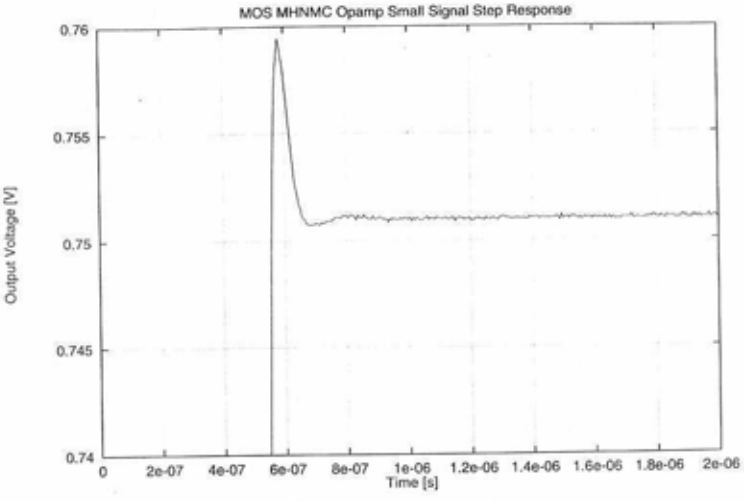


Fig. 7-42. *Step response of the CMOS ultimate low-voltage opamp with Multipath Hybrid Nested Miller compensation.*

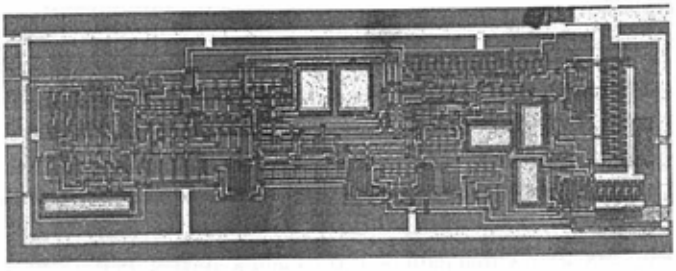


Fig. 7-43. *Die photomicrograph of the bipolar low-voltage opamp.*

Table 7-3. *Measurement results of the low-voltage CMOS opamps at* $I_{dd}=300\mu A$.

Symbol	Parameter	HNMC	MHNMC	Unit
T_{op}	temperature range	-55~+125	-55~+125	°C
V_{dd}	supply voltage range	1.5~5.5	1.5~5.5	V
I_{dd}	supply current	300	290	µA
I_l	maximum output current	5	5	mA
V_{out}	output voltage swing over temperature	V_{ss}+0.2~ V_{dd}-0.2	V_{ss}+0.2~ V_{dd}-0.2	V
V_{os}	offset voltage	3	4	mV
A_{ol}	open-loop voltage gain	120	120	dB
F_{ug}	unity-gain frequency	2	6	MHz
ϕ_m	phase margin over temperature	59~65	60~69	°
S_r	slew rate	5	13	V/µs
T_s	1% settling time (100mV step)	320	140	ns
B/P	bandwidth-to-power ratio	4.4	14	GHz/W

V_{dd}=1.5V, I_{dd}=300µA, R_l=10kΩ and C_l=10pF unless stated otherwise

7.3.6. Conclusions

The design of ultimate low-voltage CMOS opamps topologies requires frequency compensation methods that are suited for a cascade of inverting gain stages. Hybrid Nested Miller compensation and Multipath Hybrid Nested Miller compensation meet that demand, whereas traditional methods fails. By employing these new techniques, we have seen a four-stage CMOS opamp with a gain of 120dB and a unity-gain frequency of 6MHz. The operational amplifier operates at a supply voltage of 1.5V and consumes 300µA of supply current. The supply current can be programmed down to 16µA and 15µA, respectively, reducing the minimum supply

Table 7-4. *Measurement results of the low-voltage CMOS opamps at I_{dd}=16μA/15μA.*

Symbol	Parameter	HNMC	MHNMC	Unit
T_{op}	temperature range	-55~+125	-55~+125	°C
V_{dd}	supply voltage range	1.1~5.5	1.1~5.5	V
I_{dd}	supply current	16	15	μA
I_l	maximum output current	0.5	0.5	mA
V_{out}	output voltage swing over temperature	V_{ss}+0.1~ V_{dd}-0.1	V_{ss}+0.1~ V_{dd}-0.1	V
V_{os}	offset voltage	3	4	mV
A_{ol}	open-loop voltage gain	120	120	dB
F_{ug}	unity-gain frequency	0.2	0.6	MHz
ϕ_m	phase margin over temperature	58~64	45~53	°
S_r	slew rate	0.2	0.7	V/μs
T_s	1% settling time (100mV step)	2.9	1.6	μs
B/P	bandwidth-to-power ratio	11	36	GHz/W

V_{dd}=1.1V, I_{dd}=300μA, R_l=10kΩ and C_l=10pF unless stated otherwise

voltage even further to 1.1V. Furthermore, the opamp topology is suited for use in future manufacturing processes with substantially lower threshold voltages.

The use of Multipath Hybrid Nested Miller compensation is also beneficial to bipolar low-voltage operational amplifiers. This section presented an opamp operating at 1V that shows both a significantly better bandwidth-to-power ratio and lower circuit complexity compared to existing circuits.

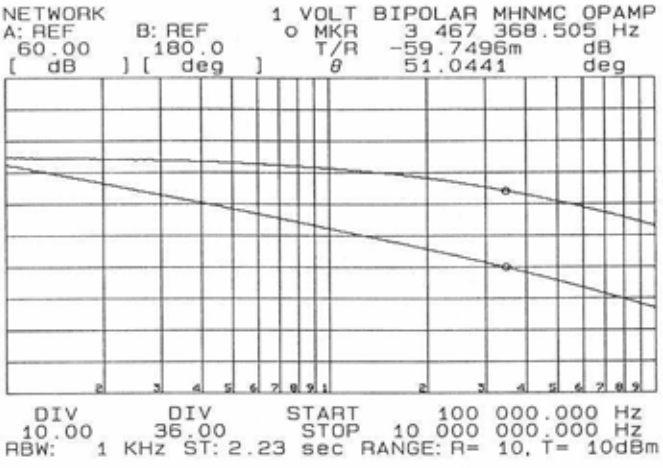


Fig. 7-44. *Measured frequency response of the bipolar low-voltage opamp.*

7.4. Multipath Conditionally Stable amplifier

With a slight modification, the bipolar Multipath Hybrid Nested Miller compensated circuit of Section 7.3.4 can function as a Multipath Conditionally Stable amplifier (see Section 6.5). The alteration merely involves the removal of the outer Miller capacitor C_{m3}. Figure 7-45 presents the circuit diagram. The circuit is almost identical to that of Fig. 7-36, except for the Miller capacitor.

The conditionally stable frequency response of the circuit considerably improves the amount of feedback up to the unity-gain frequency. This is shown in the simulated Bode plots Fig. 7-46. The lower trace in this plot represents the frequency response of the absolutely stable circuit of Fig. 7-36, the upper trace, that of the conditionally stable circuit of Fig. 7-45. The difference of the two plots makes clear the increase of the midband loop gain. At 100kHz the improvement amounts to 30dB. Figure 7-47 shows the measured high end part of the frequency response of the circuit. As expected, the phase characteristic closely approaches

Table 7-5. *Measurement results of the bipolar low-voltage opamp.*

Symbol	Parameter	Value	Unit
T_{op}	temperature range	-55~+125	°C
V_{dd}	supply voltage range	1.0~14	V
I_{dd}	supply current	250	μA
I_l	maximum output current	15	mA
V_{out}	output voltage swing over temperature	V_{ss}+0.1~ V_{dd}-0.1	V
V_{os}	offset voltage	1	mV
A_{ol}	open-loop voltage gain	120	dB
F_{ug}	unity-gain frequency	3.5	MHz
ϕ_m	phase margin over temperature	45~60	°
S_r	slew rate	1.1	V/μs
T_s	1% settling time (100mV step)	300	ns
B/P	bandwidth-to-power	14	GHz/W

V_{dd}=1V, I_{dd}=250μA, R_l=1kΩ and C_l=100pF unless stated otherwise

180° in the middle of the frequency band, but rises back to approximately 120° at the unity-gain frequency, yielding a phase margin of 60°. The amplifier will exhibit sufficient phase margins for closed-loop gains ranging between 1 and 10.

Finally, Fig. 7-48 shows the photomicrograph of the realized amplifier. The process in which it was fabricated enabled f_t = 3GHz NPN transistors and f_t = 1GHz vertical PNPs. The chip measures 0.5mm^2 (700mil^2). Apart from the conditionally stable frequency response, the characteristics of the circuit in Fig 7-45 are identical to those of the low-voltage opamp of Fig. 7-36. Therefore, the summarized specifications of Table 7-5, unaltered, hold for the conditionally stable circuit.

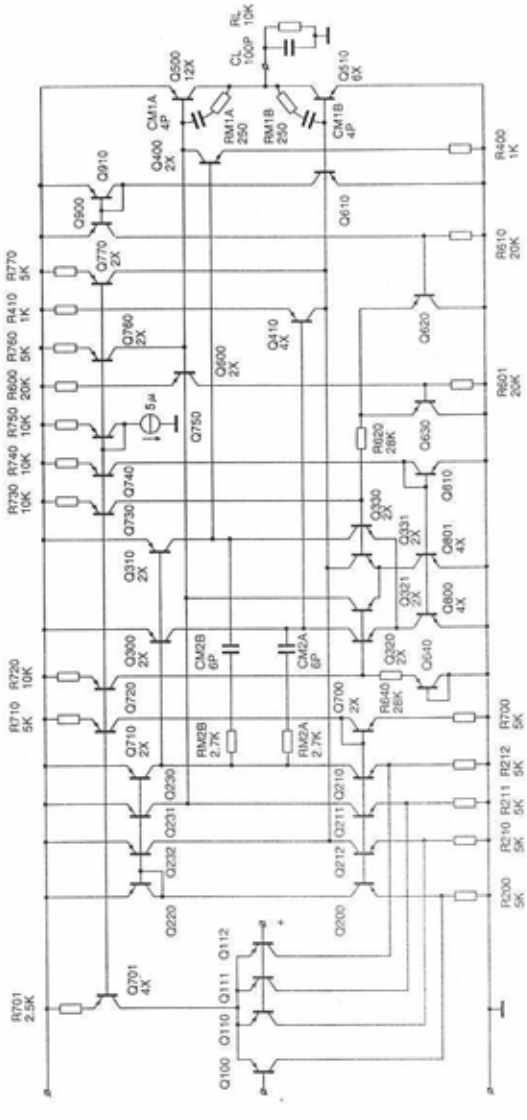


Fig. 7-45. *Circuit diagram of the Multipath Conditionally Stable amplifier.*

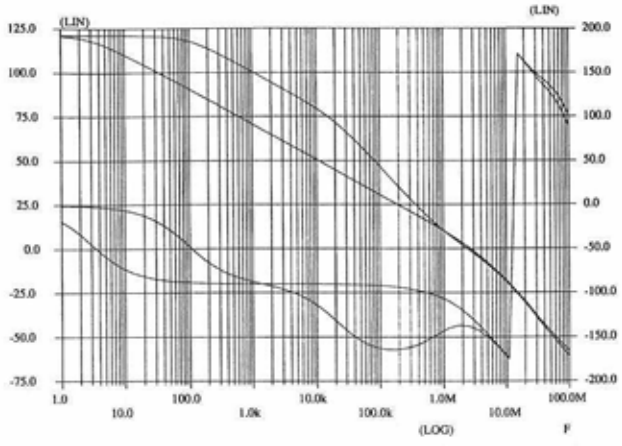


Fig. 7-46. *Simulated frequency response of (top) the Multipath Conditionally Stable amplifier and (bottom) an equivalent absolutely stable amplifier.*

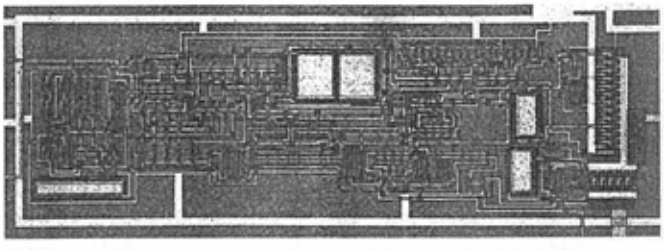


Fig. 7-48. *Photomicrograph of the Multipath Conditionally Stable amplifier.*

7.5. Conclusions

In this chapter, we met several realizations of amplifiers using the compensation methods described earlier in this work. Both bipolar and CMOS cir-

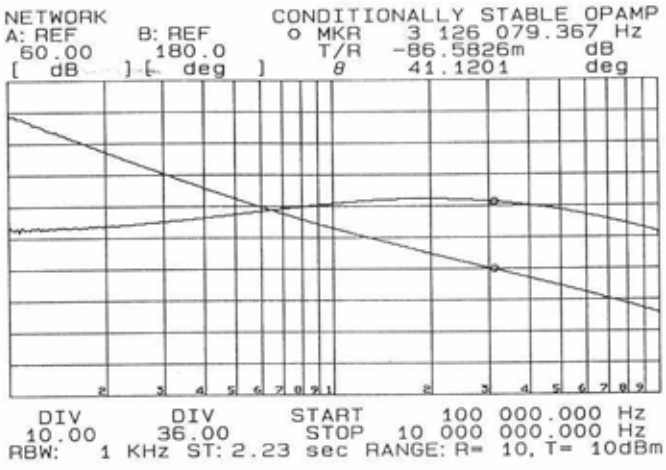


Fig. 7-47. *Measured frequency response of the Multipath Conditionally Stable amplifier.*

cuits are discussed. In each of the technologies general purpose operational amplifiers with an absolutely stable frequency response predominate. The unity-gain bandwidth of these amplifiers ranges from 400kHz for a micro-power CMOS opamp with Hybrid Nested Miller compensation up to 100MHz for a bipolar precision opamp with Multipath Nested Miller compensation. These largely differing bandwidth figures clearly stress the importance of considering the bandwidth-to-power ratio instead of the bandwidth alone. To illustrate, the power consumption of the 400kHz circuit is a mere 17μW, while its 100MHz counterpart consumes approximately 95mW.

A silicon realization of a conditionally stable amplifier demonstrated the improvement of the midband loop gain that can be obtained by leaving the absolutely stable frequency response requirement. The conditionally stable frequency characteristic is obtained through the adoption of the Multipath Conditionally Stable frequency compensation scheme. This technique guarantees stability of the amplifier even if the operating conditions vary, either due to statistical variations of the I.C. process or changes of external parameters such as the load impedance.

Furthermore, this chapter presented an implementation of the Multipath Miller Zero Cancellation technique that counteracts the bandwidth impeding zero in the right half of the complex s-plane typical to Miller compensated amplifiers. Compared to the traditional methods for Right Half Plane zero removal, Multipath Miller Zero Cancellation proves to be a robust solution, independent from variations of the current through the output transistor. This makes the technique especially suited for low-power class AB amplifiers. Such low-power circuits generally have a large quiescent-to-maximum output current ratio to obtain a high power efficiency.

Bibliography

[1] Bode, H.W., "Network Analysis and Feedback Amplifier Design", D. Van Nostrand Company Inc., New York, 1949.

[2] Klaassen, K.B., "Elektrotechnisch Meten", Delftse Uitgevers-maatschappij b.v., Delft, 1986.

[3] Dorf, R.C., "Modern Control Systems", Addison-Wesley Publishing Company, Reading, Massachusetts, 1986.

[4] Nyquist, H., "Regeneration Theory", *Bell System Technical Journal*, Jan. 1932.

[5] Strang, G, "Linear Algebra and its Applications", Harcourt Brace Jovanovich, San Diego, 1988.

[6] Cherry, E.M., Hooper, D.E., "Amplifying Devices and Low-pass Amplifier Design", John Wiley and Sons, Inc., New York, 1968.

[7] Carlson, A.B., "Communication Systems", McGraw-Hill Book Company, New York, 1988.

[8] Kamath, B.Y., Meyer, R.G., and Gray, P.R., "Relationship between Frequency Response and Settling Time of Operational Amplifiers", *IEEE Journal of Solid-State Circuits*, Vol. SC-9, Dec 1974, pp. 347-352.

[9] Miller, J.M., "Dependence of the Input Impedance of a Three-Electrode Vacuum Tube upon the Load in the Plate Circuit", *Nat. Bur. Stand. Sci. Papers*, No. 351, 367, 1919-1920.

[10] Tsividis, Y.P., "Operation and Modeling of the MOS Transistor", McGraw-Hill, New York, 1987.

[11] Solomon, J.E., "The Monolithic Opamp: a Tutorial Study", *IEEE Journal of Solid State Circuits*, vol. sc-9, no. 6, Dec 1974.

[12] Bababezhad, J.N., "A Low-Output-Impedance Fully Differential Op Amp with Large Output Swing and Continuous-Time Common-Mode Feedback", *IEEE Journal of Solid State Circuits*, vol. 26, no. 12, Dec. 1991.

[13] Huijsing, J.H., "Multi-stage amplifier with capacitive nesting for frequency compensation", U.S. Pat. Appl. Ser. No. 602234, filed April 19, 1984.

[14] Eschauzier, R.G.H., Kerklaan, L.P.T., Huijsing, J.H., "A 100-MHz 100-dB Operational Amplifier with Multipath Nested Miller Compensation Structure.", *IEEE Journal of Solid State Circuits*, vol. 27, no. 12, Dec. 1992.

[15] Eschauzier, R.G.H., Huijsing, J.H., "Multistage Amplifier with Hybrid Nested Miller Compensation", U.S. patent appl. ser. no. 93201779.1, filed Feb. 10, 1994.

[16] Eschauzier, R.G.H., Huijsing, J.H., "A Programmable 1.5V CMOS Class-AB Operational Amplifier with Hybrid Nested Miller Compensation for 120dB Gain and 6MHz UGF", *IEEE Journal of Solid-State Circuits*, vol. 29, no. 12, Dec. 1994.

[17] Cherry, E.M., "A High Quality Audio Power Amplifier", *Monitor-proceedings of the IREE Aust.*, Jan/Feb '78, 1-8.

[18] NE 5534 Data Sheet, Signetics, 1978; TDA 1034 Data Sheet, Philips, April 1976.

[19] Gray, P.R., "MOS Operational Amplifier Design - a Tutorial Overview", *IEEE Journal of Solid State Circuits*, vol. sc-17, no. 6, Dec. 1982.

[20] Huijsing, J.H., Fonderie, J., "Multi Stage Amplifier with Capacitive Nesting and Multi-path-Driven Forward Feeding for Frequency

Compensation", U.S. Pat. Appl. Ser. No. 654.855, filed February 11, 1991.

[21] Fonderie, J., Huijsing, J.H., "1-V Operational Amplifier with Multipath-Driven Output Stage", *IEEE Journal of Solid State Circuits*, vol. 26, no. 12, Dec. 1991.

[22] Huijsing, J.H., Eschauzier, R.G.H., "Amplifier Arrangement with Multipath Miller Zero Cancellation", U.S. pat. appl. ser. no. 08/197,529, filed June 21, 1993.

[23] Eschauzier, R.G.H., Huijsing, J.H., "An Operational Amplifier with Multipath Miller Zero Cancellation for RHP Zero Removal", *proceedings ESSCIRC 1993*, Editions Frontières, Gif-sur-Yvettes, France.

[24] Nauta, B., "Analog CMOS Filters for Very-High Frequencies", University of Twente, Enschede, 1991.

[25] Lebel, C.J., "Slewing Distortion in Digital-to-Analog Conversion", Journal of the Audio Engineering Society, vol. 25, no. 4, April '77, 178-183.

[26] Huijsing, J.H. and Tol F., "Monolithic operational amplifier design with improved HF behaviour", *IEEE Journal of Solid-State Circuits*, Vol. SC-11, April 1976, pp. 323-328.

[27] Seevinck, E., De Jager, W. and Buitendijk, P., "A low-distortion output stage with improved stability for monolithic power amplifiers", *IEEE Journal of Solid-State Circuits*, Vol. SC-23, June 1988, pp. 794-801.

[28] Meijer, G.C.M., "Integrated Circuits and Components for Bandgap References and Temperature Transducers", Internal Report Delft University of Technology, 1982

[29] Ahuja, B.K., "An Improved Frequency Compensation Technique for CMOS Operational amplifiers", *IEEE Journal of Solid State Circuits*, vol. sc-18, no. 6, Dec 1983.

[30] Hogervorst, R. et al., "CMOS Low-Voltage Operational Amplifier with Constant Gm Rail-to-Rail Input Stage", *proc. 1992 IEEE International Symposium on Circuits and Systems*, San Diego, May 1992.

[31] Allen, P.E., Holberg, D.R., "CMOS Analog Circuit design", Holt, Rinehart and Winston, Inc., Fort Worth, 1987.

Index