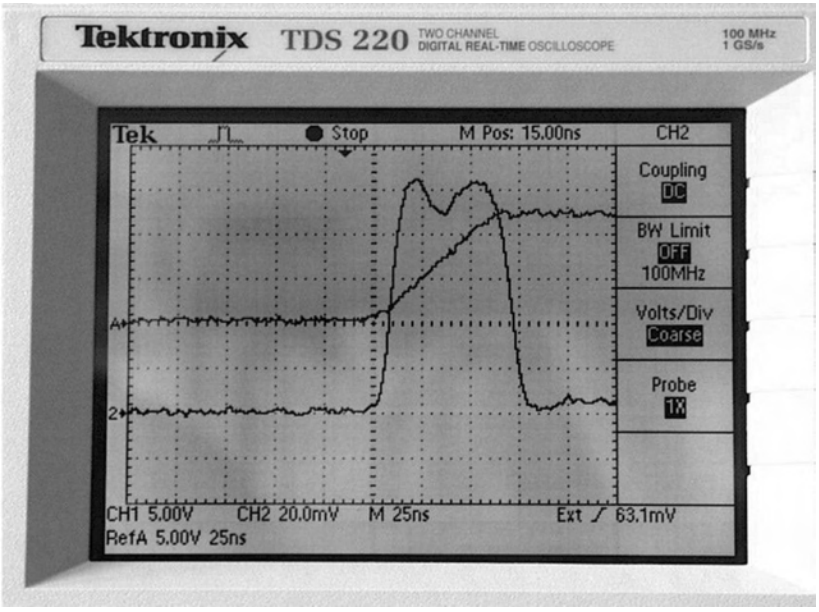
Analog Circuit Design Series Volume 3

Designing High-Performance Amplifiers



D. Feucht
Innovatia Laboratories

SciTech Publishing, Inc.
Raleigh, NC.

Published by SciTech Publishing, Inc.
911 Paverstone Drive, Suite B
Raleigh, NC 27615
(919) 847-2434, fax (919) 847-2568
scitechpublishing.com

Editor: Dudley R. Kay
Production Manager: Robert Lawless
Typesetting: SNP Best-set Typesetter Ltd., Hong Kong
Cover Design: Aaron Lawhon
Printer: Docusource

This book is available at special quantity discounts to use as premiums and sales promotions, or for use in corporate training programs. For more information and quotes, please contact the publisher.

Printed in the United States of America

10 9 8 7 6 5 4 3

ISBN: 9781891121845
Series ISBN: 9781891121876

Library of Congress Cataloging-in-Publication Data
Feucht, Dennis.
Designing high-performance amplifiers / D. Feucht.
p. cm. – (Analog circuit design series ; v. 3)
Includes bibliographical references and index.
ISBN 978-1-891121-84-5 (pbk. : alk. paper) – ISBN 978-1-891121-87-6 (series)
1. Amplifiers (Electronics)–Design and construction. 2. Electronic circuit design. I. Title.
TK7871.2.F478 2010
621.3815′35–dc22 2009028290

Preface

Solid-state electronics has been a familiar technology for almost a half century, yet some circuit ideas, like the transresistance method of finding amplifier gain or identifying resonances above an amplifier's bandwidth that cause spurious oscillations, are so simple and intuitively appealing that it is a wonder they are not better understood in the industry. I was blessed to have encountered them in my earlier days at Tektronix but have not found them in engineering textbooks. My motivation in writing this book, which began in the late 1980s and saw its first publication in the form of a single volume published by Academic Press in 1990, has been to reduce the concepts of analog electronics as I know them to their simplest, most obvious form, which can be easily remembered and applied, even quantitatively, with minimal effort.

The behavior of most circuits is determined most easily by computer simulation. What circuit simulators do not provide is knowledge of what to compute. The creative aspect of circuit design and analysis must be performed by the circuit designer, and this aspect of design is emphasized here. Two kinds of reasoning seem to be most closely related to creative circuit intuition:

1. Geometric reasoning: A kind of visual or graphic reasoning that applies to the topology (component interconnection) of circuit diagrams and to graphs such as reactance plots.
2. Causal reasoning: The kind of reasoning that most appeals to our sense of understanding of mechanisms and sequences of events. When we can trace a chain of causes for circuit behavior, we feel we understand how the circuit works.

These two kinds of reasoning combine when we try to understand a circuit by causally thinking our way through the circuit diagram. These insights, obtained

by *inspection*, lie at the root of the quest. The sought result is the ability to write down accurate circuit equations by inspection. Circuits can often be analyzed multiple ways. The emphasis of this book is on development of an intuition into how circuits work with a perspective that can be applied more generally to circuits of the same class.

The previous two volumes of this Analog Circuit Design series, *Designing Amplifier Circuits* and *Designing Dynamic Circuit Response*, provide the needed concepts for the investigation of fast and precise, or *Designing high-performance*, amplifiers as covered in this third volume. Much of the technology appearing here has been derived from the tradition of excellence in engineering found at Tektronix in the development of oscilloscopes and other electronic test and measurement instruments. It begins by considering the parameters of merit by which amplifier performance can be assessed and then the optimization of stage gain in a multistage amplifier of limited stage bandwidth. The method of characterizing amplifier performance in the complex-frequency or *s*-domain is continued, along with an intuitive method for determining poles of circuits without grinding through too much algebra. Bandwidth extension techniques using inductors is next, including a little-known method originating at Tektronix, applied notably to oscilloscope products by Carl Battjes, Bob Ross, and John Addis, known as *T-coil* peaking – another marvelous performance enhancer that has yet to become widely known. Some circuits deserve special attention, such as the source follower and shunt-feedback amplifier.

After concentrating on speed, the topics turn toward the other dimension of high performance, that of precision. Noise and distortion are topics as are the circuits of precision, operational amplifiers (op-amps), and instrumentation amplifiers. The bipolar junction transistor (BJT) differential pair is given more detailed treatment, as is thermal distortion, a topic that must be considered when setting the static operating point (or *bias*) of amplifier circuits. The often used complementary emitter-follower as a power driver output stage is given particularly detailed analysis and simplification so that the various oddities in its response have a simple explanation.

After this, speed and precision are combined to produce the highest in amplifier performance, deserving of a full chapter. It starts with well-known current-input and current-feedback amplifiers and then covers another lesser-known topic: composite amplifiers. These have multiple paths to the output and are

introduced first in block diagram form and then with some wonderful circuits, many originating at Tektronix. At center stage of these methods are feedforward amplifiers and Barrie Gilbert's translinear circuits, a conceptual breakthrough during the 1960s. Translinear circuits are simplified by reducing what historically was exponential math to ratios of sums and differences, and hopefully more engineers will venture into their use, for they are "powerful." Buffers receive special treatment, and a variety of other nifty circuits are also presented. Multipliers and programmable-gain amplifiers (PGAs) end this volume.

Much of what is in this book must be credited in part to others from whom I picked up essential ideas about circuits at Tektronix, mainly in the 1970s. I am particularly indebted to Bruce Hofer, a founder of Audio Precision Inc.; Carl Battjes, who founded and taught the Tek Amplifier Frequency and Transient Response (AFTR) course; Laudie Doubrava, who investigated power supply topics; and Art Metz, for his clever contributions to a number of designs, some extending from the seminal work on translinear circuits by Barrie Gilbert, also at Tek at the same time. Then there is Jim Woo, who, like Battjes, was another oscilloscope vertical amplifier designer; Ian Getreu and Bob Nordstrom, from whom I learned transistors; and Mike Freiling, an artificial intelligence researcher in Tektronix Laboratories whose work in knowledge representation of physical systems influenced my broader understanding of electronics.

In addition, in no particular order, are Fred Beckett, Lee Jalovec, Wayne Kelsoe, Cal Diller, Marv LaVoie, Keith Lofstrom, Peter Staric, Erik Margan, Tim Sauerwein, George Ermini, Jim Geddes, Carl Hollingsworth, Chuck Barrows, Dick Hung, Carl Matson, Don Hall, Phil Crosby, Keith Ericson, John Taggart, John Zeigler, Mike Cranford, Allan Plunkett, Neldon Wagner, and Paul Magerl. These and others I have failed to name have contributed personally to my knowledge as an engineer and indirectly to this book. Most of all, I am indebted to the creator of our universe, who made electronics possible. Any errors or weaknesses in this book, however, are my own.

Contents

1

Wideband Amplification

Some amplifiers are performance limited mainly by speed. Oscilloscope vertical amplifiers, pulse and function generator output amplifiers, and video and nuclear signal-processing amplifiers are often speed limited. Fast amplifiers are usually open-loop, limited-gain stages, such as those analyzed at low frequency in *Designing Amplifier Circuits*, with one or two transistors per stage. New techniques have increased the speed of operational amplifer (op-amp) circuits, but the fastest amplifiers consist of limited-gain stages. For the fastest speed, these amplifiers are integrated to reduce parasitic reactances. Examined first is the strategy of amplifier design before analyzing various bandwidth extension techniques.

Multiple-Stage Response Characteristics

The speed of an amplifier can be expressed by its response to a step input. For a single-pole amplifier with pole $p = 1/\tau$, the response to a unit step input can be characterized by its risetime. A single-pole amplifier has a transfer function of the form

$$A(s) = K \cdot \frac{1}{s/p + 1}$$

with pole at $-p$. The pole is also at the bandwidth

$$\omega_{bw} = p$$

From the risetime formula, the single-pole risetime is

$$t_r = \tau \cdot \ln(9) \cong 2.2 \cdot \tau = \frac{2.2}{p} \cong \frac{0.35}{f_{bw}}$$

The unit step response is

$$v_{step}(t) = K \cdot \left(1 - e^{-pt}\right)$$

These equations assume a linear amplifier (or that small-signal analysis is valid). Large-signal amplifier behavior occurs when a waveform reaches the limit of its linear range. What often results is waveform slewing, in which waveform change is rate limited and characterized by its *slew rate*. The maximum slew rate of a waveform is determined by its maximum instantaneous slope and amplitude. The *full-power bandwidth* f_{BW} of an amplifier is related to the maximum slew rate of a sinusoid that spans the dynamic range of the amplifier. Differentiating the sinusoid and solving for the maximum value,

$$\text{maximum slew rate} = \frac{d}{dt}(V_m \cdot \sin \omega_{BW} t)\big|_{\max} = \omega_{BW} \cdot V_m$$

The sinusoid changes over its full range in the slewing time:

$$t_{slew} = \frac{2V_m}{\omega_{BW} V_m} \cong \frac{0.32}{f_{BW}}$$

When $f_{bw} = f_{BW}$, t_{slew} is nearly the same as the risetime given above. A more general comparison of large- and small-signal risetime follows by finding the time it takes to slew from 10% to 90% of the final value. This time is

$$\text{slewing } t_r = \frac{(0.8) \cdot 2 \cdot V_m}{\omega_{BW} \cdot V_m} \cong \frac{0.26}{f_{BW}}$$

When an amplifier operates with some slewing, the bandwidth lies between f_{BW} and f_{bw} where always $f_{BW} < f_{bw}$.

Another quantity that characterizes speed is time delay t_d, defined as the time that the response to a unit step input takes to reach half of its final value. It is found by setting v_{step} to 0.5 and solving for t:

$$t_d = \tau \cdot \ln(2) = \frac{\ln 2}{p} \cong \frac{0.69}{p} \cong \frac{0.11}{f_{bw}}$$

Delay time is useful for measuring the propagation delay of linear logic circuits, such as logic gates.

Fast amplifiers usually consist of several cascaded gain stages. An amplifier with n single-pole stages with poles at $-p$ has a transfer function of the form

$$A(s) = K \cdot \frac{1}{(s/p + 1)^n}$$

The unit step response of $A(s)$ is

$$n\text{-stage } v_{step} = \mathcal{L}^{-1}\left\{A(s) \cdot \frac{1}{s}\right\} = Kp^n\left(1 - e^{-pt}\sum_{k=0}^{n-1}\frac{t^k}{k!}\right)$$

Calculation of the risetime from this equation can be difficult. A simpler alternative is to derive expressions for bandwidth. The bandwidth of a single-pole amplifier is the pole frequency, or

$$\omega_{bw} = p$$

The magnitude at bandwidth of a single-pole stage with static gain of K is found by setting ω to p:

$$\|A(jp)\| = K \cdot \frac{1}{\sqrt{(\omega/p)^2 + 1}}\Big|_{\omega = p} = K \cdot \frac{1}{\sqrt{2}}$$

An n-stage amplifier with single-pole stages has a transfer function of the form

$$n\text{-stage } A(s) = K \cdot \frac{1}{(s/p + 1)^n}$$

When the magnitude of $A(j\omega)$ has rolled off to that of a single-pole stage, as in $\|A(jp)\|$, this is the n-stage bandwidth:

$$\|A(j\omega)\| = K \cdot \frac{1}{\left[(\omega/p)^2 + 1\right]^{n/2}} = K \cdot \frac{1}{\sqrt{2}}$$

Solving for $\omega = \omega_{bw}$, and expressing it as a fraction of p by defining the *bandwidth shrinkage* or *bandwidth reduction factor*, S, then

$$S(n) = \frac{\omega_{bw}}{p} = \sqrt{2^{1/n} - 1}$$

Fast amplifiers usually have stages with quadratic pole factors in their transfer functions. The same kind of derivation of $S(n)$ assumes an n-stage amplifier transfer function of the form

$$A(s) = K \cdot \frac{1}{\left(s^2\tau_n^2 + 2\zeta\tau_n s + 1\right)^n}$$

Setting the magnitude of $A(s)$ to that of a single-pole amplifier at bandwidth results in the expression

$$\|A(j\omega)\| = K \cdot \left(\frac{1}{\left(1 - \tau_n^2\omega^2\right)^2 + (2\zeta\tau_n\omega)^2} \right)^{n/2} = K \cdot \frac{1}{\sqrt{2}}$$

The pole factor of $A(s)$ does not contain p, and the bandwidth is related to the single-pole stage by ω_n. (See *Designing Dynamic Circuit Response*, "Optimization of Time-Domain and Frequency-Domain Response" for derivation of bandwidth.) Solving for ω in terms of ω_n, the result is

$$S(n) = \frac{\omega_{bw}}{\omega_n} = \left(1 - 2\zeta^2 + \sqrt{4\zeta^4 + 4\zeta^2 + 2^{1/n}}\right)^{1/2}$$

For n critically damped stages,

$$S(n) = \sqrt{2^{1/2n} - 1}, \quad \zeta = 1$$

A one-stage critically damped amplifier has $S = 0.64$, whereas for four stages S is about 0.3. For maximally flat envelope delay (MFED) stages ($\zeta = \sqrt{3}/2$), with $n = 1$, $S \cong 0.786$; with $n = 4$, $S \cong 0.4$. When the number of stages increases to 10, $S \cong 0.25$. For MFED response, n stages have the approximate S of

$$S_{\text{MFED}} \cong \frac{0.786}{\sqrt{n}}$$

With these developments, we return to consider the risetime of multistage amplifiers. The transfer-function magnitude for a general amplifier of n poles is of the form

$$\|A(j\omega)\| = \frac{K}{\left[1 + \omega^2 \sum_i (1/p_i^2) + \omega^4 \sum_i \sum_j (1/p_i^2 p_j^2) + \cdots\right]^{1/2}} \cong \frac{K}{\left(1 + \omega^2 \sum_i \tau_i^2\right)^{1/2}}$$

where $1/p_i = \tau_i$. The higher-order terms in ω are negligible for widely separated poles at much higher frequencies. The sum of time constants in the ω^2 term can be regarded as an equivalent single-pole time constant of

$$\tau \cong \sqrt{\sum_i \tau_i^2}$$

From t_r, an approximate risetime is therefore

$$t_r \cong 2.2\tau = 2.2\sqrt{\sum_i \tau_i^2} = \sqrt{\sum_i (2.2\tau_i)^2} = \sqrt{\sum_i t_{ri}^2}$$

In other words, the approximate risetime of a multistage amplifier is the square root of the sum of the squares of the risetimes of the individual stages. For n identical stages, risetime degrades by approximately $\sqrt{n}$ that of a single stage.

Example: Oscilloscope Risetime

A 100 MHz oscilloscope has a probe with a 2 ns risetime. The total risetime is found by first calculating the risetime of the oscilloscope. Applying the t_r equation,

$$t_r \cong \frac{0.35}{10^8 \text{ Hz}} = 3.5 \text{ ns}$$

Then the total risetime is approximately

$$t_r \cong \sqrt{(3.5 \text{ ns})^2 + (2 \text{ ns})^2} = 4 \text{ ns}$$

An accompanying approximation to bandwidth can also be made since τ is an equivalent single-pole time constant. From single-pole bandwidth,

$$\omega_{bw} \cong \frac{1}{\tau} = \sqrt{\sum_i \frac{1}{p_i^2}}$$

For n repeated poles, bandwidth, like risetime, degrades by approximately $\sqrt{n}$.

Amplifier Stage Gain Optimization

As the number of amplifier stages increases, bandwidth decreases. For a fast-amplifier design strategy, therefore, the number of stages should be minimized. However, most amplifier designs also require a given overall gain. Reducing the stage count demands increased gain per stage. Amplifier stages have a gain-bandwidth product, f_T, affected mainly by the active device. An increase of stage gain decreases stage bandwidth. An optimum stage gain, $A_1(s)$, that maximizes amplifier bandwidth, ω_{bw}, for a given amplifier gain, $A(s)$, is derived by first noting that

$$\omega_{bw} = S \cdot \omega_1$$

where ω_1 is the single-stage bandwidth. Then the gain-bandwidth product of the amplifier is expressed in relation to its stages as

$$A^{1/n} \cdot \omega_{bw} = A_1 (S \cdot \omega_1)$$

assuming the n-stages have the same gain,

$$A_1 = A^{1/n}$$

Solving for ω_{bw} yields

$$\omega_{bw} = A_1 \cdot A^{-1/n} \cdot S(n) \cdot \omega_1$$

The optimum number of stages is found by differentiating ω_{bw} with respect to n to maximize bandwidth:

$$\frac{d}{dn} \omega_{bw} = A_1 \cdot \omega_1 \cdot \frac{d}{dn} \left[S(n) \cdot A^{-1/n} \right]$$

For single-pole stages, S is substituted. S can be expressed differently by noting that

$$2^{1/n} = e^{\ln(2)/n} = \sum_{k=0}^{\infty} \frac{(\ln 2/n)^k}{k!} \cong 1 + \frac{\ln 2}{n}$$

Then

$$S(n) \cong \left(\frac{\ln 2}{n} \right)^{1/2} = \sqrt{\ln 2} \cdot n^{-1/2}$$

Substituting for S in the derivative above, the right side becomes

$$\begin{aligned} & A_1 \cdot \omega_1 \cdot \frac{d}{dn} \left(A^{-1/n} \cdot \sqrt{\ln 2} \cdot n^{-1/2} \right) \\ & = A_1 \cdot \omega_1 \sqrt{\ln 2} \cdot \left(A^{-1/n} \cdot \left(-\frac{1}{2} \right) \cdot n^{-3/2} + A^{-1/n} \cdot n^{-1/2} \cdot \frac{1}{n^2} \cdot \ln A \right) \end{aligned}$$

To find the optimum number of stages n_{opt}, set the derivative to zero and solve. Then

$$n_{opt} = 2 \cdot \ln A$$

The optimum stage gain is

$$\text{optimum } A_1 = A^{1/n_{opt}} = A^{1/2\ln A} = e^{1/2} = \sqrt{e} \cong 1.65$$

This is not a large voltage or current gain. In practice, the optimum gain is somewhat larger than this value, usually around 2 to 3, due to bandwidth loss from interstage coupling.

Multistage amplifier frequency-response magnitude approaches a gaussian function as the number of stages increase. This gaussian response is quickly approached in practice by a few stages. It is derived by first rewriting $\|A(j\omega)\|$ in terms of ω_{bw} from S. Given $p = \omega_{bw}/S$, then

$$\|A(j\omega)\| = \frac{K}{\left[(\omega/\omega_{bw})^2\left(2^{1/n}-1\right)+1\right]^{n/2}}$$

Next, as $n \to \infty$ in $2^{1/n}$, and using the first two terms of the exponential series expansion,

$$2^{1/n} - 1 \to \frac{\ln 2}{n}, \quad n \to \infty$$

Substituting this into $\|A(j\omega)\|$, then

$$\|A(j\omega)\| \cong \frac{K}{\left[1+(\omega/\omega_1)^2(\ln 2/n)\right]^{n/2}} = \frac{K}{\left[1+(2/n)(\omega/\omega_1)^2(\ln 2/2)\right]^{n/2}}$$

$\|A(j\omega)\|$ is of the exponential form,

$$e^x = \lim_{x\to\infty}\left(1-\frac{x}{n}\right)^{-n}$$

So as $n \to \infty$,

$$\|A(j\omega)\| = K \cdot e^{-(\ln 2/2)\cdot\left(\frac{\omega}{\omega_1}\right)^2}, \quad n \to \infty$$

This e^{x^2} form of $\|A\|$ is the gaussian response function.

The maximum achievable bandwidth of an amplifier with a gaussian response is derived based on the unity-power-gain frequency f_{MAX}. If $\|A\|$ is a power gain, then it can be expressed in decibel scaling as

$$\|A(j\omega)\|_{\text{dB}} = 10 \cdot \log K - 10 \cdot \left(\frac{\ln 2}{2}\right) \cdot (\log e) \cdot \left(\frac{f}{f_1}\right)^2 = \|A(0)\|_{\text{dB}} - c \cdot \left(\frac{f}{f_1}\right)^2$$

where c reduces to

$$c = 5 \cdot \log 2 \cong 1.51$$

The maximum bandwidth is achieved when $\|A\|$ passes through f_{MAX} at unity power gain (0 dB). From $\|A(j\omega)\|_{\text{db}}$,

$$\|A(0)\|_{\text{dB}} - c \cdot \left(\frac{f_{MAX}}{f_1}\right)^2 = 0$$

Solving for c and substituting it into $\|A(j\omega)\|_{\text{db}}$,

$$\|A\|_{\text{dB}} = \|A(0)\|_{\text{dB}} \cdot \left(1 - \frac{f^2}{f_{MAX}^2}\right)$$

The bandwidth is the frequency at which $\|A\|_{\text{dB}}$ has rolled off by –3 dB:

$$\|A\|_{\text{dB}} - \|A(0)\|_{\text{dB}} = -3 = \|A(0)\|_{\text{dB}} \cdot \left(-\frac{f^2}{f_{MAX}^2}\right)$$

The solution of this equation for f is the power-gain bandwidth f_{bw} and is

$$f_{bw} = f_{MAX} \cdot \sqrt{\frac{3}{\|A(0)\|_{\text{dB}}}}$$

Example: Oscilloscope Vertical Amplifier

A wideband analog oscilloscope has a vertical deflection sensitivity of 2 V/cm at the cathode-ray tube (CRT) and a deflection plate termination resistance of 350 Ω. The input sensitivity is 50 mV/div into 100 Ω from the source when terminated by the 50 Ω scope input. The power gain is

$$\|A(0)\|_{\text{dB}} = 10 \cdot \log\left(\frac{(2\,\text{V/div})^2/350\,\Omega}{(50\,\text{mV/div})^2/100\,\Omega}\right) \simeq 26.6\,\text{dB}$$

The f_{MAX} is 2 GHz. The maximum f_{bw} is

$$f_{bw} = 2\,\text{GHz} \cdot \sqrt{\frac{3\,\text{dB}}{26.6\,\text{dB}}} = 672\,\text{MHz}$$

The actual bandwidth of the amplifier (without the CRT) is 550 MHz. The use of maximum bandwidth as a performance index can be taken as the ratio of actual to maximum theoretical bandwidth, or

$$\frac{550\,\text{MHz}}{672\,\text{MHz}} \cong 82\%$$

That is, 82% of the maximum achievable bandwidth is realized in the vertical amplifier, within 18% of the theoretical limit.

Pole Determination by Circuit Inspection

Circuit complexity increases as the number of reactive elements increases and makes derivation of the transfer function more difficult. At some complexity

threshold (which varies among engineers), the urge to simulate the circuit by computer overwhelms the desire to achieve an intuitive understanding of it. Even for complexity that requires simulation, it is necessary to know what to simulate. Until circuit design is computerized, the choice of numeric values of circuit elements must be based on estimation techniques and qualitative reasoning.

Most circuits can be decomposed into *modules* with well-defined interfaces. Intrastage behavior is relatively free of interaction with other noncoupled stages. Interaction among modules can be considered apart from interaction within modules. The dynamic behavior of each circuit module can thus be determined individually, reducing the complexity of analysis.

A technique described by Cochrun and Grabel (1973) and streamlined by Rosenstark (1986) makes estimation of pole locations in active *RC* circuits simpler than solving the circuit for the transfer function. The degree of the characteristic equation (the transfer function denominator set to zero) equals the number of poles and the number of reactive circuit elements. Each capacitor in an *RC* circuit is associated with a pole. The characteristic equation in a normalized transfer function can be written as

$$D(s) = a_n s^n + a_{n-1} s^{n-1} + \cdots + a_1 s + 1 = 0$$

The technique allows determination of the a_n from inspection of the circuit. The coefficients, in terms of circuit elements, are found as follows. The equation for a_1 is

$$a_1 = \sum_{i=1}^{n} R_i(\text{open}) \cdot C_i = \sum \tau_i(\text{open})$$

a_1 is calculated as a procedure as follows

a_1 Procedure

1. Order the C by numbering them.
2. For each C, beginning with C_1, find the equivalent resistance across its terminals with all other C open. This is R_i (open). Multiply R_i (open) by C_i for τ_i (open).
3. Sum the τ_i (open) to obtain a_1.

This procedure is expedited by writing the time constants in the first column of a table beginning with τ_1 (open) in the top row and τ_n (open) in the bottom row.

Next, for a multi-capacitor circuit, a_2 is needed and is

$$a_2 = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} R_i(\text{open}) \cdot C_i \cdot R_j(C_i \text{ shorted}) \cdot C_j = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \tau_i(\text{open}) \cdot \tau_j(C_i \text{ shorted})$$

The procedure for a_2 continues the table by filling in the second column and then using the τ_i (open) from the first column. All C not shorted are open when R is being found.

a_2 Procedure

1. For each C_i, do the following:
   a. Short C_i. For each C after C_i (in the order they were numbered in the a_1 procedure), find the terminal R for C (C_j). Multiply this $R_j(C_i$ shorted) by C_j for $\tau_j(C_i$ shorted).
   b. Multiply the $\tau_j(C_i$ shorted) by τ_i (open).
2. Add the time-constant products from step 1b to obtain a_2.

Each entry in the first column of the table from the a_1 procedure will have $n - i$ entries in the second column for each C_j.

For a_3, three summations are made, extending the a_2 procedure. For the third column, two capacitors are shorted at a time (indices i and j), and k is indexed:

$$a_3 = \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^{n} R_i(\text{open}) \cdot C_i \cdot R_j(C_i \text{ shorted}) \cdot C_j \cdot R_k(C_i, C_j \text{ shorted}) \cdot C_k$$

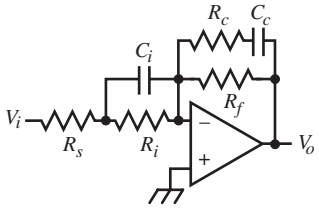
A way to keep the indexing straight is to base the entire procedure around the time-constant table. This *Rosenstark table* for three capacitances is sketched below.

	a_1	a_2	a_3
C_1	τ_1 (open)	τ_2 (C_1 shorted)	τ_3 (C_1, C_2 shorted)
		τ_3 (C_1 shorted)	
C_2	τ_2 (open)	τ_3 (C_2 shorted)	
C_3	τ_3 (open)		

The column numbering is for the i index. The procedure amounts to filling in the table and then, for a_1, summing the τ_i in column 1; for a_2, summing the products of τ in column 2 with τ from column 1; for a_3, summing the products of τ in column 3 with τ from columns 2 and 1. The summation always involves the τ of a column multiplied by the τ of the columns to the left. When a τ is found, the capacitors indexed in the columns to the left are shorted.

Active devices can change the resistance at a capacitor, for example, due to the Miller effect, and must be taken into account when finding equivalent resistances.

Example: Op-Amp Circuit Poles from the Cochrun-Grabel Method



$$A_v = -\frac{R_f}{R_i + R_s}\,\frac{(sR_cC_c+1)(sR_iC_j+1)}{(s[R_i \| R_s]C_i+1)(s[R_f+R_c]C_c+1)}$$

This op-amp circuit, with transfer function as shown, can be analyzed using the Cochrun-Grabel method. The Rosenstark table is given below.

	a_1	a_2
C_i	$(R_i \parallel R_s)\cdot C_i$	$(R_c + R_f)\cdot C_c$
C_c	$(R_c + R_f)\cdot C_c$	

The ordering of capacitors, as shown in the table, is C_i, C_c. This ordering is arbitrary. The a_1 column is filled in, beginning at the C_i row, by finding the open-circuit resistance across the terminals of C_i. An active device (i.e., the op-amp) requires first a determination of its effect on resistance. The V_- input is a virtual ground for the ideal op-amp. Knowing this, the C_i terminals have across them $R_i \| R_s$ because V_i has zero resistance. The time constant for the first entry, C_i row, a_1 column, is complete. For the C_c row, a_1 column, examine C_f. R_c is in series with the C_c terminal and goes to ground. From the other terminal, we determined that the V_o node has a resistance of R_f to R_c. Thus, the total resistance across C_c is $R_c + R_f$. The second entry is complete.

Now, begin with the a_2 column. Short C_i (from the first column) and determine resistance across C_c. Again this is $R_c + R_f$. The table is complete. The a_n are now found from the table. The sum of the first column is

$$a_1 = (R_i \| R_s) \cdot C_i + \left(R_c + R_f\right) \cdot C_c$$

Then a_2 is found from the second column by multiplying its entries by the first column and adding them. Since there is only one, no addition is needed here, and

$$a_2 = \left(R_c + R_f\right) \cdot C_c \cdot (R_i \| R_s) \cdot C_i$$

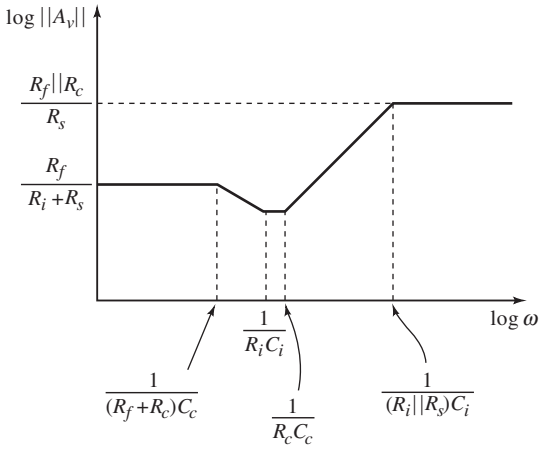
The characteristic equation is

$$a_2 s^2 + a_1 s + 1 = 0$$

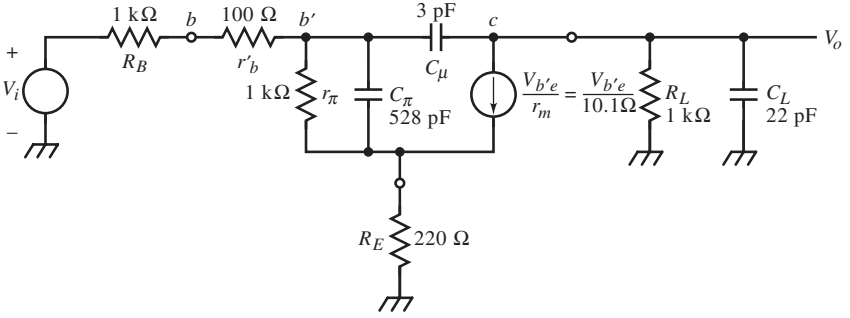
This quadratic equation is easily solved by noting that the a_1 terms are the a_2 factors, and the poles are therefore

$$p_1 = \frac{-1}{(R_i \| R_s) \cdot C_i}, \quad p_2 = \frac{-1}{\left(R_c + R_f\right) \cdot C_c}$$

This result agrees with the reactance chart method, as shown below.



Example: BJT Amplifier Poles from the Cochrun-Grabel Method

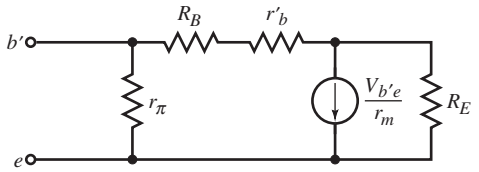


(a)

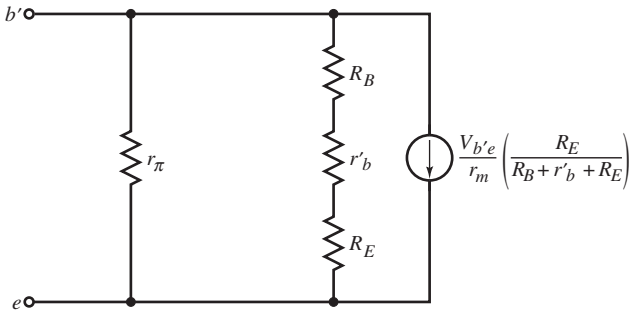
	a_1	a_2	a_3	ns units
C_π	28.9	6.30	3.41	
		22.0	a_3 = 621 ns^3	
C_μ	19.7	0.479		
C_L	22.0	a_2 = 827 ns^2		
	a_1 = 70.6 ns			

(b)

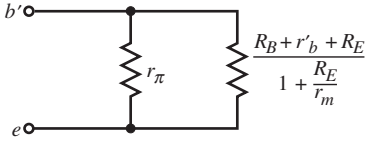
The poles of the BJT amplifier of (a) are found by the Cochrun-Grabel method using the Rosenstark table in (b) to obtain the characteristic equation. The third-degree equation can be solved by computer or approximated by a lower-degree equation that retains an approximation of the dominant poles.



Begin with R of C_π in the first column. With all the other Cs open, the resistance is found by the following steps. First, the equivalent circuit is Nortonized.



Then the substitution theorem is applied and two branches are combined.

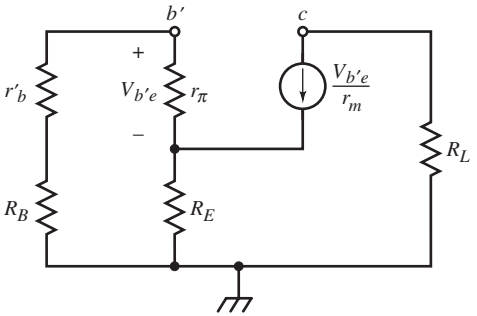


The resistance is

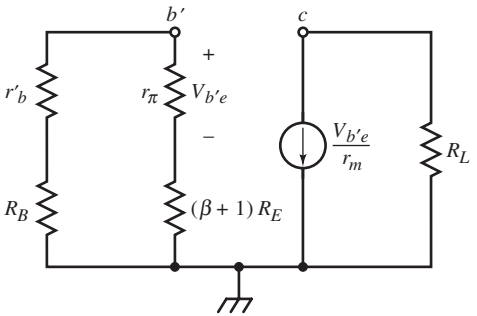
$$R_\pi = r_\pi \left\| \frac{(R_E + R_B + r_b')}{(1 + R_E/r_m)} \right.$$

This has the form of Miller's theorem, in which $K = R_E / r_m$. $R_\pi = 54.8\ \Omega$, and the resulting time constant is 28.9 ns.

For C_μ, the equivalent circuit is shown below.



Applying the β transform to R_E results in the equivalent circuit:



The resistance at the b' node to ground is

$$R_{b'} = (r_b' + R_B) \| (\beta + 1) \cdot (r_e + R_E)$$

The collector current source is controlled by $V_{b'e}$. From the base loop,

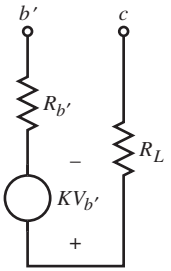
$$V_{b'e} = V_{b'} \cdot \left(\frac{r_\pi}{r_\pi + (\beta + 1) \cdot R_E} \right) = V_{b'} \cdot \left(\frac{r_e}{r_e + R_E} \right)$$

At the collector,

$$V_c = -R_L \cdot \frac{V_{b'e}}{r_m} = -\alpha \cdot \frac{R_L}{r_e + R_E} \cdot V_{b'} = -K \cdot V_{b'}$$

Current injected into the b' node causes the voltage across $b'c$ to change by $1 + V_c / V_{b'} = 1 + K$. This Miller effect causes the voltage across the injecting current source to be larger by $1 + K$ times. This makes the effective resistance $1 + K$ times larger also.

Next, change the Norton circuit of the collector loop to a Thevenin source as shown below using the collector equation.



The voltage across $R_{b'}$ is now $(1 + K)\ V_{b'}$, making $R_{b'}$ appear $(1 + K)$ times larger. This Miller resistance is in series with R_L. Thus, the resistance sought is

$$R_\mu = R_{b'} \cdot (1 + K) + R_L = (r_b' + R_B) \| (\beta + 1) \cdot (r_e + R_E) \cdot \left[1 + \alpha \cdot \frac{r_e}{r_e + R_E} \right] + R_L$$

Substituting circuit values gives $R_\mu = 6.57$ kΩ and $\tau = 19.7$ ns.

The time constant for a_1 due to C_L is $R_L \cdot C_L = 22.0$ ns. For the a_2 column, the first capacitance C_π is shorted, and R across $b'c$ is again determined. With C_π shorted, $V_{b'e} = 0$, and the transistor current source is nulled. This simplifies the resistance to

$$R_\mu (C_\mu \text{ shorted}) = (r_b' + R_B) \| R_E + R_L$$

which is 1.18 kΩ. The time constant is 6.30 ns. Next, C_π remains shorted as we find the resistance across C_L. It is R_L, and $\tau = 22.0$ ns. The last entry in the a_2

column is found by shorting C_μ and finding the resistance across C_L. The collector current source is now across the $b'e$ branch, and the substitution theorem reduces it to r_m. Applying $r_m \| r_\pi = r_e$,

$$R_L(C_\mu \text{ shorted}) = R_L \| (r_e + R_E) \| (r_b' + R_B)$$

This resistance is 160 Ω and $\tau = 0.479$ ns.

The final entry, for a_3, is the resistance across C_L with C_π and C_μ shorted.

$$R_L(C_\pi, C_\mu \text{ shorted}) = (r_b' + R_B) \| R_E \| R_L$$

This value is 155 Ω and $\tau = 3.41$ ns. The table is complete.

Next, find the a_n as follows:

$$a_1 = (28.9 + 19.7 + 22.0)\,\text{ns} = 70.6\,\text{ns}$$

$$a_2 = [(6.30)(28.9) + (22.0)(28.9) + (0.479)(19.7)]\,\text{ns}^2 = 827\,\text{ns}^2$$

$$a_3 = (3.41)(6.30)(28.9)\,\text{ns}^3 = 621\,\text{ns}^3$$

The characteristic equation is therefore

$$(621\,\text{ns}^3)\cdot s^3 + (827\,\text{ns}^2)\cdot s^2 + (70.6\,\text{ns})\cdot s + 1 = 0$$

This equation was solved by computer, producing real roots at

$$-2.83\,\text{MHz}, \quad -11.6\,\text{MHz}, \quad -198\,\text{MHz}$$

A SPICE simulation shows a damped response due to the dominant real pole with magnitude roll-off of –3 dB at 2.7 MHz. The two slowest poles combine to yield an approximation bandwidth of 2.75 MHz.

If a computer is not used to solve the characteristic equation for the poles, some approximations can be made by ignoring higher-degree terms. By dropping the s^3 term, two poles are found at 2.85 MHz and 10.7 MHz. By dropping the quadratic term also, the single pole is at 2.25 MHz, a 16% error. This error is acceptable for many pole estimates and leads to a simplified version of the Cochrun-Grabel approach; instead of building a table, build only the first

column. That is, sum the open-circuit time constants for each capacitor and invert it for the radian pole frequency. The hard work is in finding R_π and R_μ, but we have done that already, and the equations for them can be used for BJT analysis (and, for field-effect transistors, with the BJT-to-FET transform) generally.

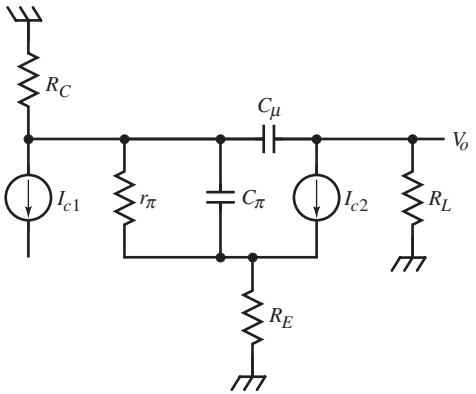
The f_T specified by transistor manufacturers is defined as the frequency at which β is one with the collector dynamically (ac) shorted to the emitter. Then C_μ shunts C_π. This implies that

$$\text{mfg } f_T = \frac{1}{2\pi\tau_T}, \quad \tau_T = r_\pi \cdot (C_\pi + C_\mu) \cong r_\pi \cdot C_\pi$$

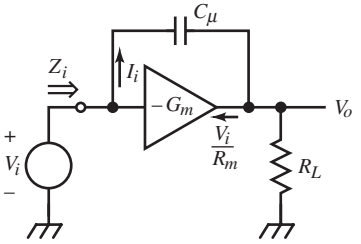
The manufacturer's f_T for the transistor in the previous example is 300 MHz and $\beta_o = 99$. For the single-pole BJT model (see *Designing Dynamic Circuit Response*, "Derivation of BJT High-Frequency Model") f_T is defined with $C_\mu = 0$ to make the resulting theory simpler. This should cause no problem if the manufacturer's f_T and C_μ (given as C_{ob}) values are used to compute f_β and f_T as defined here. In most cases, f_T is close enough already. In this example, the error is 0.6%.

The Cochrun-Grabel method produces only poles. One technique for determining zeros begins by first writing the nodal equations of the circuit. A flow graph is especially helpful here. Those transmittances that lead from input to output are examined for evidence of zeros.

Pole estimation is often applied to interstage coupling, to the pole formed by the load resistance of a common-emitter (CE) or common-base (CB) stage and the following CE stage input.



The input impedance of the loading BJT stage Z_i has branches through C_μ and Z_π.



Consider first the branch involving C_μ. The circuit is idealized above to eliminate the effect of the Z_π path. The BJT is represented by a transconductance amplifier. Its output current, representing collector current, is shown flowing into the amplifier with value $G_m \cdot V_i = V_i/R_m$. The input current to the amplifier flows into C_μ and is

$$I_i = \frac{V_i - V_o}{1/sC_\mu} = sC_\mu \cdot (V_i - V_o) = \frac{V_o}{R_L} + \frac{V_i}{R_m}$$

Solving this Kirchhoff's current law (KCL) equation for V_o, substituting it into the first expression for I_i, and solving for V_i/I_i,

$$Z_{i\mu} = \frac{1}{s[1 + (R_L/R_m)] \cdot C_\mu} + \frac{R_L}{[1 + (R_L/R_m)]} = \frac{1}{s(K_v + 1) \cdot C_\mu} + \frac{R_L}{K_v + 1}$$

Besides the Miller capacitance, R_L is reduced by $K_v + 1$. For large voltage gain, this branch presents a nearly capacitive impedance.

The impedance through Z_π is

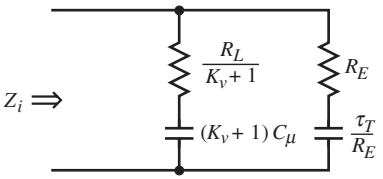
$$Z_{i\pi} = (\beta_o + 1) \cdot (r_e + R_E) \cdot \frac{s\alpha_o \cdot \tau_T \cdot (R_E/(r_e + R_E)) + 1}{s\tau_\beta + 1}$$

For $R_E >> r_e$ and $\alpha_o \cong 1$, $Z_{i\pi}$ is the hf $Z_b(R_E)$. If the series-peaking $\omega_n << \omega_T$, we can ignore the zero in $Z_{i\pi}$. The result is a shunt RC with R_i shunting

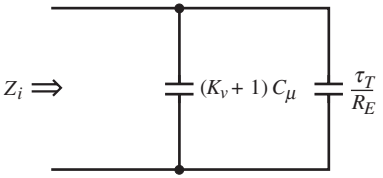
$\tau_\beta / R_i \cong \tau_T / R_E$. R_i is large and usually presents negligible shunting. Consequently, $Z_{i\pi}$ reduces to τ_T / R_E. It shunts $Z_{i\pi}$ so that

$$Z_i = Z_{i\mu} \| Z_{i\pi}$$

When the branches are combined, Z_i as shown below results.



This simplifies, under these assumptions, to the following circuit.



Unless the transistor is very slow (large τ_T) or R_E is small (not much larger than r_e or less), the only significant capacitance is the Miller capacitance. Therefore,

$$Z_i \cong \frac{1}{s(K_v + 1) \cdot C_\mu}, \quad R_E >> r_e, \quad \beta_o >> 1, \quad f << f_T, \quad \frac{R_L}{K_v + 1} \cong 0$$

Another assumption of this equation is that the capacitance loading R_L from the stage following it is negligible. If not, shunt capacitance across R_L further reduces the impedance in series with C_μ, making this a better approximation.

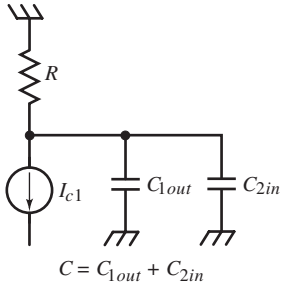
A more exacting estimate is based on Z_i of the previous (more exact) network above. Let the elements be designated more generally as R_1 in series with C_1 shunting R_2 in series with C_2. Then,

$$Z_i = \frac{(sR_1C_1 + 1)\cdot(sR_2C_2 + 1)}{[s(C_1 + C_2)]\cdot[s(R_1 + R_2)\cdot(C_1 \| C_2) + 1]}$$

When element values are substituted, the zeros are at frequencies of $1/\tau_T$ and $1/R_L \cdot C_\mu$; the poles are at the origin, and $1/(R_1 + R_2) \cdot (C_1 \| C_2)$ (where $\|$ is a math operator, not a topological designator). The capacitance $C_1 + C_2$ dominates Z_i until $1/R_L \cdot C_\mu$. The second pole causes Z_i to appear capacitive out to f_T.

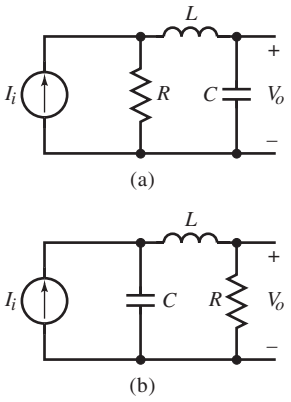
Inductive Peaking

Interstage coupling often degrades bandwidth due to parasitic reactances. For example, collector output capacitance shunts the input base capacitance of the next stage in the circuit shown below.



The load resistor R is shunted by $C = C_{1out} + C_{2in}$. An unwanted pole is created at $1/R \cdot C$.

The addition of an inductor can extend the bandwidth by creating a series or parallel resonant circuit with a peak in the frequency or transient responses – hence the technique name of *inductive peaking*. Below are some *series peaking* circuits.



The transfer function is not changed by exchanging R and C. In both cases, L is in series with C. The transfer function for series peaking has a quadratic pole factor,

$$\frac{V_o}{I_i} = R \cdot \frac{1}{s^2 LC + sRC + 1}$$

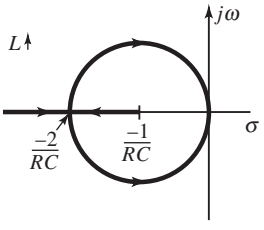
This has a familiar quadratic-pole response. The basic parameters are

$$\omega_n = \frac{1}{\sqrt{a}} = \frac{1}{\sqrt{LC}} = \frac{2\zeta}{RC}; \quad \zeta = \frac{b}{2\sqrt{a}} = \frac{R}{2Z_n}, \quad Z_n = \sqrt{\frac{L}{C}}$$

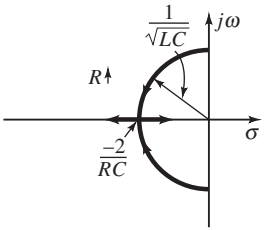
Usually, R is chosen to set the gain, and C is parasitic. This leaves L as the design parameter. For a desired ζ,

$$L = \frac{R^2 \cdot C}{4 \cdot \zeta^2}$$

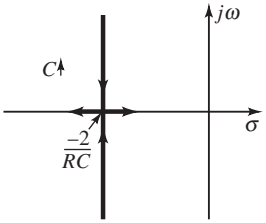
Because L is in a (the s^2 coefficient) only, the poles move with increasing L as shown below.



It is worth noting that if R were varied instead, the pole locus would be as shown below.



And for increasing C, the direction of the root locus is vertical downward, splitting along the real axis.



When the poles become complex as L increases, pole radius ω_n shrinks and ζ decreases. The most desirable pole locations for a wideband amplifier are in a range slightly off the real axis, near the critically damped pole location $-2/RC$. Here, ω_n is maximum and ζ is in a range that gives a desired response.

For variation in R, ω_n remains constant as ζ changes proportionally. Variation in C causes the most trouble. While ζ varies with the square root of C, ω_n varies inversely. Consequently, with C usually parasitic, control over its range of values is least, and though response peaking (in time or frequency) is not so much affected by ΔC, the risetime and bandwidth are. Causes for C, such as transistor process parameters and circuit-board layout, are significant in control of pole radius.

How much improvement in bandwidth can series peaking offer? To determine this, compare bandwidth improvement with the uncompensated RC circuit by expanding the meaning of the bandwidth reduction factor S to include

bandwidth extension. Both definitions of S are useful here and are separately denoted as

$$S_n \equiv \frac{\omega_{bw}}{\omega_n}, \quad S_p \equiv \frac{\omega_{bw}}{p} = \frac{\omega_{bw}}{(1/RC)}$$

S_n compares bandwidth with respect to pole radius; S_p compares it with the uncompensated RC circuit.

In the time domain, $t_r \cdot \omega_n$ has been used to express relative risetime. Comparison against the risetime of the RC circuit also is a measure of improvement. The risetime improvement factor is

$$\frac{t_r}{t_r(RC)} \cong \frac{t_r}{2.2 \cdot RC}$$

All of these performance indicators are combined with those already derived in *Dynamic Response Compensation*, "Optimization . . .", in the following series-peaking summary table.

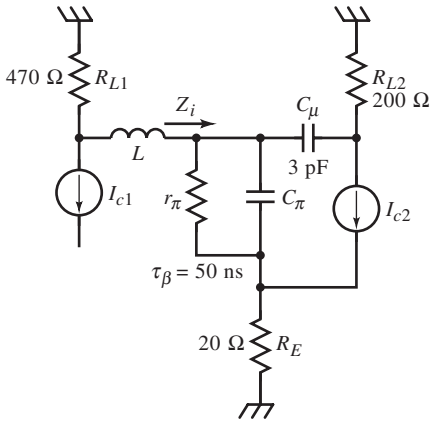
ζ	L	ω_n	$\frac{\omega_{bw}}{\omega_n}$	$\frac{\omega_{bw}}{p}$	M_p,%	$t_r \cdot \omega_n$	$\frac{t_r}{2.2 \cdot RC}$	
1.00	$\frac{R^2C}{4}$	$2/RC$	0.644	1.288	0	3.36	0.765	critical damping
0.866	$\frac{R^2C}{3}$	$1.73/RC$	0.786	1.361	0.433	2.73	0.717	MFED
0.707	$\frac{R^2C}{2}$	$1.41/RC$	1.000	1.414	4.32	2.15	0.692	maximally flat amplitude
0.500	R^2C	$1/RC$	1.272	1.272	16.3	1.64	0.746	$\phi = 60°$

For $\zeta = 0.5$, the poles are at a 60° angle and $\alpha = 1/RC$, the same as the single-pole case. From the root-locus plots above, at critical damping, both poles are at $-2/RC$ and have twice the pole radius of a single-pole RC circuit. As L varies,

$$\left(\frac{\omega_{bw}}{p}\right) = \left(\frac{\omega_{bw}}{\omega_n}\right) \cdot (2 \cdot \zeta) \Rightarrow S_p = 2 \cdot \zeta \cdot S_n$$

By adding L for series peaking, the critically damped ($\zeta = 1$) bandwidth increases by 29% and the risetime by 31%. This is a significant improvement caused by the addition of one component, but greater improvement is possible.

Example: Series Peaking



The amplifier stage output is loaded by the input impedance of a CE stage. The CE transistor has $\tau_T = 500$ ps and $\beta_o \cong 100$. The collector capacitance to ground of the first BJT is negligible. A MFED response is desired.

The first step is to find the input impedance Z_i of the loading stage. The poles involved in inductive peaking will be in the hf region of the CE BJT stage. The emitter branch hf capacitance is 500 ps/20 Ω or 25 pF. Z_π in series with it is negligible. The voltage gain is about

$$\frac{R_{L2}}{R_E} = 10$$

and the Miller collector capacitance is $11 \cdot (3\text{ pF}) = 33$ pF. Because R_E is so small, τ_T / R_E is significant. The collector output resistance, taking into account the Miller effect, is less than 20 Ω. The capacitance in series with L is thus

$$C_i \cong 25\,\text{pF} + 33\,\text{pF} = 58\,\text{pF}$$

The uncompensated bandwidth is

$$f_{bw}(\text{uncomp}) = \frac{1}{2\pi \cdot RC} \cong \frac{1}{2\pi \cdot (470\,\Omega) \cdot (58\,\text{pF})} = 5.8\,\text{MHz}$$

And risetime is

$$t_r \cong 2.2 \cdot RC = 60\,\text{ns} \quad \text{or} \quad t_r \cong \frac{0.35}{f_{bw}} = 60\,\text{ns}$$

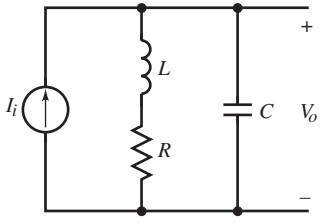
From the inductive peaking table, for MFED response,

$$L = \frac{R^2 \cdot C}{3} \cong \frac{(470\,\Omega)^2 \cdot (58\,\text{pF})}{3} = 4.3\,\mu\text{H}$$

The value of $R = R_L$ assumes negligible series resistance in Z_i. Each path in Z_i has about 20 Ω. Z_i becomes resistive at $1/R_L \cdot C_\mu$ or 265 MHz. This is about 50 times larger than f_{bw}(uncomp), and the assumption that Z_i is purely capacitive over the frequency range of interest is valid.

The compensated bandwidth (from the table) is 1.36 times higher, or 7.9 MHz, and the risetime is 43 ns. The series resonance is at f_n = 10 MHz.

An alternative to series peaking is *shunt peaking*, as shown below. The addition of L in series with R places it in parallel with C and creates a parallel resonance. For a step of input current, most of it charges C at first because current does not change instantaneously in an inductor. Consequently, C charges faster and response speed increases.



The transfer function of the shunt peaking circuit is

$$\frac{V_o}{I_i} = R \cdot \frac{s(L/R)+1}{s^2LC + sRC + 1}$$

The addition of a zero over series peaking improves response speed but also peaks the response more. To compare shunt with series peaking, we need formulas for the performance parameters of a two-pole, one-zero circuit. We now digress to derive them generally and then apply them to shunt peaking.

The two-pole, one-zero transfer function can be generally expressed in terms of $\tau_n = 1/\omega_n$ and Q as

$$\frac{V_o}{I_i} = R \cdot \frac{sQ\tau_n + 1}{s^2\tau_n^2 + (\tau_n/Q)s + 1}$$

In narrow-band amplifier terminology,

$$Q \equiv \frac{1}{2\zeta}$$

This quantity quantifies the amount of peaking and occurs frequently in resonant-circuit equations. The time constant of the zero of V_o/I_i is

$$\tau_z = Q \cdot \tau_n \Rightarrow Q = \left(\frac{\tau_z}{\tau_n}\right)$$

Bandwidth is found in the usual way by setting the magnitude of V_o/I_i to $1/\sqrt{2}$. The general result is

$$\omega_{bw} = \omega_n \sqrt{1 - \frac{1}{2Q^2} + 2Q^2 + \sqrt{\left(1 - \frac{1}{2Q^2} + 2Q^2\right)^2 + 1}}$$

The overshoot, M_p, is expressed in ζ as

$$M_p = \frac{1}{2\zeta} \cdot \exp\left[\frac{-\zeta}{\sqrt{1-\zeta^2}} \cdot \left(\frac{3\pi}{2} - \cos^{-1}\zeta - \tan^{-1}\left\{ \frac{\sqrt{1-\zeta^2}}{\zeta} \right\} + \sin^{-1}(-\zeta) \right) \right]$$

where $3\pi/2 = 270°$.

The unit-step response for $\zeta < 1$ is

$$v_{step}(t) = 1 - \frac{1}{2\zeta\sqrt{1-\zeta^2}} \cdot e^{-\zeta w_n t} \sin\left(\omega_n t\sqrt{1-\zeta^2} + \cos^{-1}\zeta + \tan^{-1}\left\{ \frac{\sqrt{1-\zeta^2}}{\zeta} \right\} \right)$$

The 10% and 90% times are found by numerical computer solution for shunt peaking as $t_r \cdot \omega_n$. This is a convenient representation since $\omega_n t$ is the independent variable.

For $\zeta = 1$, the poles are repeated, and the step response is

$$v_{step} = \left[\left(\frac{p}{z} - 1 \right) \cdot p \cdot t - 1 \right] \cdot e^{-p \cdot t} + 1, \quad \zeta = 1$$

This function also has no closed-form solution and is numerically solved by computer.

We now apply these general results to shunt peaking and V_o/I_i, where the zero is

$$z = \frac{1}{\tau_z} = \frac{1}{L/R} = 2 \cdot \zeta \cdot \omega_n$$

and the repeated poles have a frequency of

$$2p = \frac{2}{RC} = \omega_n$$

It then follows that, when $\zeta = 1$,

$$\frac{p}{z} = \frac{1}{2}$$

and v_{step} reduces to that for shunt peaking:

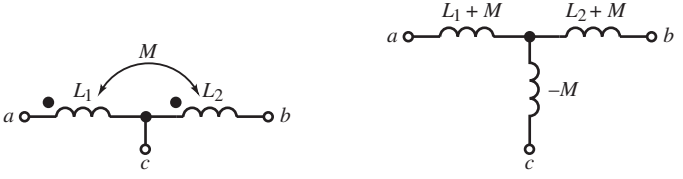
$$v_{step} = \left(-\frac{1}{2} \cdot \omega_n \cdot t - 1 \right) \cdot e^{-\omega_n t} + 1$$

The values of $t_r \cdot \omega_n$ are numerically computed from this equation.

A table similar to that for series peaking can now be constructed for shunt peaking.

ζ	Q	$\frac{\omega_{bw}}{\omega_n}$	$\frac{\omega_{bw}}{p}$	M_p,%	$t_r\omega_n$	$\frac{t_r}{2.2 \cdot RC}$	ϕ
1.00	0.500	0.786	1.572	0	3.071	0.699	0°
0.866	0.577	1.086	1.881	0.620	2.319	0.609	30°
0.707	0.707	1.554	2.198	6.70	1.559	0.502	45°
0.500	1.000	2.279	2.279	29.8	0.940	0.428	60°

Shunt peaking is faster than series peaking for the same pole parameters but is less damped in response. It achieves an 88% increase in bandwidth over the *RC* circuit and 39% decrease in risetime for a MFED pole response. Comparing the shunt and series peaking tables directly can be misleading, however. A pole angle of, say, 30° is *not* an MFED response for shunt peaking because of the zero. The values of ζ for MFED and maximally flat amplitude (MFA) responses must be derived as in *Designing Dynamic Circuit Response*, "Graphical Representation . . .". For an MFED response, $\zeta = 0.881$, and for MFA, $\zeta \cong 0.777$. These values are somewhat higher than those without the zero.



Greater speed improvement can be achieved by using a *T-coil* (a). This is a transformer with controlled coupling and a common connection at *c*. An equivalent circuit is shown in (b) where the polarity of coupling determines the

polarity of the mutual inductance $-M$ to terminal c. With the coupling as shown,

$$L = L_{ab} = (L_1 + M) + (L_2 + M) = L_1 + L_2 + 2M$$

The addition of $-M$ in the equivalent circuit produces the correct self-inductances:

$$L_{ac} = (L_1 + M) - M = L_1,\ L_{bc} = (L_2 + M) - M = L_2$$

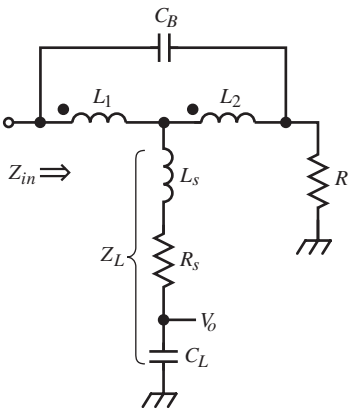
If terminals a and b are shorted, the inductance from a or b to c is

$$L_{ab,c} = (L_1 + M) \| (L_2 + M) - M$$

Let the mutual inductance be signed. Then the coupling coefficient is always positive and is

$$k = \left| \frac{M}{\sqrt{L_1 \cdot L_2}} \right|$$

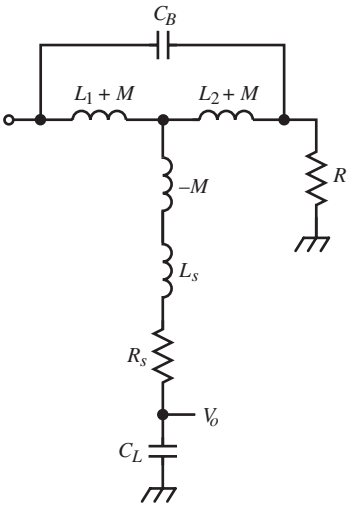
The use of the T-coil for bandwidth extension has resulted in the general form of the *bridged T-coil circuit*, shown below.



The coil is terminated in R, but the load is connected to the centertap. A general load,

$$Z_L = sL_s + R_s + \frac{1}{sC_L}$$

(or series RLC) is similar to the input impedance of BJT stages. The equivalent circuit is shown below.



At 0 Hz, the input impedance of the T-coil is R. Because of the bridging capacitor C_B, at high frequencies it is also R. For a given Z_L and by proper choice of L_1, L_2, M, and C_B, $Z_{in} = R$ and is independent of frequency. For this circuit behavior, the resulting design equations are

$$L_1 = \frac{C_L}{4} \cdot \left(1 + \frac{1}{4\zeta^2}\right) \cdot (R + R_s)^2 - RR_sC_L - L_s$$

$$L_2 = \frac{C_L}{4} \cdot \left(1 + \frac{1}{4\zeta^2}\right) \cdot (R + R_s)^2 - L_s$$

$$M = \frac{C_L}{4} \cdot \left(R^2 - R_s^2 - \frac{1}{4\zeta^2}(R + R_s)^2\right) + L_s$$

$$C_B = \frac{C_L}{16\zeta^2} \cdot \left(1 + \frac{R_s}{R}\right)^2$$

In addition to these design equations, the equivalent inductor element values are

$$L_1 + M = \frac{RC_L}{2} \cdot (R - R_s) - L_s$$

$$L_2 + M = \frac{RC_L}{2} \cdot (R - R_s) - L_s$$

$$L = R^2 C_L - 2L_s$$

The transfer function has two poles at

$$p_{1,2} = -\frac{4\zeta^2}{RC_L} \pm j\frac{4\zeta}{RC_L} \cdot \sqrt{\zeta^2 - 1}$$

The form of transfer function is the same as for series peaking but with twice the speed improvement! For MFED response, $\omega_{bw}/p = 2.72$, nearly three times better than the original *RC* circuit. The greatest improvement is 2.83 for an MFA response.

For $Z_L = 1/sC_L$, the transfer function for the load is

$$\frac{V_o}{I_i} = R \cdot \frac{1}{\frac{1}{4} \cdot \left(\frac{1-k}{1+k}\right) \cdot R^2 C_L^2 s^2 + \frac{1}{2} \cdot RC_L s + 1}$$

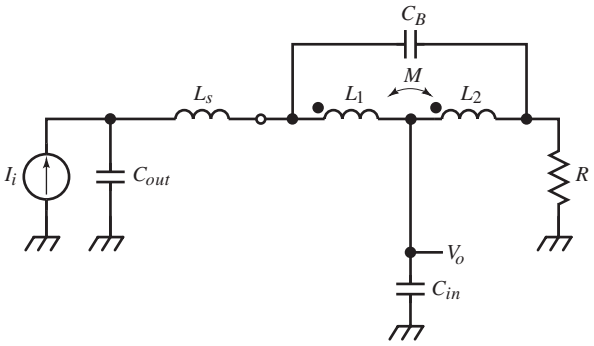
$$C_B = \frac{1}{4} \cdot \left(\frac{1-k}{1+k}\right) \cdot C_L, \quad L = R^2 C_L, \quad L_1 = L_2 = \frac{L}{2 \cdot (k+1)}$$

As k increases, the pole angle decreases for complex poles. With perfect coupling, $k = 1$, and the s^2 term is zero, leaving a single-pole response but with twice the bandwidth of a simple *RC* circuit. This is a simpler, lower-performance T-coil

circuit with no bridging capacitance and with $L_1 = 4 \cdot L_2$. MFED response is achieved when $k = 1/2$, a relatively loose coupling not hard to implement. Then $C_B = C_L/12$, $L_1 = 3 \cdot L_2$, and $L = R^2 \cdot C_L$. A balanced T-coil ($L_1 = L_2$) that meets the above conditions has no coupling between the coils, $C_B = C_L/4$, and the pole angle is 60°. $L_1 > L_2$ is necessary to meet these conditions with a capacitive load.

The basic theory of T-coils, including the full derivation of the above equations (and others), was originally worked out at Tektronix by Robert I. Ross and Carl Battjes (1982), inventor of T-coil compensation. Subsequently, Peter Starič and Erik Margan presented the derivation in *Wideband Amplifiers* (2006).

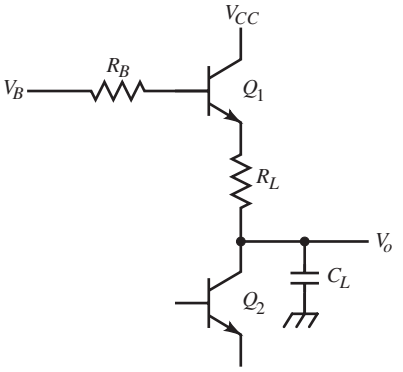
Even greater bandwidth improvement is possible by taking advantage of the constant-resistance input of the T-coil circuit. Series peaking can be cascaded in front of the T-coil, as shown below.



The input and output capacitances of the uncompensated circuit are separated and become part of different peaking circuits. Because the interstage coupling satisfies the requirement for series peaking, the bandwidth improvement of each circuit remains unchanged. Thus the total improvement is the product of the individual improvement factors.

Amplifiers with bandwidths under 100 MHz usually have had T-coils constructed as tapped cylindrical coils wound on a plastic bobbin. The magnetic path is through air and plastic. For higher frequencies, a common T-coil is made of a bifilar-wound loop of magnet wire. The two wires are twisted together and then formed into a loop, or circuit-board traces can be spiraled on opposite

sides of the board and connected at the center via plated-through holes. Two traces can be run next to each other to form coupled inductors. For very-high-speed circuits, integrated circuit (IC) bonding wires have even been used to form T-coils.



An inductive peaking circuit used in ICs, sometimes referred to as "emitter peaking," realizes a shunt inductance by the high-frequency gyration of a BJT base resistor. The emitter appears inductive. (See *Designing Dynamic Circuit Response*, "High-Frequency Impedance Transformations.") The adjustment of R_B adjusts the emitter inductance $R_B \cdot \tau_T$.

Example: T-Coil Compensation

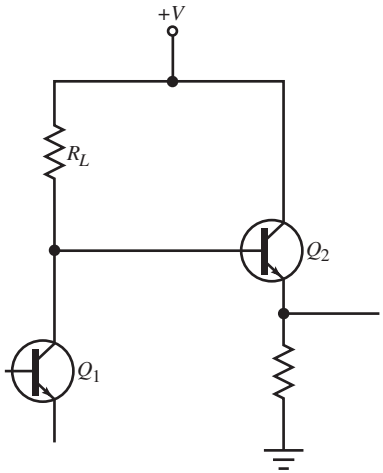
The circuit of the previous example is T-coil compensated. The T-coil formulas are applied directly. The loading is $C_L = 58$ pF, and $R_s = L_s = 0$. Also, $R = R_L = 470\ \Omega$. For MFED response, $\zeta = 0.866$. The results are

$$L_1 = L_2 = 4.3\,\mu\text{H}, \quad M = 2.1\,\mu\text{H}, \quad C_B = 4.8\,\text{pF}$$

From the equation for k, the coupling of the inductors is $k = 0.5$. This is loose coupling and is easily implemented. The bandwidth has improved to 15.8 MHz and risetime to approximately 22 ns.

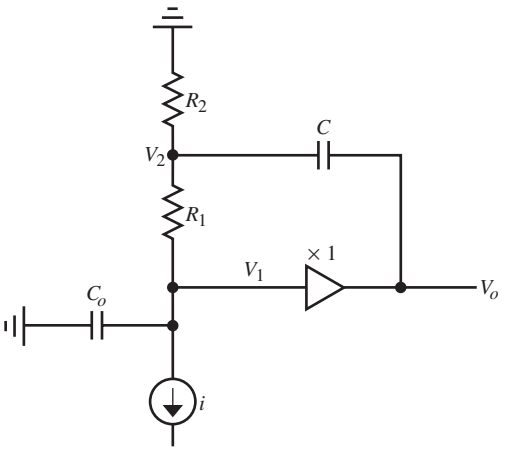
Bootstrap Speed-Up Circuit

How can the speed of an amplifier be increased without increasing its power dissipation? Amplifier bandwidth can be extended by a simple technique that requires no inductors or complicated adjustments. In its simpler form, a collector- or drain-loaded transistor amplifier, as shown below, has a single-load resistor, R_L, followed by a buffer amplifier, usually an emitter- or source-follower.



The buffer stage keeps the output from directly loading R_L by providing current gain and approximately a ×1 voltage gain. This two-stage cascaded amplifier is typical in high-speed circuit design as a place to start before increasing amplifier "speed" (bandwidth).

What mainly can limit speed is the capacitance, C_o, at the collector node of the first stage. It forms a time constant with R_L that slows the dynamic response. To increase speed, it is possible to introduce various inductive peaking techniques. However, for IC and some discrete design, inductors are problematic. A speed-up technique that uses capacitance instead is the *bootstrap speed-up circuit*, shown below, where Q_1 has been replaced by a current source, i, and the Q_2 stage by an ideal ×1 buffer amplifier. R_L has been split into R_1 and R_2, and a bootstrap capacitor, C, added between the output and the split-R_L node.



The idea behind this circuit is that if the voltage at the top of R_1 can track the voltage at the bottom of R_1, the voltage change across R_1 will be zero. Then none of i will be diverted into the load resistance. With all of i flowing into C_o, it charges faster and the circuit response is quicker. In effect, C *bootstraps* the voltage across R_1 to accomplish this, and the buffer amplifier output has the needed current drive to provide R_2 current.

To analyze the dynamic response of this circuit, the *s*-domain (pole-zero) expressions need to be derived. The full analysis takes some algebraic effort. To provide further insight into the circuit and guidance for checking the full result later, a simpler analysis omits C_o. By setting $C_o = 0$, the output-node voltage is

$$v_1 = i \cdot \left(R_1 + R_2 \| \frac{1}{sC} \right) + v_o \cdot \left(\frac{R_2}{R_2 + 1/sC} \right); \quad v_o = v_1$$

After some algebraic simplification, this becomes

$$v_1 = i \cdot \left(R_1 + \frac{R_2}{sR_2C + 1} \right) + v_o \cdot \left(\frac{sR_2C}{sR_2C + 1} \right)$$

Then the first-stage amplifier gain, a transresistance (current in, voltage out), is

$$\frac{v_o}{i} = \frac{R_1 + \dfrac{R_2}{sR_2C + 1}}{1 - \dfrac{sR_2C}{sR_2C + 1}} = (R_1 + R_2) \cdot (s[R_1 \| R_2] \cdot C + 1)$$

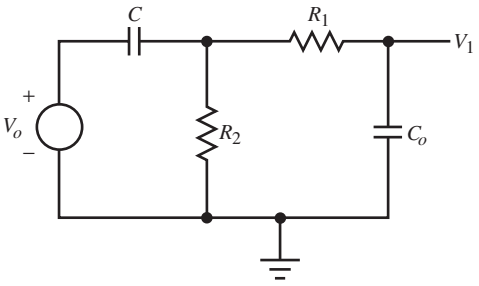
From this result, $R_L = R_1 + R_2$ is the static gain, followed by a zero at $z = -1/(R_1 \| R_2) \cdot C$. The absence of a pole is due to bootstrapping. The pole formed by R_2 and C is cancelled by the ×1 buffer gain. If its gain were a value of K other than one, a finite pole factor would appear at $s \cdot (K-1) \cdot R_2 \cdot C + 1$.

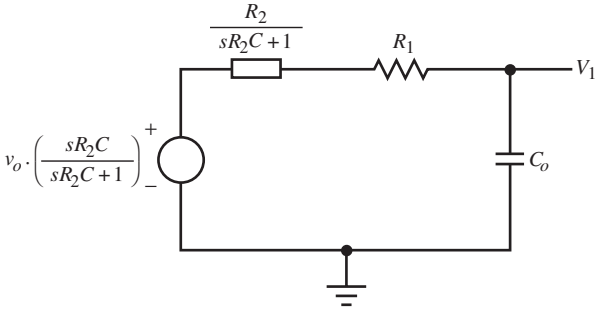
Now, include C_o in the analysis such that $C_o \neq 0$. Then the circuit can be modeled as shown below. The input current source is omitted but attaches to the output node, at v_1. The two-loop circuit reduces to a single loop by Thevenizing v_o, C, and R_2, as shown in the second circuit. Then superposition of the two sources, v_o and i, results in the following two equations:

$$v_1 = v_o \cdot \left(\frac{sR_2C}{sR_2C+1} \right) \cdot \left(\frac{sR_2C+1}{s^2(R_1R_2CC_o) + s[(R_1+R_2)\cdot C_o + R_2C] + 1} \right), \quad i = 0$$

$$v_1 = i \cdot \left(\frac{1}{sC_o} \| (R_1 + R_2 \| 1/sC) \right)$$

$$= i \cdot (R_1 + R_2) \cdot \frac{s(R_1 \| R_2)C + 1}{s^2 R_1 R_2 C C_o + s[(R_1+R_2)C_o + R_2C] + 1}, \quad v_o = 0$$





The first equation simplifies quickly to

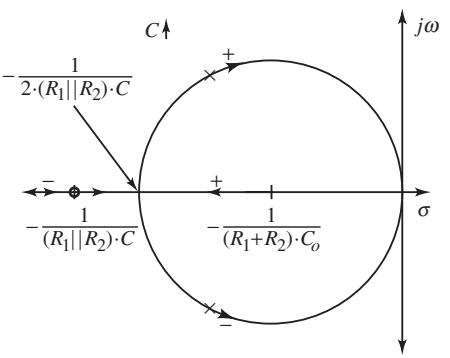
$$v_1 = v_o \cdot \left(\frac{sR_2C}{s^2(R_1R_2CC_o) + s[(R_1 + R_2)\cdot C_o + R_2C] + 1} \right), \quad i = 0$$

Applying the buffer condition, $v_o = v_1$, and by superposition,

$$\frac{v_o}{i} = (R_1 + R_2)\cdot \frac{s(R_1 \| R_2)C + 1}{s^2R_1R_2CC_o + s(R_1 + R_2)C_o + 1}$$

The bootstrap capacitor, C, provides for the additional zero. Without it $(C = 0)$, the uncompensated circuit pole remains at $-1/(R_1 + R_2)\cdot C_o$, as in the uncompensated amplifier. The design question is now one of determining the optimal value of C and the split between R_1 and R_2.

The contour (not root-locus) plot of the poles of this circuit is shown below. C appears only in the quadratic term, leading to a locus that varies with increasing C as shown.



The conjugate poles are marked with their polarities to show that the positive pole originates at the center of the pole-pair circle, $-1/(R_1 + R_2)\cdot C_o$ (the pole location of the uncompensated amplifier), and the negative pole at $-\infty$. The contour plot itself begins at the frequency of the zero, $-1/(R_1 \| R_2)\cdot C$, and at the circle center. As C increases, the poles move toward each other and meet at $-1/2(R_1 \| R_2)\cdot C$ before splitting off the real axis and eventually terminating at

the origin for excessively large (infinite, shorted) C. For equal poles on the real axis,

$$2(R_1 \| R_2) \cdot C = \tfrac{1}{2} \cdot (R_1 + R_2) \cdot C_o$$

from which the design constraints are obtained: $R_1 = R_2$ and $C = C_o$.

Optimal amplifier response usually has the poles off the real axis for a compromise between time- and frequency-domain response performance. For poles alone, the MFED angle is 30°. The relevant formulas are as follows. The damping factor is

$$\zeta = \frac{b}{2 \cdot \sqrt{a}} = \frac{1}{2} \cdot \sqrt{\frac{(R_1 + R_2) \cdot C_o}{(R_1 \| R_2) \cdot C}}$$

where the pole angle is

$$\phi = \cos^{-1} \zeta$$

The pole radius is

$$\omega_n = \frac{1}{\sqrt{a}} = \frac{1}{\sqrt{R_1 R_2 C C_o}}$$

Then for a pole angle of $\phi = 30°$, $\zeta = \cos(30°) = \sqrt{3}/2$ and

$$\frac{(R_1 + R_2) \cdot C_o}{(R_1 \| R_2) \cdot C} = 3$$

Given R_1, R_2, and an estimate for C_o,

$$C = \frac{(R_1 + R_2) \cdot C_o}{3 \cdot (R_1 \| R_2)}$$

For $R_1 = R_2$, then

$$C = \frac{4}{3} \cdot C_o$$

C_o is an undesirable parasitic circuit element, is typically small, and takes on a small range of possible values. Because of its only approximate value, C might need to be adjustable to tune the pole angle for optimal response.

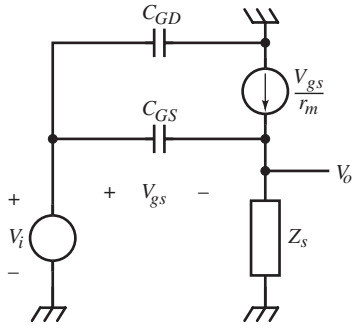
Because of its small size, C is also small, approximately the value of C_o. Consequently, small-variable-C methods might need to be applied. One simple discrete-circuit approach is to use what used to be called a "gimmick." Twist a length of insulated wire, strip and solder-coat one pair of ends as C terminals, then snip the length of the pair for optimal C. Varnish or glue C to retain its geometry and its C value. This works for relatively slow high-speed circuits. Better methods can also be applied. If approximate dynamic behavior is adequate, or by using a response trim elsewhere, no adjustment is needed. In this case, a circuit-board or IC capacitance, though approximate, can be sufficient.

SOURCE-FOLLOWER COMPENSATION

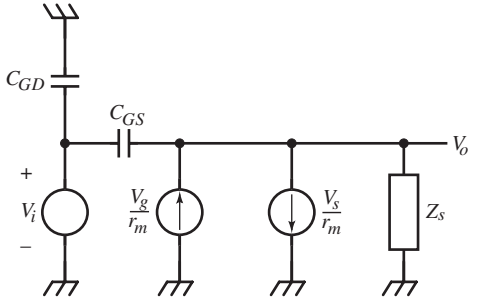
Common-source (CS) stages are often used at the input of instrument amplifiers to minimize resistive loading. Unlike their BJT counterpart (the CC), FET C_{GS} is typically much smaller than C_π of BJTs. Consequently, the Z_π term of Z_b (described in *Designing Dynamic Circuit Response*, "High-Frequency Impedance Transformations") cannot be ignored as it usually can for the BJT. Applying the transformation for FETs and including the effect of C_{GD},

$$Z_g = \left[\frac{1}{sC_{GS}} + Z_S \cdot \left(\frac{s\tau_T + 1}{s\tau_T}\right)\right] \Big\| \frac{1}{sC_{GD}}$$

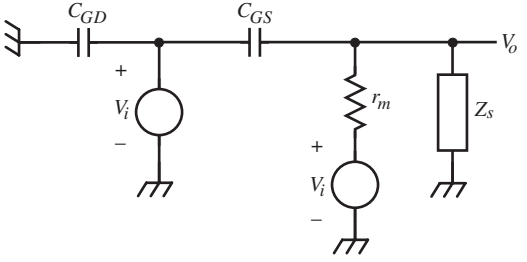
where $\tau_T = r_m \cdot C_{GS}$. The equivalent circuit is shown below.



It is transformed by applying the substitution theorem to the circuit shown below.



Then Norton-to-Thevenin conversion of the current source results in the following circuit.



For a capacitive load of C_S, V_i drives a capacitive divider that causes an input voltage step to immediately rise to a fraction of the step amplitude,

$$\frac{C_{GS}}{C_{GS}+C_S}$$

and then continues to rise exponentially, due to r_m, to the input step value. The transfer function of the above circuit is

$$\frac{V_o}{V_i}=\frac{sr_mC_{GS}+1}{sr_m\left(C_{GS}+C_S\right)+1}=\frac{s\tau_T+1}{s\left(\tau_T+r_mC_S\right)+1}$$

The pole is less than the zero, resulting in the response of a phase-lag circuit. The initial response step is p/z with a time constant of $r_m \cdot (C_{GS} + C_S)$. The

response can be compensated by adding a phase-lead circuit at another stage in the amplifier.

A second anomaly of the CS is its input impedance. From Z_g, the gyrating factor can be expressed in topological form as

$$Z_S \cdot \left(\frac{s\tau_T + 1}{s\tau_T} \right) = Z_S \cdot \frac{1}{1 - 1/(s\tau_T + 1)}$$

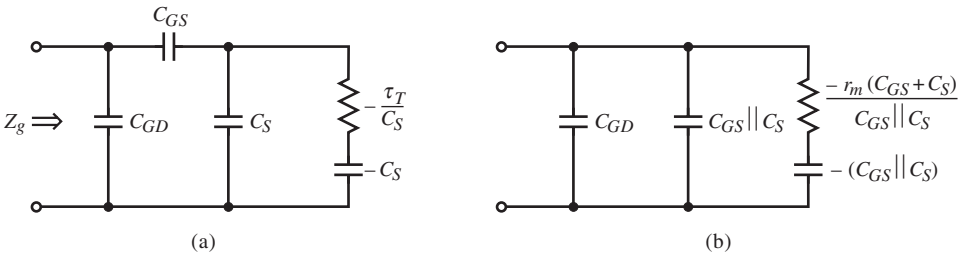
Substituting $Z_S = 1/sC_S$, Z_g reduces to

$$Z_g = \frac{1}{sC_{GD}} \left\| \left(\frac{1}{sC_{GS}} + \frac{1}{sC_S} \left\| \left(-\frac{\tau_T}{C_S} - \frac{1}{sC_S} \right) \right. \right. \right)$$

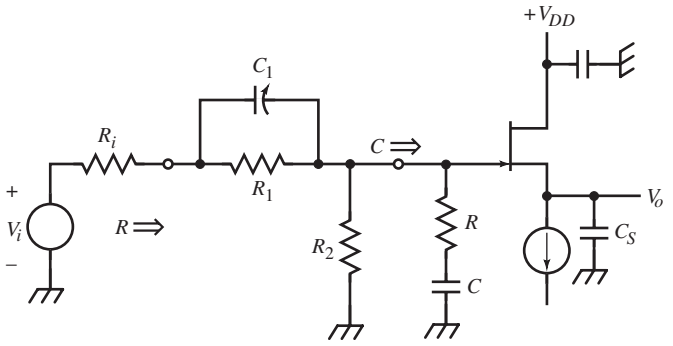
A more useful form is

$$\begin{aligned} Z_g &= \frac{1}{sC_{GD}} \left\| \left(\frac{s(C_{GS} + C_S)\tau_T + C_{GS}}{s^2 C_{GS} C_S \tau_T} \right) \right. \\ &= \frac{1}{sC_{GD}} \left\| \left(\frac{1}{s(C_{GS} \| C_S)} \right) \right\| \left(-r_m \cdot \frac{C_{GS} + C_S}{C_{GS} \| C_S} - \frac{1}{s(C_{GS} \| C_S)} \right) \end{aligned}$$

This equivalent hf gate impedance is shown below and is compensated by the method used to compensate the CC, that of shunting the gate with a series *RC*, which produces a purely capacitive input. (Keep in mind that ‖ is a mathematical operator, not a topological descriptor.) The values of the compensating elements are derived from the expression for Z_g.



A source follower is usually used to prevent loading of a high-impedance source. If the source impedance is resistive, then it forms an uncompensated voltage divider with Z_g. This can be compensated by introducing a shunt RC in series with the input.



It forms a compensated divider with Z_g, resulting in a resistive input.

The last CS problem to be considered is distortion due to large-signal effects. When a large-amplitude square-wave is applied to the compensated CS, C_{GS} and r_m both change significantly between levels. If C_1 is adjusted for compensation of the positive transition, then for the negative transition, r_m and C_{GS} increase causing τ_T to increase. An increase of C_{GS} increases the step fraction, causing negative overshoot or undershoot. The transfer function is also affected. Both pole and zero decrease, but with significant C_S, the pole decreases less. Consequently, p/z decreases, causing the compensator to overcorrect and produce undershoot. FETs with large C_{GS} have reduced undershoot, and if they also have a large τ_T, undershoot error diminishes more quickly. Similarly, FETs with large pinch-off voltages have less r_m variation with V_{GS}.

Example: CD Input Buffer Compensation

A source-follower FET has a transconductance of 50 mS ($r_m = 200\ \Omega$), $C_{GS} = 6$ pF and $C_{GD} = 2$ pF. Maufacturer data sheets give FET capacitances as

$$C_{iss} = C_{GS} + C_{GD},\ C_{rss} = C_{GD}$$

With a load capacitance of $C_S = 10$ pF, a series RC shunting the gate compensates the FET input. To calculate their values, τ_T is needed:

$$\tau_T = r_m \cdot C_{GS} = 1.25\,\text{ns}$$

Then, from the above circuit diagram,

$$R = r_m \cdot \left(\frac{C_{GS} + C_S}{C_{GS} \| C_S} \right) = 853\,\Omega \Rightarrow 820\,\Omega,$$

$$C = C_{GS} \| C_S = 3.75\,\text{pF} \Rightarrow 3.9\,\text{pF}$$

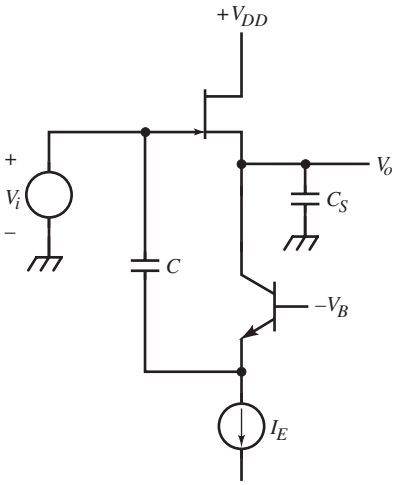
The input now is a shunt RC, where $C_g = C + C_{GD} = 5.9$ pF and $R_2 = 1$ MΩ. To compensate this pole, a shunt RC is placed in series with the input (as in the figure above). In applications in which the input comes from a probe or passive attenuator, the compensating RC is in the probe body so that the probe itself contains the top part of the voltage divider. For a 10 MΩ input,

$$R_1 = 9\,\text{M}\Omega, \quad C_1 = \frac{R_2 C_g}{R_1} = \frac{C_g}{9} = 0.66\,\text{pF}$$

For such large R_1, C_1 is an extremely small capacitance. The R_1 resistor probably has more parasitic shunt capacitance than the value C_1. As a consequence, practical values of capacitors make it infeasible to try to compensate the input divider. That is why, for example, oscilloscope vertical inputs are marked with labelings such as 1 MΩ, 22 pF.

The following circuit takes an alternative approach to common-drain (CD) compensation. It has two paths: the main path through the FET and a compensation path through the CB BJT. The FET path transfer function is given by V_o / V_i, whereas for the compensation path, C forms a divider with $Z_\pi/(\beta + 1)$. The BJT-path voltage gain is

$$\text{BJT } \frac{V_o}{V_i} = \alpha_0 \cdot \frac{sr_{m\text{FET}} \cdot C}{\left[sr_{m\text{FET}} \cdot (C_{GS} + C_S) + 1\right]\left[sr_e \cdot (C_\pi + C) + 1\right]}$$



The paths add to produce the total transfer function. Both paths share the pole with time constant

$$r_{m\text{FET}} \cdot (C_{GS} + C_S) + \tau_{T\text{FET}} + r_{m\text{FET}} C_S$$

Then if the BJT time constant

$$r_e \cdot (C_\pi + C) = \tau_{T\text{BJT}} + r_e \cdot C$$

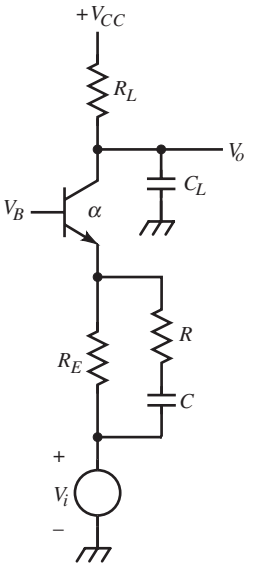
is much smaller than that for the FET path, its pole can be ignored and the transfer functions of the paths added:

$$\frac{V_o}{V_i} \cong \frac{s(\tau_{T\text{FET}} + \alpha_o r_{m\text{FET}} C) + 1}{s(\tau_{T\text{FET}} + r_{m\text{FET}} C_S) + 1}, \quad \tau_{T\text{BJT}} + r_e C \to 0$$

For flat response, the time constants are equated and

$$\alpha_o \cdot C = C_S \Rightarrow C \cong C_S, \quad \alpha_o \cong 1$$

For step inputs with fast edges, the voltage differentiation of C can cause currents that exceed I_E and drive the BJT into cutoff.



EMITTER COMPENSATION

An impedance in series with the emitter (or source) of CB and CE (or CG and CS) amplifiers creates series feedback (see *Designing Amplifier Circuits*, "Noninverting Feedback Amplifier Examples") and can improve speed. Compensation networks can be connected to the emitter node that correct for speed limitations at the collector. The figure shows a CB stage with capacitive output loading.

A series RC is placed in parallel with R_E to provide correction. In the lf region (or for $\tau_T \to \infty$), the transfer function is

$$\frac{V_o}{V_i} = \alpha \cdot \frac{R_L}{r_e + R_E} \cdot \frac{s(R_E + R)\cdot C + 1}{(sR_L C_L + 1)\cdot[s(R + r_e \| R_E)C + 1}$$

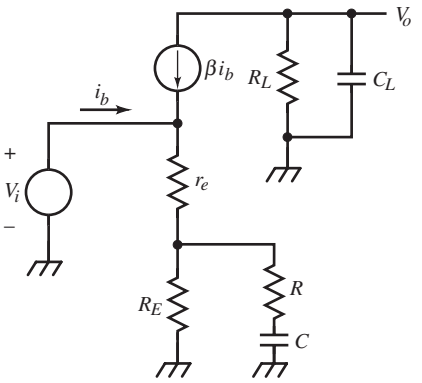
For compensated response, the zero cancels the collector pole at frequency $1/R_L \cdot C_L$, leaving a much higher-frequency pole. For a flat frequency response, the compensating elements must have the values

$$R = R_L - R_E, \quad C = C_L$$

and for bandwidth extension,

$$(R + r_e \| R_E) \cdot C << R_L \cdot C_L$$

For $r_e << R$ and R_E, then $R << R_L$.



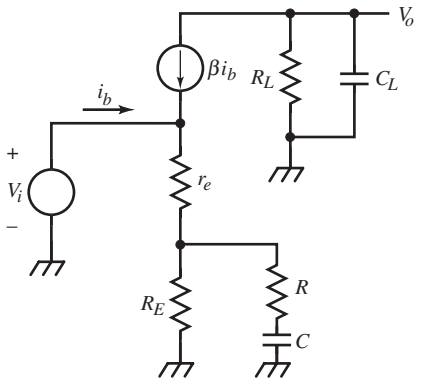
Compensation of CE (or CS) amplifiers is similar. In the circuit shown above, a similar network is connected to the emitter, resulting in a negative V_o/V_i. In both cases, transistor reactances have been ignored, and the resulting equations are useful for amplifiers for which the output pole is below f_β.

Analysis in the hf region uses the hybrid-π BJT model (and its extension to FETs). The CB stage has the advantage over the CE of no Miller effect. However, Z_π forms an uncompensated voltage divider with R_E, requiring an additional shunt C_E around R_E for compensation. At the output, C_μ contributes to C_L.

The CE suffers from the Miller effect, the cause of its dominant pole. Also, the input impedance includes a hf-gyrated emitter impedance. Networks in the

emitter circuit that compensate for output poles have the side effect of creating input anomalies. To design a fast amplifier with CE input, the Miller effect must be minimized and input impedance controlled. The Miller effect is essentially eliminated by following the CE with a CB (thus creating a cascode) or operating the CE as a shunt-feedback amplifier.

First, consider the input side of the CE by assuming a cascode configuration. This simplifies analysis and allows us to ignore C_μ for a while and regard the CE as a transadmittance amplifier.



The transresistance method (now generalized to the transimpedance method) applies in the hf region, using the hf BJT model. CE output current is $\beta(s)\cdot I_b(s)$. The transadmittance is derived from the equivalent circuit, shown above, and is

$$\frac{I_o}{V_i}=\frac{\beta(s)\cdot I_b}{V_i}=\frac{\beta(s)}{Z_b}=\frac{\beta(s)}{[\beta(s)+1]\cdot Z_E}=\alpha(s)\cdot Y_E=\frac{1}{s\tau_T+1}\cdot Y_E$$

Because of $\alpha(s)$, the emitter network admittance Y_E must have a $(s\tau_T+1)$ factor to cancel hf effects. A simple compensation is to let Z_E be a shunt RC:

$$Z_E=\frac{R_E}{sR_EC_E+1}$$

Then,

$$\frac{I_o}{V_i} = \left(\frac{1}{R_E}\right) \cdot \frac{sR_E C_E + 1}{s\tau_T + 1}$$

The response is flat when

$$R_E \cdot C_E = \tau_T$$

In the hf region, the input impedance of the CE is

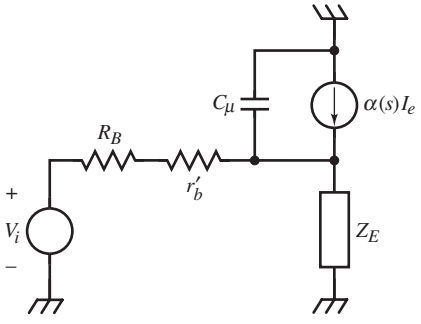
$$Z_b = [\beta(s) + 1] \cdot Z_E = \left(\frac{s\tau_T + 1}{s\tau_T}\right) \cdot \left(\frac{R_E}{sR_E C_E + 1}\right)$$

When the compensated τ_T condition is applied,

$$Z_b = \frac{1}{s(\tau_T / R_E)} = \frac{1}{sC_b}$$

The input is a capacitance of value $C_b = \tau_T / R_E = C_E$.

In the base circuit, C_b forms a pole with the base node resistance. A more complete hf model includes r_b' and C_μ at the internal (b') base node, shown below.



A general expression for the transadmittance is

$$\frac{I_o}{V_i} = \frac{I_o}{I_e} \cdot \frac{I_e}{V_e} \cdot \frac{V_e}{V_i} \cong \left(\frac{\beta}{\beta + 1} \right) \cdot \frac{1}{Z_E} \cdot \frac{\dfrac{1}{sC_\mu} \| (\beta + 1) \cdot Z_E}{\dfrac{1}{sC_\mu} \| (\beta + 1) \cdot Z_E + R_S}$$

$$= \frac{\beta}{(\beta + 1) Z_E + R_S} \cdot \frac{1}{s\left[R_S \| (\beta + 1) \cdot Z_E\right] \cdot C_\mu + 1}$$

where β is $\beta_{hf} = 1/s\tau_T$ and

$$R_S = R_B + r_b'$$

When emitter compensation is added – that is, when Z_E and compensated τ_T are applied – then

$$Z_E = \frac{R_E}{s\tau_T + 1}$$

and the transadmittance reduces to

$$\frac{I_o}{V_i} = \frac{1}{R_E} \cdot \frac{1}{sR_S\left[(\tau_T / R_E) + C_\mu\right] + 1}$$

In the low-frequency (lf) region, for $I_i = V_i / R_B$,

$$K_i = \frac{I_o}{I_i} = \alpha_o \cdot \frac{R_B}{R_S / (\beta_o + 1) + r_e + R_E}$$

The hf model has $\alpha_o = 1$ and $r_e = 0\ \Omega$. For dominant R_E the lf current gain is then

$$K_i \cong \frac{R_B}{R_E}$$

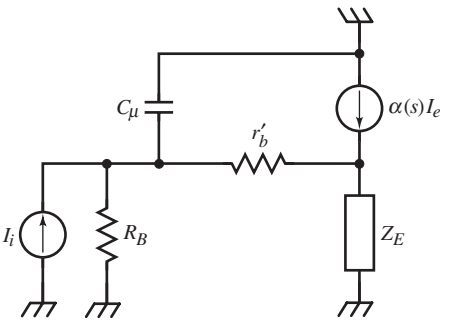
High-speed amplifiers are usually analyzed in terms of current gain because the input variable to a stage is usually a current. The low input-resistance cascode

CB effectively has a current input. The input to a CE is usually the collector of another transistor, modeled as a current source. The Thevenin voltage source input of the above circuit is changed to a Norton equivalent input by setting

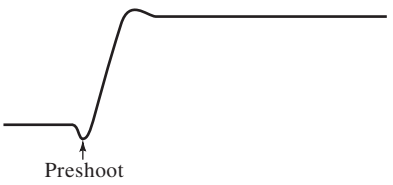
$$V_i = I_i \cdot R_B$$

Then V_o/I_i is used to express current gain as

$$\frac{I_o}{I_i} = \frac{I_o}{V_i/R_B} \cong \frac{K_i}{sR_S\left[\left(\tau_T/R_E\right) + C_\mu\right] + 1}$$



Both V_o/I_i and I_o/I_i are approximate because the path to the output through C_μ is ignored. It introduces a right half-plane (RHP) zero at $1/R_E \cdot C_\mu$. This frequency is usually much higher than the others and can be ignored. The passive path through C_μ causes an output response to occur sooner than the inverted response of the active path. This passive path current is the cause of *preshoot* in the output step response. Instead of rising, the step first dips negative.



The effect of r_b' and C_μ is to degrade speed. Without them, the time constant of the pole in I_o/I_i is at

$$R_B \cdot \frac{\tau_T}{R_E} = K_i \cdot \tau_T$$

Given K_i by design choice, there is an optimum value of R_B that minimizes the pole time constant, which can be written in terms of K_i. By multiplying out the pole time constant in I_o/I_i and setting its derivative to zero,

$$\frac{d}{dR_B}\left(K_i \cdot \tau_T + R_B C_\mu + \frac{r_b' K_i \tau_T}{R_B} + r_b' C_\mu \right) = 0$$

The optimum R_B is

$$\text{optimum } R_B = \sqrt{\frac{K_i \cdot \tau_T \cdot r_b'}{C_\mu}}$$

For this value of R_B, the current-gain pole has a time constant of

$$K_i \cdot \tau_T \cdot \left(1 + 2 \cdot \sqrt{\frac{r_b' \cdot C_\mu}{K_i \cdot \tau_T}} \right)$$

A CE with a lf voltage gain of K_v has an effective C_μ of $K_v + 1$ times, due to the Miller effect. The optimum R_B and pole time constant are modified by multiplying C_μ by $K_v + 1$.

If C_μ is located at the external base node (on the outside of r_b'), as in the previous circuit diagram, the current gain is

$$\begin{aligned} \frac{I_o}{I_i} &\cong \frac{K_i}{s^2\left(K_i \tau_T R_B C_\mu\right) + s\left[R_B\left(\tau_T/R_E + C_\mu\right) + r_b' \tau_T / R_E\right] + 1} \\ &= \frac{K_i}{s^2\left(K_i \tau_T R_B C_\mu\right) + s\left[R_B C_\mu + K_i \tau_T \left(1 + r_b'/R_B\right)\right] + 1} \end{aligned}$$

Again, the RHP zero has been ignored. The CE input now consists of two cascaded *RC* integrators. The minimum ζ is

$$\min \zeta = \sqrt{1 + \frac{r_b'}{R_B}} = \frac{\tau_n}{K_i \cdot \tau_T}$$

Pole separation is typically not significant enough to approximate the response by a dominant single pole since fast amplifiers have values of R_B not very different from r_b'. Consequently, the r_b' term in I_o/I_i cannot be ignored. The value of R_B at maximum pole radius occurs when the poles are repeated and $\zeta = 1$. This results in a fourth-degree equation in R_B. The optimum R_B can be approximated by assuming independence of the time constants of the *RC* integrators. Then fastest response occurs when their time constants are equal, or

$$R_B \cdot C_\mu = r_b' \cdot \left(\frac{K_i \cdot \tau_T}{R_B} \right)$$

Solving for R_B gives

$$\text{optimum } R_B \cong \sqrt{\frac{K_i \cdot \tau_T \cdot r_b'}{C_\mu}}$$

Interestingly, this result is the same as the previous optimum R_B. Whether C_μ is largely internal or external to the base does not strongly affect the optimum R_B value.

Peter Starič and Erik Margan have systematically developed wideband amplifier emitter compensation in *Wideband Amplifiers* (2006).

Cascode Compensation of the Common-Base Stage

Consider the cascode CB stage. Above f_β of the CB transistor, base resistance is gyrated at the emitter to an inductance that resonates with the output capacitance, C_o, of the CE shunting the input to the CB. This shunt *RLC* forms a parallel resonance with

$$\zeta = \frac{1}{2 \cdot R} \cdot \sqrt{\frac{L}{C}} = \frac{1}{2} \cdot \sqrt{\frac{\tau_T}{R_B \cdot C_o}}$$

where R_B is the CB base resistance. For a BJT with $f_T = 300$ MHz, $C_o = 3$ pF and $R_B = 100\ \Omega$, $\zeta = 0.665$ for a pole angle of 48°. With a transistor twice as fast, the pole angle is about 60°. This resonance can cause oscillation because C_o is due largely to C_μ of the CE and is connected to the base. The CE provides the gain that causes oscillation.

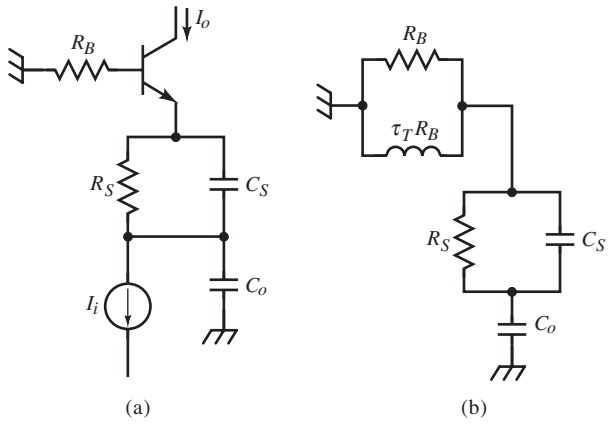
This resonance can be damped by adding resistor R_S in series with the emitter of the CB, isolating it from C_o. The series damping required for MFED response is 77 Ω. Typically, r_e' is 1 Ω, far less than the resistance required. A series resistance damps the resonance but also creates an uncompensated voltage divider with the CB-gyrated base impedance. In addition, it causes voltage gain at the collector of the CE and the Miller effect. The CB transfer function is

$$\frac{I_o}{I_i} = \frac{1}{s^2\{\tau_T R_B C_o[1 + (R_S/R_B)]\} + s(\tau_T + R_S C_o) + 1}$$

The pole radius ω_n, is reduced by

$$\frac{R_B}{R_S + R_B}$$

Compensation in the CB emitter is shown below in (a) as a shunt *RC*.



The hf equivalent emitter circuit (b) has a shunt *RL* due to the gyrated R_B. When

$$R_S \cdot C_S = \tau_T$$

the hf model yields

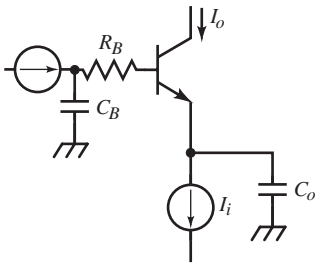
$$\frac{I_o}{I_i} = \alpha(s) I_e = \frac{1}{s^2(\tau_T R_B C_o) + s(\tau_T + R_S C_o) + 1}$$

C_o forms a current divider with the emitter branch; I_e is calculated from the current-divider formula. This compensation maintains the pole radius the same as the uncompensated CB. For design, the value of C_S is determined by the τ_T compensation formula above. R_S is expressed in ζ from I_o/I_i as

$$R_S = -\frac{\tau_T}{C_o} + 2\zeta \cdot \sqrt{\frac{\tau_T \cdot R_B}{C_o}}$$

For $R_S = R_B$, the network of (b) forms an all-pass constant resistance of R_B. The poles are located at $-1/R_B \cdot C_o$ and $-1/\tau_T$.

An estimation of C_o is required to use the expression for R_S. Current in C_μ of the CE is input current to the base. The resulting collector current is larger by the lf current gain. In effect, current in C_μ results in a total current of $K_i + 1$. Thus, the effective capacitance of C_μ is $(K_i + 1) \cdot C_\mu$. (This result suggests a form of Miller's theorem for current amplifiers.)



A CB compensation scheme introduced by John Addis is shown above. C_B shunts R_B, and gyrates to hf damping resistance at the emitter. (See "Emitter-Follower Reactance Plot . . ." in *Designing Dynamic Circuit Response.*)

If an inaccessible R_B is bypassed by C_B on its current-supply side, then a series base RC results. Using the hf BJT model, the emitter impedance is

$$Z_E = \frac{R_B + 1/sC_B}{\beta(s)+1} = \frac{\tau_T \cdot (sR_BC_B + 1)}{(s\tau_T + 1)\cdot(C_B)}$$

Z_E shunts C_o. The current gain of the CB is

$$\frac{I_o}{I_i} = \frac{1}{s^2(\tau_T R_B C_o) + s\left[\tau_T(1 + C_o/C_B)\right] + 1}$$

The pole radius is not reduced by this technique, but control of ζ is more limited:

$$\zeta = \frac{1}{2}\cdot\sqrt{\frac{\tau_T}{R_B\cdot C_o}} + \frac{1}{2}\cdot\sqrt{\frac{\tau_T\cdot C_o}{R_B\cdot C_B^2}} = \zeta_{\text{uncomp}} + \frac{1}{2}\cdot\sqrt{\frac{\tau_T\cdot C_o}{R_B\cdot C_B^2}}$$

The appearance of C_o in both terms indicates a minimum ζ dependent on C_o. Because of the unavoidable base spreading resistance, r_b', R_B is partly constrained in value by the BJT. The design value of C_B is found by solving the above equation:

$$C_B = \frac{C_o}{2\zeta/\sqrt{\tau_T/R_B\cdot C_o} - 1} = \frac{C_o}{\zeta/\zeta_{\text{uncomp}} - 1}$$

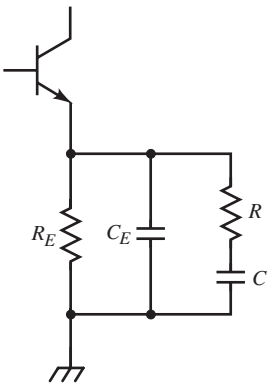
The lower bound on ζ is that

$$\zeta > \zeta_{\text{uncomp}}$$

One remaining pole in the cascode requires compensation. At the output, the CB transistor output capacitance forms a pole with the load resistance. This pole can be compensated by peaking the CE. The CE is already compensated by

$$R_E \cdot C_E = \tau_T$$

Output compensation thus requires a more complicated emitter network. The strategy is to cancel the output pole with a zero. Poles of the emitter network impedance are zeros of the cascode transfer function.



This network meets the requirements. Its impedance is

$$Z_E = R_E \cdot \frac{sRC + 1}{s^2(R_E C_E RC) + s(R_E C_E + R_E C + RC) + 1}$$

Input impedance compensation requires one of the poles of Z_E to be at τ_T. For pole-zero cancellation of the output pole, with time constant τ_L, the other pole of Z_E must be at $-1/\tau_L$. Therefore, the denominator is constrained to be of the form

$$(s\tau_T + 1)\cdot(s\tau_L + 1) = s^2\tau_T\tau_L + s(\tau_T + \tau_L) + 1$$

The zero must lie between τ_T and τ_L. The network behaves as a phase-lead compensator, shifting the load pole to the higher frequency at $1/RC$. The cascode transimpedance with Z_E compensation is

$$\frac{V_o}{I_i} = -R_L \cdot K_i \cdot \frac{1}{s^2(K_i\tau_T\tau_L) + s(K_i\tau_T + RC) + 1}$$

For design, R_L, K_i, τ_T, and τ_L are given. We can choose RC based on the desired damping ratio without affecting the pole radius:

$$RC = 2\zeta \cdot \sqrt{K_i \cdot \tau_T \cdot \tau_L} - K_i \cdot \tau_T$$

With RC determined and R_E constrained by K_i, the coefficients of the Z_E denominator and its constrained form are equated to yield

$$C_E = \frac{\tau_T \cdot \tau_L}{R_E \cdot R \cdot C}$$

$$C = \frac{\tau_T + \tau_L - R \cdot C - \tau_T \cdot \tau_L / R \cdot C}{R_E}$$

$R \cdot C$ is known from the above formula, and R is easily found once C is known. Note that with R and C added, the value of C_E is different from that given by $R_E \cdot C_E = \tau_T$.

Example: Cascode Dynamic Response Compensation

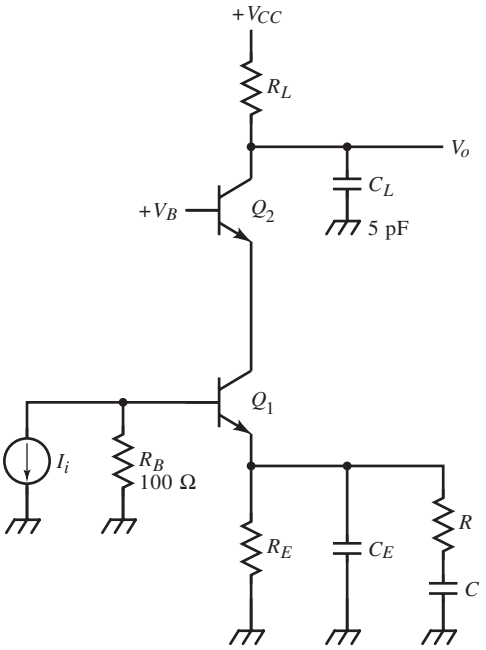
The cascode amplifier is to have approximately MFA response and a transresistance of 1 kΩ with maximum bandwidth. The transistors have f_T = 600 MHz, C_μ = 2 pF, and r_b' = 50 Ω.

The input current source terminates in a 100 Ω base resistor R_B. The output has 5 pF of load capacitance.

To analyze and compensate this amplifier, begin with some transistor calculations:

$$\tau_T = \frac{1}{2\pi \cdot f_T} = 265\,\text{ps}$$

$$C_{o2} = C_{\mu 2} + C_L$$
$$= 2\,\text{pF} + 5\,\text{pF} = 7\,\text{pF}$$



With R_B given, the optimum R_E or K_i can be determined from the optimum-R_B equation by solving for R_E (in K_i):

$$\text{optimum } R_E = \left(\frac{\tau_T}{R_B \cdot C_\mu}\right) \cdot r_b' = 66.3\,\Omega \Rightarrow 68\,\Omega$$

Then $K_i = R_B/R_E = 1.47$. With R_E calculated, R_L can be found:

$$R_m = \frac{V_o}{I_i} = \frac{V_o}{I_o} \cdot \frac{I_o}{I_i} = R_L \cdot K_i \Rightarrow R_L = \frac{R_m}{K_i} = 680\,\Omega \Rightarrow 680\,\Omega$$

The output pole is at

$$f_L = \frac{1}{2\pi \cdot \tau_L} = \frac{1}{2\pi \cdot R_L \cdot C_{o2}} = \frac{1}{2\pi \cdot (4.76\,\text{nsec})} = 33.4\,\text{MHz}$$

To maximize bandwidth, compensation of τ_L is required. Proceed to compensate for both τ_L and the input impedance of the CE. For MFA response (and assuming a single pole-pair),

$$\zeta = \sqrt{2}/2 \cong 0.707$$

Also,

$$K_i \cdot \tau_T = 390\,\text{ps}$$

Applying the formula for *RC*, the series *RC* compensator in the emitter circuit has a time constant of

$$RC = 2\zeta \cdot \sqrt{K_i \cdot \tau_T \cdot \tau_L} - K_i \cdot \tau_T = (1.414) \cdot \sqrt{(390\,\text{ps}) \cdot (4.76\,\text{ns})} - 390\,\text{ps} = 1.54\,\text{ns}$$

Continuing with C_E,

$$C_E = \frac{\tau_T \cdot \tau_L}{R_E \cdot R \cdot C} = 12.0\,\text{pF} \Rightarrow 12\,\text{pF}$$

and from *C*,

$$C = \frac{\tau_T + \tau_L - R \cdot C - \tau_T \cdot \tau_L / R \cdot C}{R_E} = 39.2\,\text{pF} \Rightarrow 39\,\text{pF}$$

With *C* known, *R* can readily be found from the previous calculation:

$$R = \frac{1.54\,\text{ns}}{39\,\text{pF}} = 39.3\,\Omega \Rightarrow 39\,\Omega$$

From series compensation, the new pole is at $1/2\pi \cdot R \cdot C = 104$ MHz.

The CB base spreading resistance of 50 Ω may require compensation. At the emitter it forms a shunt *RLC* circuit with $C_{\mu 1}$ with

$$\tau_n = \sqrt{\tau_T \cdot r_b' \cdot C_{\mu 1}} = 163\,\text{ps} \Rightarrow 978\,\text{MHz}$$

and

$$\zeta = \frac{1}{2} \cdot \sqrt{\frac{\tau_T}{r_b' \cdot C_{\mu 1}}} = 0.814 \Rightarrow 36°$$

Because this resonance has a high value of ζ relative to the MFA value and has f_n at nearly 1 GHz, this pole is not likely to affect the response much and is not compensated.

With R_E now known, we can calculate the CE input time constant. The uncompensated time constant, with only R_E in the emitter circuit, is approximated as

$$\text{uncomp } \tau_i \cong (R_B + r_b' + R_E) \cdot \left(C_\mu + \frac{\tau_T}{R_E} \right) = (218\,\Omega) \cdot (5.9\,\text{pF}) = 1.29\,\text{ns}$$

This corresponds to a frequency of 124 MHz. The compensated τ_i is

$$\text{comp } \tau_i \cong (R_B + r_b') \cdot \left(C_\mu + \frac{\tau_T}{R_E} \right) = 885\,\text{ps} \Rightarrow 180\,\text{MHz}$$

The series R_E and (τ_T / R_E) is replaced by τ_T / R_E alone, and the speed increases. The uncompensated bandwidth is calculated by single-pole approximation of the time constant, from the risetime formula,

$$\begin{aligned} \text{uncomp } \tau &= \sqrt{\tau_i^2 + \tau_{CB}^2 + \tau_L^2} = \sqrt{(1.29\,\text{ns})^2 + (1.63\,\text{ps})^2 + (4.76\,\text{ns})^2} \\ &= 4.93\,\text{ns} \Rightarrow 32.3\,\text{MHz} \end{aligned}$$

With compensation,

$$\text{comp } \tau = \sqrt{(885\,\text{ps})^2 + (163\,\text{ps})^2 + (1.54\,\text{ns})^2} = 1.78\,\text{ns} \Rightarrow 89\,\text{MHz}$$

Shunt or series inductive peaking at the output could increase the bandwidth above 100 MHz.

The choice of $\zeta = \sqrt{2}/2$ leads to MFA response for only one pole pair. However, in this circuit, the CB pole and CE input pole also influence pole angle. The combination is not exactly MFA but is slightly overpeaked from MFA due to the additional poles.

The value of C_μ is not given in manufacturer's data sheets. C_μ depends on V_{BC}; its value cannot be specified except at a given voltage. The typical value is at $V_{BC} = 0$ V and is $C_{jc}(0)$ or C_{j0}. Then C_μ is the junction capacitance C_{jc}:

$$C_{jc} = \frac{C_{j0}}{\left[1 - (V_{BC}/\phi_C)\right]^m}$$

where ϕ_C is the barrier potential and m depends on the junction grading. Typically,

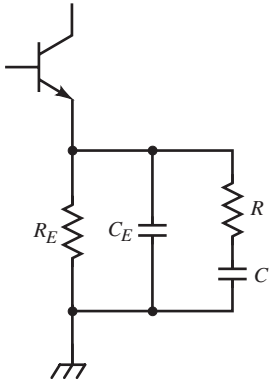
$$\phi_C = 0.75\,\text{V}, \quad m = 0.5$$

For linearly graded junctions, $m = 0.33$. The SPICE parameters corresponding to these quantities are

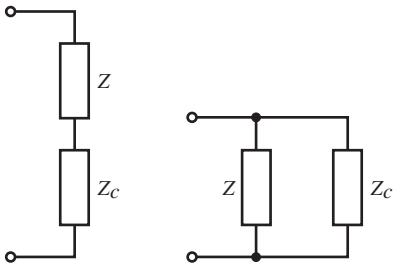
$$C_{j0} \Rightarrow \text{CJC}, \quad \phi_C \Rightarrow \text{VJC(PC)}, \quad m \Rightarrow \text{MJC(MC)}$$

For a normal-mode NPN, V_{BC} is negative, and the subtraction in C_{jc} is the addition of a positive voltage ratio. With the typical values, at 5 V reverse bias a junction has one third of the capacitance it has at zero volts.

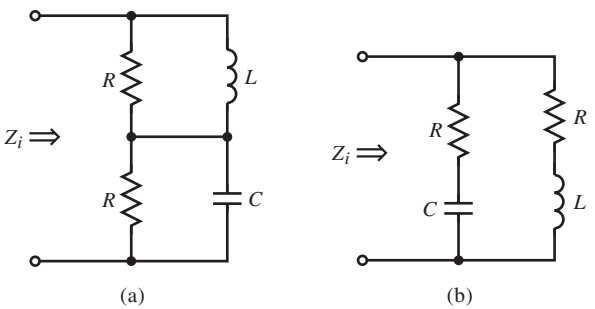
Compensation Network Synthesis



The emitter compensation network, shown above, was chosen because its impedance provided the poles and zero required for compensation. In general, compensation requirements are known in terms of poles and zeros, whereas the topology and equations for element values are unknown. The compensation of hf-gyrated impedances is simplified by deriving the equivalent circuits and noting that all-pass networks can be formed with them. Because of the need for compensation networks, a few common synthesis techniques can be useful.



One compensation technique is to make a reactive network with impedance Z resistive and thus independent of frequency. A compensating impedance Z_c is added in series (a) or parallel (b) with Z.

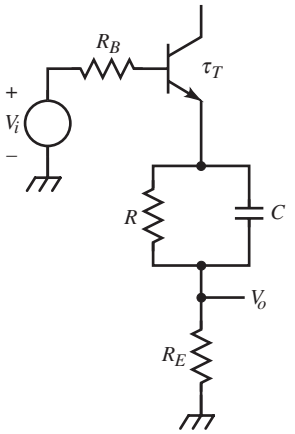


For the above networks, $Z_i = R$ whenever

$$L/C = R^2$$

A shunt *RL*, such as an emitter-gyrated base resistance, is compensated by adding in series a shunt *RC*. Then from the all-pass constraint equation,

$$R = R_B, C = \frac{\tau_T}{R_B}$$

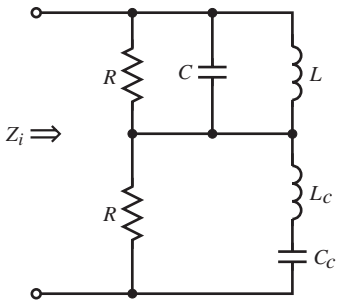


This creates a resistive voltage divider at the emitter. If the load is capacitive, *C* is made larger to compensate the divider. Then *Z* of the network from the emitter must still be equivalent to a shunt *RC*, satisfying the all-pass constraint.

More generally, a shunt *RC* can similarly be compensated by adding a shunt *RL* in series with it. Or a series *LC* can be compensated by the two resistors shunting each of them, as in (a) above. In these cases, $Z_i = R$ when the all-pass constraint is satisfied.

A series *RC*, such as a base-gyrated emitter resistance, can be compensated by shunting it with a series *RL*, as in (b). The input of a common-collector (CC) with significant C_μ at the internal base node and resistance in the collector supply return line forms a series *RC* that can compensate the series *RL* of the base. Base *R* and *L* are both parasitic (r_b' contributes to *R* and lead inductance to *L*) and can be made to appear resistive at *b*′. Adjustment of the collector and base resistance and series base inductance make it possible to satisfy the all-pass constraint. For $r_b' = 100\ \Omega$, $L_b = 10$ nH, and $C_\mu = 3$ pF, the impedance at *b*′ toward

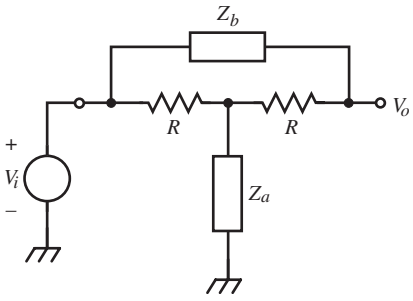
b is resistive and is 100 Ω when the collector resistance is 100 Ω and 20 nH is added to the base circuit. This added inductance might be from the inductive peaking of the previous stage.



More complicated networks are also possible, such as the one shown above. Here, the conditions for a resistive input of $Z_i = R$ are

$$L/C = L_c/C_c = R^2$$

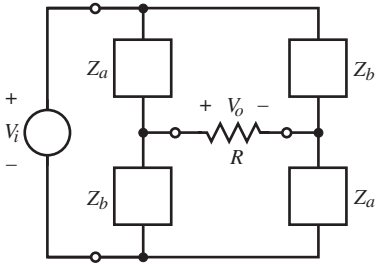
This network is a compensated shunt *RLC*, or a shunt *RC* in which the capacitor, C_c, has parasitic inductance, L_c. This is typical of electrolytic capacitors, which have resonant frequencies around 1 MHz or, for higher frequencies, any capacitors with leads. A monolithic multilayer ceramic capacitor has about 5 nH of inductance with leads of a length needed for insertion into circuit-board holes. Leadless chip capacitors are sometimes required for good high-frequency bypass or decoupling of the power supply terminals of active devices.



A common network, the *bridge-T*, is shown above. A special case of the bridge-T is applied in T-coil compensators. The input is $Z_i = R$ when

$$Z_a \cdot Z_b = R^2 \Rightarrow \frac{V_o}{V_i} = \frac{1}{Z_b/R + 1}$$

Z_a and Z_b must be dual reactances (or *reciprocal impedances*); if Z_a is capacitive, Z_b must be inductive.



Another common network, shown above, is the *lattice* or *bridge*. The lattice network has the same resistive input conditions but a different transfer function:

$$Z_a \cdot Z_b = R^2 \Rightarrow \frac{V_o}{V_i} = \frac{1 - Z_a/R}{1 + Z_a/R}$$

Two general methods can be applied to the synthesis of passive networks with a given $Z(s)$:

1. *Partial-fraction synthesis*, for factored poles or zeros.
2. *Continued-fraction synthesis*, for explicit network topology.

Partial-fraction synthesis is based on partial-fraction expansion of $Z(s)$. This requires factoring the denominator. If the numerator is easier to factor, expand $Y(s) = 1/Z(s)$ instead. Various network topologies can result, however, and other design considerations could constrain the choice of topology.

Example: Partial-Fraction Network Synthesis

A network is described by the following $Z(s)$:

$$Z = R \cdot \frac{s\tau_3 + 1}{(s\tau_1 + 1) \cdot (s\tau_2 + 1)}$$

Z can be written as

$$Z = \frac{A}{s\tau_1 + 1} + \frac{B}{s\tau_2 + 1}$$

When Z is partial-fraction expanded, A and B are

$$A = R \cdot \left(\frac{\tau_1 - \tau_3}{\tau_1 - \tau_2} \right), \quad B = R \cdot \left(\frac{\tau_2 - \tau_3}{\tau_2 - \tau_1} \right)$$

A and B are resistances. The expanded Z can be written as

$$Z = \frac{1}{s(\tau_1/A) + 1/A} + \frac{1}{s(\tau_2/B) + 1/B} = [1/s(\tau_1/A)] \| A + [1/s(\tau_2/B)] \| B$$

Z has the form of two shunt RCs in series.

Continued-fraction synthesis produces a continued-fraction form of Z. A desirable feature of continued-fraction impedances is that the topology is explicit in the form of the expression. The general procedure is to invert rational expressions that are less than one and to divide by synthetic division.

Example: Continued-Fraction Network Synthesis

An impedance of the form of the previous example is

$$Z = R \cdot \frac{sc + 1}{as^2 + bs + 1}$$

R is first multiplied to the numerator in s and the fraction inverted:

$$Z = \frac{1}{\left(\dfrac{as^2 + bs + 1}{sRc + R}\right)}$$

The denominator is now greater than one and can be divided to become

$$Z = \cfrac{1}{s\left(\dfrac{a}{Rc}\right) + \left(\dfrac{bc - a}{Rc^2}\right) + \cfrac{1}{\left(\dfrac{c^2 - bc + a}{sRc + R}\right)}}$$

The remainder is divided by $s \cdot R \cdot c + R$ and then inverted. Division is carried out once again, and the final continued fraction results:

$$Z = \cfrac{1}{s \cdot \left(\dfrac{a}{Rc}\right) + \left(\dfrac{bc - a}{Rc^2}\right) + \cfrac{1}{sR \cdot \left(\dfrac{c^3}{c^2 - bc + a}\right) + R \cdot \left(\dfrac{c^2}{c^2 - bc + a}\right)}}$$

The terms in the denominator of Z are admittances. The capacitance a/Rc is in parallel with resistance

$$R \cdot \left(\frac{c^2}{bc - a}\right)$$

and with the series RL, where the resistance is

$$R \cdot \left(\frac{c^2}{c^2 - bc + a}\right) = R_S$$

and the inductance is $c \cdot R_S$.

Continued fractions represent shunt topologies, and partial fractions represent series topologies. In continued-fraction expansion, divisions are executed the usual way, beginning with the highest power in *s*. If, instead, division begins with the lowest power, the divisor grows in powers of *s*. The remainder is then a power of *s* higher than that of the dividend. This approach does not produce the circuit topology of the desired network but can be useful in approximating a network by truncating the quotient. No s^2 term represents a circuit element, but it can be transformed into an equivalent network with negative element values. (See *Designing Dynamic Circuit Response*, "High-Frequency Impedance Transformations.")

Example: Differentiator

The circuit shown below in (a) is a differentiator with resistive input and output, suitable for transmission-line coupling. The circuit is analyzed by transforming it to (b) using the T-coil theory of inductive peaking. This topology is recognized as an approximate bridge-T when $R_L = R$, and the coupled inductors are ideal as a transformer element. Then leakage inductances $L_1 = L_2 = 0$, and $Z_a = sM$, where M is the mutual inductance of L_1 and L_2, and $Z_b = 1/sC$. Applying the all-pass constraint,

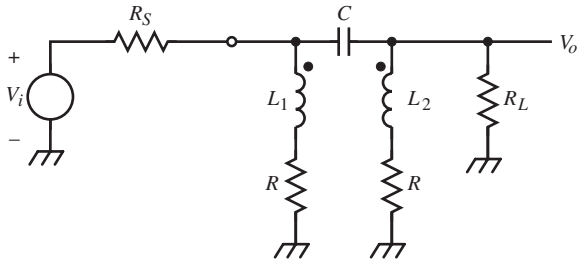
$$\frac{V_o}{V_i} = \frac{1}{Z_b/R + 1} = \frac{sRC}{sRC + 1}$$

and $Z_i = R$.

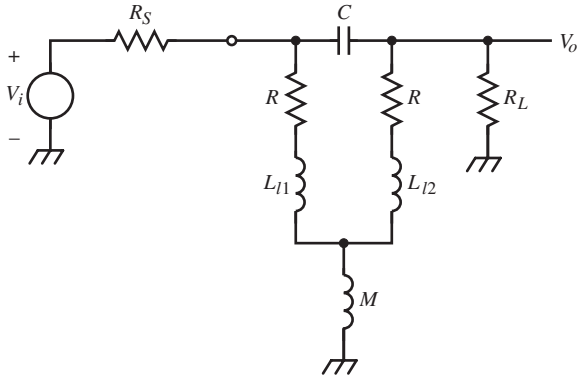
The circuit differentiates to a frequency of $1/RC$. A 50 Ω transmission line can drive the differentiator and be terminated properly when $R = 50\ \Omega$. For wideband differentiation to 100 MHz, $RC = 1/2\pi \cdot (100\ \text{MHz}) = 1.59$ ns. Then $C = 31.8$ pF and M must be the pulse transformer magnetizing inductance:

$$M = R^2C = 79.6\,\text{nH}$$

The output-terminating resistance can be the characteristic impedance of another transmission line. In other words, the differentiator can be inserted



(a)

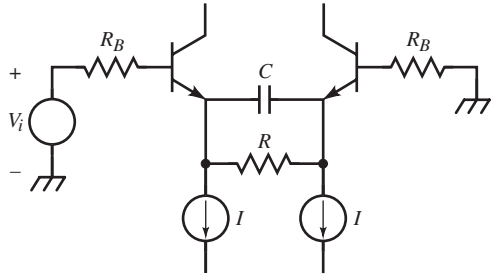


(b)

into a transmission line without causing discontinuity. For example, high-speed differentiation of a ramp can be performed by driving a 50 Ω coaxial cable into a 50 Ω test section wherein the differentiator has been built. This section then terminates in the 50 Ω input of an oscilloscope vertical amplifier.

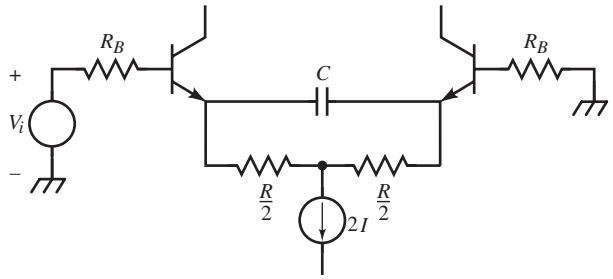
DIFFERENTIAL-AMPLIFIER COMPENSATION

The two-transistor differential amplifier (diff-amp) shown below has, at each BJT emitter, a voltage divider formed by R and Z_e of the other transistor. This divider can be compensated by shunting R with a compensating C. The approach is the same for both Π (above) and T (below) emitter networks, since they are equivalent. The design goal is to compensate

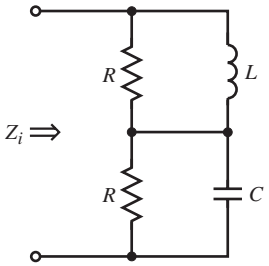


$$Z_e = \frac{Z_\pi + Z_B}{\beta(s)+1} = \frac{r_e}{s\alpha_o\tau_T + 1} + \left(\frac{Z_B}{\beta_o + 1}\right)\cdot\left(\frac{s\tau_\beta + 1}{s\alpha_o\tau_T + 1}\right)$$

Below f_T, the first term is approximately r_e. In the hf region, Z_B is gyrated.



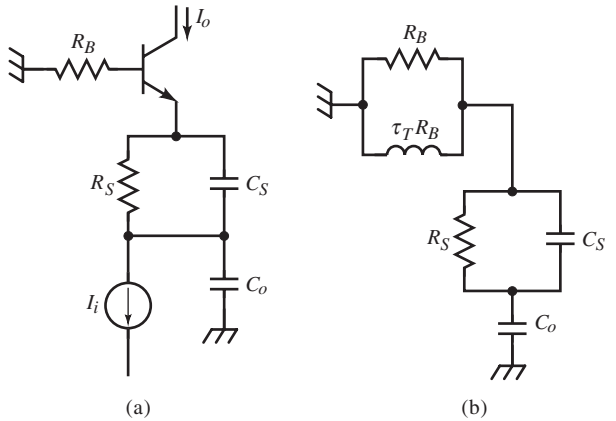
The situation is similar to that of the CB of the cascode. For $Z_B = R_B$, the resistive network shown below is formed.



To present a resistance to the emitter of the other BJT, the compensator time constant is

$$R \cdot C = \frac{\tau_T \cdot R_B}{R} \Rightarrow C = \frac{\tau_T \cdot R_B}{R^2}$$

In practice, significant stray capacitance is often present at the emitter (or current source) node(s). The circuit is then represented by the circuit shown below, in which C_o is the stray capacitance.



The previous results for the cascode can be applied. From the equation for ζ, increasing C_o decreases ζ, causing the circuit to be less damped.

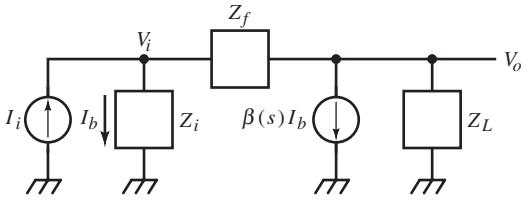
Transistor compensation techniques apply to differential as well as single-ended amplifiers. For balanced differential amplifiers, the shared networks between sides experience twice the drive of a single-ended network. Where gain is involved, their effective impedances are halved.

SHUNT-FEEDBACK AMPLIFIER DESIGN

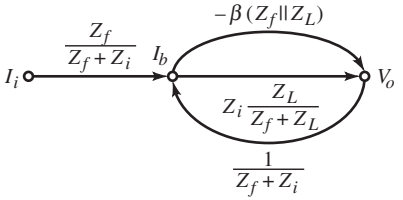
The frequency-independent shunt-feedback amplifier was analyzed in *Designing Amplifier Circuits* and the closed-loop transresistance was derived. A shunt-feedback topology was also considered in *Designing Dynamic Circuit Response* with capacitive Z_f. The general topology has a transimpedance of

$$\frac{V_o}{I_i} = (Z_f \| Z_i) \cdot \frac{(Z_f \| Z_L)(Z_f \| -R_m)}{1 + [(Z_f \| Z_L)/(Z_f \| -R_m)] \cdot [Z_i/(Z_f + Z_i)]}$$

A wideband realization of a shunt-feedback amplifier is the BJT amplifier shown below, with frequency-dependent β and general impedances.



This amplifier is equivalent to the previously developed topology but is explicit in $\beta(s)$ so that hf behavior can be analyzed.



Assuming a general $\beta(s)$ at first, the transimpedance is found by solving its flow graph.

$$\frac{V_o}{I_i} = Z_L \cdot \frac{Z_i - \beta \cdot Z_f}{Z_f + Z_i + (\beta + 1) \cdot Z_L}$$

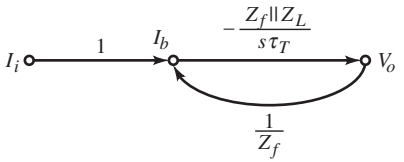
The first term in the numerator represents the passive noninverting path to the output, and the second term represents the active path through the BJT. The input impedance is derived after r_{in} in *Designing Amplifier Circuits*, "Shunt-Feedback Amplifier Feedback Analysis":

$$Z_{in} = \frac{Z_i \| Z_f}{1 + GH} = Z_i \cdot \frac{Z_f + Z_L}{Z_f + Z_i + (\beta + 1) \cdot Z_L}$$

Similarly, following the quasistatic development,

$$Z_{out} = Z_L \left\| \left(\frac{Z_f + Z_i}{\beta + 1} \right) = Z_L \cdot \frac{Z_f + Z_i}{Z_f + Z_i + (\beta + 1) \cdot Z_L} \right.$$

The hf BJT model has $Z_\pi = 0$. If Z_π is not Z_i, then it is a shunt contributor to it, and when set to zero causes Z_i to be zero. The hf approximation of $\beta(s)$ is $1/s\,\tau_T$. Making these hf approximations to the flow graph, the hf flow graph for the hf transimpedance results.



Solving for the transimpedance, based on this flow graph,

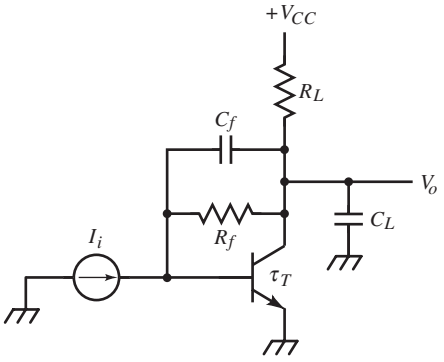
$$\left.\frac{V_o}{I_i}\right|_{\text{hf}} = -\frac{Z_f}{s\tau_T \cdot (1 + Z_f/Z_L) + 1} = -Z_L \| \alpha_{\text{hf}} \cdot Z_f$$

Z_{in} is trivially zero, and the hf output impedance is

$$Z_{out}(\text{hf}) = \frac{s\tau_T \cdot Z_f}{s\tau_T \cdot (1 + Z_f/Z_L) + 1} = -\frac{(V_o/I_i)|_{\text{hf}}}{\beta_{\text{hf}}}$$

These general results are applied to a less general single-BJT shunt-feedback amplifier, shown below, where

$$Z_f = R_f \left\| \frac{1}{sC_f} = \frac{R_f}{sR_fC_f + 1}; \quad Z_L = R_L \right\| \frac{1}{sC_L} = \frac{R_L}{sR_LC_L + 1}$$



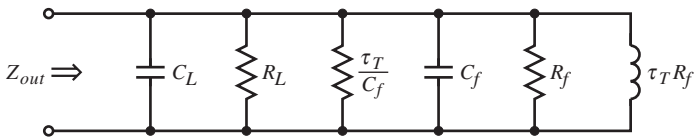
Substituting into $(V_o/I_i)_{\text{hf}}$ and simplifying yields

$$\frac{V_o}{I_i} = \frac{R_f}{s^2\tau_T R_f\left(C_f + C_L\right) + s\left[\tau_T\left(1 + R_f/R_L\right) + R_f C_f\right] + 1}$$

Before analyzing this transimpedance, easily obtained from Z_{out}(hf) is the expression for output impedance:

$$\begin{aligned} Z_{out} &= \frac{s\tau_T R_f}{s^2\tau_T R_f\left(C_f + C_L\right) + s\left[\tau_T\left(1 + R_f/R_L\right) + R_f C_f\right] + 1} \\ &= \left(R_f \left\| R_L \left\| \frac{\tau_T}{C_f} \right.\right.\right) \left\| \frac{1}{s\left(C_f + C_L\right)} \right\| s\tau_T \cdot R_f \end{aligned}$$

This impedance is represented topologically below as a shunt *RLC*.



From V_o/I_i, the complex pole-pair damping ratio is

$$\zeta = \frac{b}{2\sqrt{a}} = \frac{\tau_T \cdot \left(1 + R_f/R_L\right) + R_f \cdot C_f}{2\sqrt{\tau_T \cdot R_f \cdot \left(C_f + C_L\right)}}$$

The desired response is set by choosing ζ and solving for the element that is free to be varied. For a given transistor, τ_T is fixed, and the required gain for the stage is also determined by the amplifier design strategy. This sets R_f. Biasing constraints can set R_L, and C_f is partly determined by C_μ of the BJT. This leaves C_L; its minimum is determined by the capacitive loading of the next stage. The best design insight is gained from the loci of poles when various elements are allowed to vary parametrically. The loci are determined by extending the technique of "Loci of Quadratic Poles" in *Designing Dynamic Circuit Response*.

The pole locus is described by geometric equations in the real and imaginary s coordinates, σ and ω. The two basic equations are

$$\sigma = -\left(\frac{b}{2a}\right)$$

$$\omega^2 = \frac{1}{a} - \left(\frac{b}{2a}\right)^2$$

and

$$\omega^2 + \sigma^2 = \frac{1}{a}$$

Starting first with R_f as parameter, substitute into σ from V_o/I_i:

$$\sigma = -\frac{b}{2a} = -\frac{\tau_T\left(1 + R_f/R_L\right) + R_f C_f}{2\tau_T R_f\left(C_f + C_L\right)}$$

Solving for R_f,

$$R_f = \frac{-\tau_T R_L}{\tau_T + R_L C_f + 2\tau_T R_L\left(C_f + C_L\right)\sigma}$$

The $1/a$ equation leads to

$$\omega^2 + \sigma^2 = \frac{1}{a} = \frac{1}{\tau_T R_f\left(C_f + C_L\right)} = -\frac{\tau_T + R_L C_f + 2\tau_T R_L\left(C_f + C_L\right)\sigma}{\tau_T^2 R_L\left(C_f + C_L\right)}$$

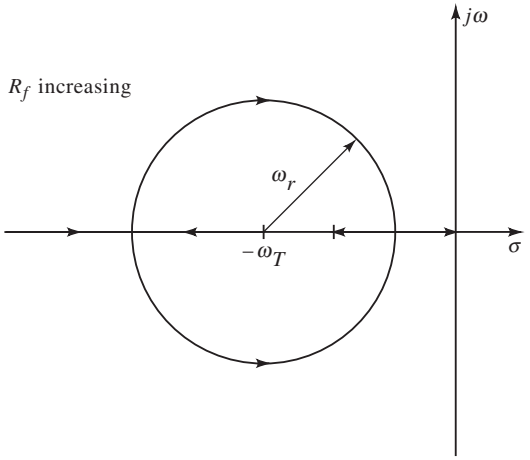
Simplifying the right side and collecting σ terms on the left side yields

$$\omega^2 + \left(\sigma^2 + \frac{2\sigma}{\tau_T}\right) = -\frac{\tau_T + R_L \cdot C_f}{\tau_T^2 \cdot R_L \cdot \left(C_f + C_L\right)}$$

Completing the square in σ by adding $1/\tau_T^2$ to both sides,

$$\omega^2 + \left(\sigma + \frac{1}{\tau_T}\right)^2 = -\frac{\tau_T + R_L \cdot C_f}{\tau_T^2 \cdot R_L \cdot (C_f + C_L)} + \frac{1}{\tau_T^2} = \left(\frac{1}{\tau_T} \cdot \sqrt{\frac{R_L \cdot C_L - \tau_T}{R_L \cdot (C_f + C_L)}}\right)^2 = \omega_r^2$$

This equation describes a circular locus centered at $\sigma_0 = -1/\tau_T = -\omega_T$ with a radius of ω_r. Unlike previous loci, the circle does not contain the origin but is offset to the left, as shown below.



In practice, usually,

$$R_L \cdot C_L >> \tau_T$$

and ω_r simplifies to

$$\omega_r \cong \frac{1}{\tau_T} \cdot \sqrt{\frac{C_L}{C_f + C_L}}, \quad R_L \cdot C_L >> \tau_T$$

At $R_f = 0$, the poles are at $-\infty$ and $-\omega_T$. As R_f increases, they move together and become complex, following the circular locus with decreasing σ. At the σ axis they split; as $R_f \to \infty$, one goes to zero and the other to

$$-\frac{\tau_T / R_L + C_f}{\tau_T \cdot (C_f + C_L)}$$

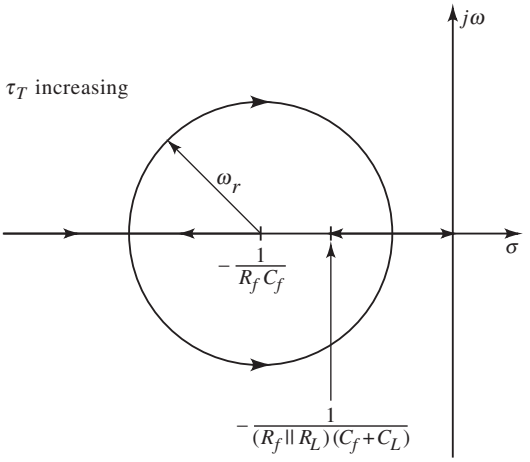
The locus equation for parameter τ_T is derived similarly to that for R_f. It is also circular, centered at $-1/R_f \cdot C_f$ with

$$\omega_r = \frac{1}{R_f \cdot C_f} \cdot \sqrt{1 - \frac{R_f \cdot C_f}{(R_f \| R_L) \cdot (C_f + C_L)}}, \quad R_f \cdot C_f < R_L \cdot C_L$$

This constraint is required for real ω_r. When R_L is replaced by a current source, this simplifies to

$$\omega_r = \frac{1}{R_f \cdot C_f} \cdot \sqrt{\frac{C_L}{C_f + C_L}}, \quad R_L \to \infty$$

At $\tau_T = 0$, the poles are at $-1/R_f \cdot C_f$ and $-\infty$. As τ_T increases, the poles move together.



The low-frequency pole actually increases in frequency with a slower transistor. The poles form a circular locus with σ decreasing until they separate along the σ axis. As $\tau_T \to \infty$, one pole is at the origin and the other is at

$$-\frac{1}{(R_f \| R_L) \cdot (C_f + C_L)}$$

The locus of C_L is derived similarly and has the same form as before. The center of the circle is in the RHP at

$$\sigma_0 = \left(\frac{C_L}{C_f}\right) \cdot \frac{1}{R_f \cdot C_f - \tau_T \cdot (1 + R_f/R_L) \cdot (C_L/C_f)}$$

with

$$\omega_r = \frac{1}{R_f \cdot C_f - \tau_T \cdot (1 + R_f/R_L) \cdot (C_L/C_f)} \cdot \sqrt{\frac{R_f \cdot C_f - \tau_T \cdot (1 + R_f/R_L) \cdot (C_L/C_f) + \tau_T \cdot (C_L/C_f)^2}{\tau_T}}$$

Often it is the case that

$$C_L << C_f, \quad R_f \cdot C_f >> \tau_T \cdot \left(1 + \frac{R_f}{R_L}\right)$$

and Z_f dominates the response. The locus for C_L simplifies to

$$\sigma_0 \cong \left(\frac{C_L}{C_f}\right) \cdot \frac{1}{R_f \cdot C_f}, \quad \omega_r \cong \frac{1}{\sqrt{\tau_T \cdot R_f \cdot C_f}}$$

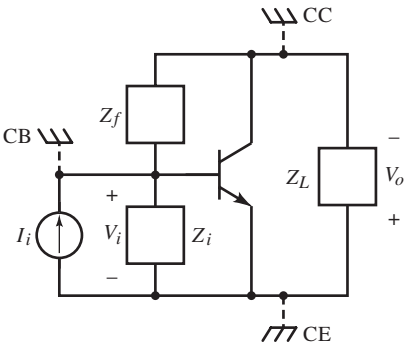
For left half-plane (LHP) poles, $\omega_r > \sigma_0$, and the equation for σ_0 satisfies this condition for typical values.

The locus for C_f would also be useful but cannot be put in a form similar to the previous parameters. When r_b' is taken into account, the transimpedance gains a pole, and the denominator is cubic. The effect of r_b' is to slightly undamp the amplifier. Under the above conditions,

$$\frac{b}{2a} = \frac{1}{2\tau_T} = \text{constant}$$

Variation in R_f or C_f moves the poles along the vertical locus. An increase in C_f or R_f reduces pole angle, though pole radius is also reduced somewhat. When $C_L = 0$ and $R_L \to \infty$, the poles are located at $-1/R_f \cdot C_f$ and $-\omega_T$.

The astute observer will recognize that Z_{out} from the equivalent circuit is similar to what would be expected of emitter impedance due to a gyrated shunt *RC* in the base. The coincidence is not accidental. Bruce Hofer observed that the topology of the shunt-feedback amplifier and emitter-follower are identical.

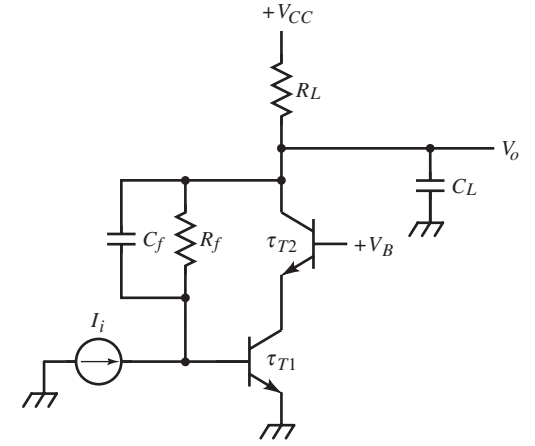


The figure above shows a general BJT circuit with impedances shunting each BJT terminal pair. Which of the three configurations (CE, CB, or CC) is represented depends on where ground is placed, as shown. Since ground is an arbitrary 0 V reference node, the port impedance of the V_o node is the same for CC and CE. For the CC, Z_f is a shunt base impedance; for the shunt-feedback CE, the output port is the same, except in relation to ground. Because port impedances are independent of grounding conventions, Z_{out} is the same.

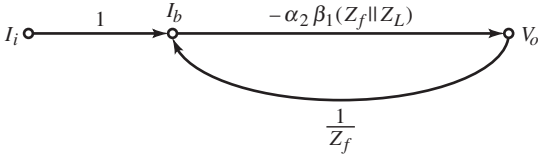
This observation also applies to inverting and noninverting op-amp configurations; their topology is identical. In the case of the noninverting op-amp, the input is added (in series) with the amplifier.

Shunt-Feedback Cascode and Darlington Amplifiers

The unavoidable presence of C_μ in the shunt-feedback amplifier has led to a minimization of parasitic feedback capacitance by use of a cascode amplifier as the forward path *G*, shown in (a) below.



(a)



(b)

This involves the additional factor α_2 of the CB. The flow graph (b) reduces to a transimpedance of

$$\frac{V_o}{I_i} = -R_f \Big/ \Big\{ s^3 \big[\tau_{T1}\tau_{T2} R_f \big(C_f + C_L \big) \big] + s^2 \big[\tau_{T1}\tau_{T2} \big(1 + R_f / R_L \big) + \tau_{T1} R_f \big(C_f + C_L \big) \big] + s \big[\tau_{T1} \big(1 + R_f / R_L \big) + R_f C_f \big] + 1 \Big\}$$

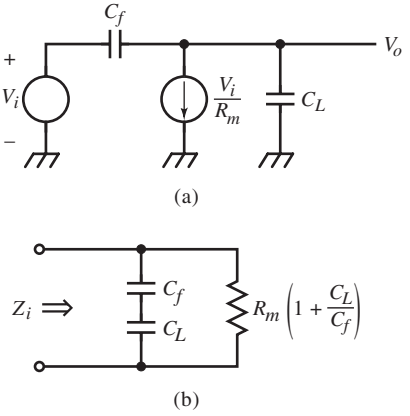
An additional pole due to $\alpha_2(s)$ results in a cubic denominator. For

$$R_f \cdot C_f >> \tau_{T1}, \tau_{T2}$$

the denominator of V_o/I_i can be factored approximately to yield

$$\frac{V_o}{V_i} \cong -\frac{R_f}{\big(sR_f C_f + 1 \big) \big(s^2 \big[\tau_{T1}\tau_{T2} \big(1 + C_L / C_f \big) \big] + s \big[\tau_{T1} \big(1 + C_L / C_f \big) \big] + 1 \big)}$$

In the complex pole factor, τ_{T2} is present only in a and not b. This results in a circular locus for τ_{T2} and a vertical (constant-σ) pole locus for the other parameters.



An amplifier of transconductance $1/R_m$ with load capacitance but with no R_f is shown in (a). The voltage gain for a general load impedance Z_L is

$$\frac{V_o}{V_i} = -\left(\frac{Z_L}{R_m}\right)\cdot\left(\frac{-sR_mC_f+1}{sZ_LC_f+1}\right)$$

For $Z_L = 1/sC_L$,

$$\frac{V_o}{V_i} = -\frac{-sR_m\cdot C_f+1}{sR_m\cdot\left(C_f+C_L\right)}$$

This result has a familiar RHP zero. The equation can be envisioned as an uncompensated voltage divider in which the upper impedance is a shunt RC consisting of C_f and $-R_m\cdot(1 + V_o/V_i)$, and the lower impedance is due to C_L.

The input impedance is

$$Z_{in} = \frac{V_i}{I_i} = \frac{1}{sC_f\cdot\left(1-V_o/V_i\right)}$$

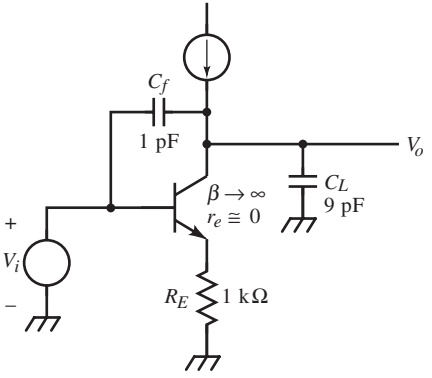
and has the form of a Miller capacitance. After substitution of V_o/V_i and simplification,

$$Z_{in} = \left[\frac{1}{s(C_f \| C_L)}\right] \| R_m \cdot \left(1 + \frac{C_L}{C_f}\right)$$

The equivalent input network is shown in (b) above. This is a surprising result because the input has a shunt resistance, but the actual circuit has only a capacitive connection to the input node. This circuit models CE amplifiers with significant C_μ and load capacitance.

Example: CE with Load and Shunt-Feedback Capacitances

The BJT amplifier has an input impedance determined by the previous equation for Z_{in}. Assume that the transistor has $\alpha = 1$ and $r_e \cong 0$. Then the transresistance of the BJT is approximately R_E, or 1 kΩ. The base input impedance is infinite and can be ignored. The analysis applies to the lf region and is not due to hf effects.



An intuitive explanation begins by noting that the path through C_f directly presents a capacitance of the series combination of C_f and C_L, or 0.9 pF, to the input. Second, if a 1 V step is applied to the input, it generates 1 mA of collector current. This current divides between C_f and C_L since they form a capacitive current divider. For $C_f = 1$ pF and $C_L = 9$ pF, 0.9 mA flows through C_L whereas

0.1 mA flows from the input. This component of current corresponds to the resistive path in Z_{in} circuit (b) above. The resistance is

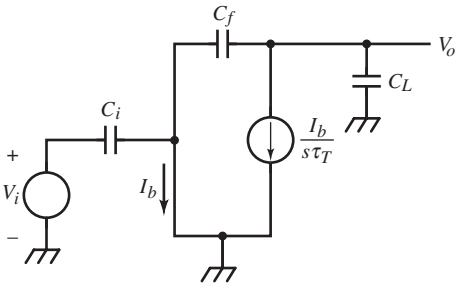
$$\frac{1\,\text{V}}{0.1\,\text{mA}} = 10\,\text{k}\Omega$$

Now to check this result, use the equation for Z_{in}. The series combination of capacitances follows immediately. The resistance should be

$$R_m \cdot \left(1 + \frac{C_L}{C_f}\right) = 1\,\text{k}\Omega \cdot \left(1 + \frac{9\,\text{pF}}{1\,\text{pF}}\right) = 10\,\text{k}\Omega$$

and the two solutions agree.

The amplifier of (a) above is now modified to conform to a BJT shunt-feedback circuit for the hf region.



For a general output impedance Z_L, the voltage gain is

$$\frac{V_o}{V_i} = -\frac{s\beta \cdot C_i}{s(\beta+1)\cdot C_f + 1/Z_L}$$

Substituting $\beta_{\text{hf}} = 1/s\tau_T$ and $Z_L = 1/sC_L$,

$$\frac{V_o}{V_i} = -\left(\frac{C_i}{C_f}\right) \cdot \frac{1}{s\tau_T \cdot (1 + C_L/C_f) + 1}$$

To avoid C_μ, the amplifier can be made a cascode. Then the above circuit is modified by multiplying α_2 of the CB to the current source. This leads to a complex pole-pair in the voltage gain:

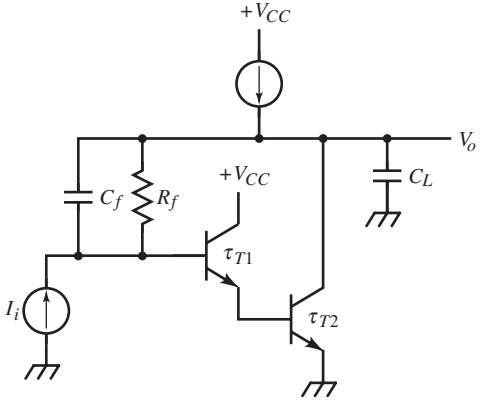
$$\frac{V_o}{V_i} = -\left(\frac{C_i}{C_f}\right)\cdot\frac{1}{s^2\tau_{T1}\tau_{T2}\left(1+C_L/C_f\right)+s\tau_{T1}\left(1+C_L/C_f\right)+1}$$

The response can be designed in the usual way for complex pole pairs. The damping ratio is

$$\zeta = \frac{1}{2}\cdot\sqrt{\left(1+\frac{C_L}{C_f}\right)\cdot\left(\frac{\tau_{T1}}{\tau_{T2}}\right)}$$

Comparing this with ζ for a single-BJT shunt-feedback amplifier, C_f has the opposite effect of damping the response. In this circuit, increasing C_f undamps it.

Finally, the idea of isolating Z_f from stray capacitance can be extended to the Darlington configuration, shown below.

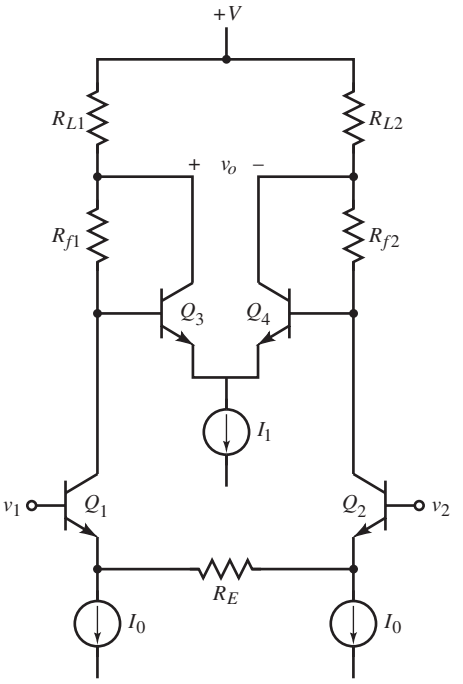


The transimpedance has a cubic denominator and a zero at ω_{T1}. Again assuming a dominant time constant of R_fC_f, the approximate transimpedance is

$$\frac{V_o}{I_i} \cong \frac{R_f\left(s\tau_{T1}+1\right)}{\left(sR_fC_f+1\right)\left\{s^2\tau_{T1}\tau_{T2}\left[1+\left(C_{\mu 2}+C_L\right)/C_f\right]+s\tau_{T1}\left[1+\left(C_{\mu 2}/C_f\right)\right]+1\right\}}$$

The complex pole-pair has the same radius as the shunt-feedback cascode but has a different expression for ζ. Because C_L appears only in a, its root locus has a circle in the LHP whereas the other parameters have a constant-σ locus. Therefore, C_L is the component of choice for adjustment of response.

CLOSURE



Fast amplifiers consist of several fast stages in cascade. Various combinations are used in fast amplifier designs, such as the above diff-amp in cascade with a differential shunt-feedback amplifier. Subsequent chapters continue development of the fast amplifier repertoire, adding more precise, yet fast, amplifier stages.

2

Precision Amplification

Circuit speed is basic to analog circuit performance, and designers seek to extend bandwidth "from dc to daylight." Yet speed is not the only performance criterion. In audio, bandwidth that is much beyond human hearing degrades performance due to increased noise; the important measure of performance is *fidelity*, the precise reproduction of the input waveform. In this looser sense, *precision* means ideal analog waveform processing.

Causes of Degradation in Precision

For amplification, any effect beyond scaling of the input quantity adds error to the scaling function and degrades precision. The output quantity for amplification can be expressed as

$$X_{out} = \sum_{i=0}^{\infty} a_i X_{in}^i + X_{\text{noise}}$$

where, ideally,

$$X_{out} = k \cdot X_{in} = (a_1 + \varepsilon) \cdot X_{in}$$

and k is the exact scaling coefficient. Basic causes of error are scaling inaccuracy ε and nonlinear terms of X_{in}, called *distortion*. Any contribution to X_{out} that is not from X_{in} is noise. Noise generated by the circuit itself is *intrinsic*; noise from other electrical activity (electromagnetic interference, or EMI) that interferes with circuit activity is *extrinsic*. A special case of error is the constant term a_0, called *offset error*, due to bias element inaccuracy, static thermal effects, or even noise.

Heat causes noise and distortion. Thermal effects due to the ambient temperature of the circuit environment (or *thermal drift*) cause offset error and affect dynamic circuit parameters. Changes in power dissipation with waveform variation cause self-heating of elements whose dynamic parameters change with temperature and cause dynamic thermal effects, or *thermals*.

Bandwidth limitations also degrade precision by failing to scale all frequency components of X_{in} equally (due to magnitude roll-off) and by shifting them relative to each other in time (nonlinear group delay). This causes $X_{out}(t)$ to have a different waveshape from $X_{in}(t)$, and functional accuracy is degraded. A fundamental trade-off occurs between accuracy and bandwidth due to the finite gain-bandwidth product (f_T) of active devices. Greater accuracy requires more settling time, resulting in a lower effective bandwidth. This also applies to feedback amplifiers because accuracy is related to loop gain. As loop gain decreases with frequency, loop accuracy degrades. Therefore, larger bandwidth is sometimes necessary to achieve low-frequency accuracy.

Intrinsic Noise

Intrinsic noise is generated by the components of a circuit. Noise is characterized in the frequency domain by its *spectral density*, or power spectrum. This is the Fourier transform of its autocorrelation function,

$$R_{XX}(\tau) = \int_{-\infty}^{+\infty} x(t) \cdot x(t+\tau) \cdot dt$$

For random functions, the greater the time separation of two points on the waveform, the more likely they are to be independent. When $\tau = 0$, the points coincide and R_{XX} (0) is maximum. For continuous $x(t)$, the closer two points are chosen in time, the more likely they are related in amplitude and tend to reinforce, resulting in larger R_{XX}.

The rms noise voltage or current x_n is related to its noise spectral density $\tilde{x}_n$ by

$$x_n^2 = \int_{-\infty}^{+\infty} \tilde{x}_n^2(f) \cdot df$$

where the expected value of x_n (or the average x_n) is zero. For a limited frequency band of $\Delta f = f_h - f_l$, the limits of this integration are the band limits.

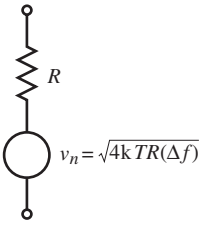
Thermal noise arises due to the random motion of particles. Thermal energy is kinetic, and particles move in random paths as they collide. From statistical thermodynamics, the average kinetic energy in any one direction per particle is proportional to kT, the same kT as in the diode v-i relation, where

$$k = \text{Boltzmann's constant} = 1.38 \times 10^{-23}\ \text{J/K}$$

and T is absolute temperature in degrees Kelvin. At $T = 300$ K (27°C), $kT = 4.14 \times 10^{-21}$ J. This kind of noise occurs in electrical circuits because the thermal vibration of ions in a crystalline lattice causes them to collide with free electrons and exchange energy. This is manifested at the macrolevel as resistance. With no electric field applied, the lattice is at thermal equilibrium, and the average current is zero. Yet the instantaneous voltage fluctuates about zero and produces a rms voltage across the resistance of

$$\text{rms thermal } v_n = \sqrt{4 \cdot k \cdot T \cdot R \cdot (\Delta f)}$$

where Δf is the frequency band in which the noise occurs. A resistance model that accounts for thermal noise is shown below.

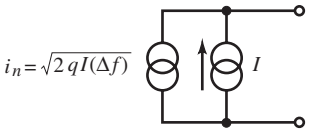


Thermal noise is broadband and has a flat spectral density.

With an applied field, average current is due to the average motion of the electrons. Since current is the aggregate motion of many charged particles, it also fluctuates randomly, as does pressure in gaseous systems. This noise due to current is *shot noise*. Fluctuation in the instantaneous current has an rms value of

$$\text{rms shot } i_n = \sqrt{2 \cdot q_e \cdot I \cdot (\Delta f)}$$

where $q_e = 1.60 \times 10^{-19}$ C (the electron charge), and I is the average current. Shot noise has a flat spectral density up to optical frequencies. A shot-noise current source is modeled below.



Both thermal and shot noise varies by the square root of the bandwidth. Manufacturers usually specify noise by its spectral density, or $\text{noise}/\sqrt{\text{frequency}}$, in units of $\text{V}/\sqrt{\text{Hz}}$ or $\text{A}/\sqrt{\text{Hz}}$. The bandwidth Δf is that of an ideal bandpass filter, not an actual circuit. The *noise equivalent bandwidth* Δf is consequently different from any actual bandwidth. For a single-pole bandwidth of f_{bw}, the noise equivalent bandwidth is

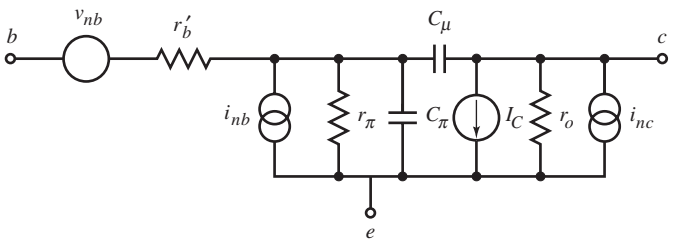
$$\int_0^\infty \left(\frac{1}{1+(f/f_{bw})^2} \right) \cdot df = \left(\frac{\pi}{2} \right) \cdot f_{bw} \cong 1.57\ f_{bw}$$

where the integrand is the square of the single-pole transfer function.

The total noise due to multiple sources is combined according to the rms summation formula:

$$x_{\text{rms}} = \sqrt{\sum_i x_{\text{rms}}^2(i)}$$

where the $x_{\text{rms}}(i)$ are independent (or more generally, uncorrelated) rms noises.



The bipolar junction transistor (BJT) noise model has three noise sources, with spectral noise densities:

$\tilde{v}_{nb}$ = thermal noise of base resistance, $r_b' = \sqrt{4kTr_b'}$

$\tilde{i}_{nb}$ = shot noise due to base current, $I_B = \sqrt{2q_eI_B}$

$\tilde{i}_{nc}$ = shot noise due to collector current, $I_C = \sqrt{2q_eI_C}$

The BJT model can be solved for output noise voltage due to the BJT alone, with a shorted input (so that $R_S = 0$) and open output (so that $R_L \to \infty$). Ignore C_μ. The BJT output noise voltage density $\tilde{v}_{no}$ has three terms corresponding to the three noise sources just listed:

$$\tilde{v}_{no}^2 = \left(\tilde{i}_{nc}\cdot r_o\right)^2 + \left\{\frac{\mu^2}{\left[\omega\cdot r_{b\pi}\cdot C_\pi\right]^2+1}\right\}\cdot\left[\left(\tilde{v}_{nb}\cdot\left(\frac{r_\pi}{r_\pi+r_b'}\right)\right)^2+\left(\tilde{i}_{nb}\cdot r_{b\pi}\right)^2\right]$$

$$\uparrow$$

noise voltage-gain^2

where $r_{b\pi} = r_\pi \| r_b'$ and $\mu = r_o/r_m$. The noise-gain break frequency is at $1/r_{b\pi}\cdot C_\pi$; the noise equivalent bandwidth, $\omega \cong 1.57/r_{b\pi}\cdot C_\pi$ or $f = 1/(4\cdot r_b'\cdot C_\pi)$. The thermal noise voltage of r_b' is attenuated by the divider formed by r_b' and r_π in the base circuit.

In the BJT circuit, any external resistance in the input (base-emitter) loop, when referred to the base, is r_S. It includes r_b' and replaces it in the above noise-voltage equation. Also, for finite load resistance R_L, collector resistance $r_o \| R_L$ replaces r_o. To avoid calculation of $\tilde{i}_{nc}\cdot r_o$, the μ transform can be applied to refer this noise to the emitter:

$$\frac{\tilde{i}_{nc}\cdot r_o}{\mu+1} = \tilde{i}_{nc}\cdot\left(r_m\|r_o\right) \cong \tilde{i}_{nc}\cdot r_m\Big|_{\mu>>1} = \tilde{i}_{nc}\cdot\frac{kT}{q_eI_C}, \quad \mu >> 1$$

Substituting the noise current density, the *emitter shot-noise* voltage density results:

$$\tilde{v}_n = \sqrt{2\cdot q_e\cdot I_C}\cdot\left(\frac{k\cdot T}{q_e\cdot I_C}\right) = k\cdot T\cdot\sqrt{\frac{2}{q_e\cdot I_C}}$$

This noise voltage decreases with increasing emitter current until *input thermal noise* in r_b' dominates.

$$\tilde{v}_n = \sqrt{4k \cdot T \cdot r_S} \cdot \left(\frac{r_\pi}{r_\pi + r_S} \right)$$

The voltage-divider factor can usually be omitted because it is desirable to keep r_S small to reduce thermal noise.

Finally, the third noise term is the *input shot noise* of r_S:

$$\tilde{v}_n = \sqrt{2q_e \cdot I_B \cdot r_S}$$

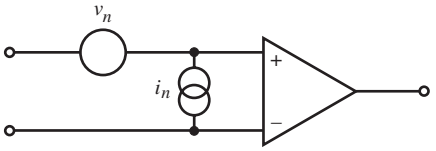
The total noise voltage density is the rms sum of these three noise sources.

Given $I_C = \beta_o \cdot I_B$, the emitter and input shot noise terms both vary with I_C. An optimum I_C can be found by differentiating the total noise voltage density squared and solving for the current when set to zero. Minimum-noise

$$I_C = \sqrt{\beta_o} \cdot \left(\frac{k \cdot T}{q_e \cdot r_S} \right)$$

The higher r_m of field-effect transistors (FETs) produces more noise than BJTs, especially at low frequencies (below 10 Hz), except when source resistance is high. Then base current causes dominant shot noise in the BJT source resistance.

A BJT-input amplifier noise model is shown below.



The BJT input v_n is due to collector shot noise in r_m and thermal noise in r_b'. The shot noise in r_S, which is external to the amplifier, is accounted for by $i_n \cdot r_S$. An optimum r_S contributes equal amounts of shot and thermal noise, or

$$\sqrt{4k\cdot T\cdot r_S} = \sqrt{2q_e\cdot I_I}\cdot r_S \Rightarrow \text{optimum } r_S = 2\cdot\left(\frac{kT}{q_e\cdot I_I}\right)$$

At a circuit temperature of 300 K (27°C), the thermal noise voltage density is

$$\text{thermal } \tilde{v}_n \text{ at 300 K} = \left(129\,\text{pV}/\sqrt{\text{Hz}\Omega}\right)\cdot\sqrt{R} \cong 4\text{nV}/\sqrt{\text{Hz}}, 1\text{k}\Omega$$

and the shot-noise current density is

$$\text{shot } \tilde{i}_n = \left(566\times10^{-12}\sqrt{A}/\sqrt{\text{Hz}}\right)\cdot\sqrt{I} \cong 18\text{pA}/\sqrt{\text{Hz}}, 1\text{mA}$$

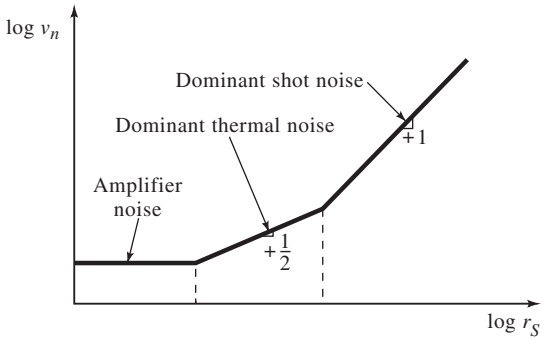
Example: Op-Amp Input Noise

An op-amp has an equivalent input noise voltage density of $20\,\text{nV}/\sqrt{\text{Hz}}$ and noise current density of $0.1\,\text{pA}/\sqrt{\text{Hz}}$. In the noninverting configuration, no resistors are required for a × 1 buffer; the total noise voltage is $20\,\text{nV}/\sqrt{\text{Hz}}$. A ×(−1) inverting op-amp configuration with 100 kΩ input and feedback resistors has a total equivalent input noise voltage at 300 K of

$$\sqrt{\left(20\,\text{nV}/\sqrt{\text{Hz}}\right)^2 + 2\left(40.7\,\text{nV}/\sqrt{\text{Hz}}\right)^2 + \left(5.00\,\text{nV}/\sqrt{\text{Hz}}\right)^2} = 61.1\,\text{nV}/\sqrt{\text{Hz}}$$

or about three times (10 dB) as much noise. The input noise gain is that of the noninverting configuration, as it also is for offset voltage, because the noise source is in series with the op-amp input terminals.

The op-amp bias current also generates shot noise in the resistors, which must be included if significant. For 1 nA of bias current, the shot noise current is $17.9\,\text{fA}/\sqrt{\text{Hz}}$ through 100 kΩ, producing a shot-noise voltage of $1.8\,\text{nV}/\sqrt{\text{Hz}}$. This is negligible compared with thermal noise voltage. If the bias current splits between the input and feedback resistors, the shot noise remains unchanged.



The graph shows amplifier noise voltage as a function of r_S on a log-log plot. The amplifier noise sets a baseline independent of external resistance. Then as r_S increases, the thermal noise becomes significant at the thermal noise corner, where the curve slopes upward with a slope of one half. At the shot-noise corner, shot-noise voltage dominates, is proportional to r_S, and the slope is one. The shot-noise corner shifts with input current and can lie below the thermal-noise corner.

Thermal and shot noise are wideband, or *white* (just as white light is wideband), and are associated with resistance. Another kind of noise, associated with semiconductor surface leakage and conductors in general, is $1/f$ noise, also called *flicker* noise. It rolls off to a break frequency at the circuit white-noise level. This break frequency is typically 1 to 10 Hz for BJT noise voltage and 10 Hz to 1 kHz for FET noise current. FET $1/f$-noise break frequencies are typically 50 to 100 times higher, and that for CMOS is around 100 times higher.

The $1/f$ noise rolls off at a slope of $-1/2$ and breaks at a frequency where it intersects white noise. This break or corner frequency, f_f, specifies both voltage and current $1/f$ noise. For either voltage or current white-noise density, $\tilde{x}_{nw}$, $1/f$ noise density, $\tilde{x}_{nf}(f)$, is

$$\tilde{x}_{nf} = \tilde{x}_{nw} \cdot \sqrt{\frac{f_f}{f}}$$

From the integral expression for x_n^2, the band-limited noise is

$$x_n^2 = \int_{f_l}^{f_h} \tilde{x}^2 df = \tilde{x}_{nw}^2 \cdot f_f \cdot \ln\left(\frac{f_h}{f_l}\right)$$

The total noise is the rms sum of white and flicker noise, or

$$\tilde{x}_n = \sqrt{\tilde{x}_{nf}^2 + \tilde{x}_{nw}^2} = \tilde{x}_{nw} \cdot \sqrt{\frac{f_f}{f} + 1}$$

Manufacturers' data sheet noise specifications are based on this equation. Finally, x_n, the noise quantity, is found by substituting the density of x_n into x_n^2 and integrating. This results in

$$x_n = \tilde{x}_{nw} \cdot \sqrt{f_f \cdot \ln\left(\frac{f_h}{f_l}\right) + (f_h - f_l)}$$

For static (dc) amplifiers, $f_l = 0$, but this results in infinite noise. A practical lower limit is the thermal drift frequency, usually a fraction of a hertz. A typical f_l is 10 mHz. Below this, low-frequency noise is indistinguishable from drift.

With formulas for calculating rms noise, we sometimes are interested in what value the peak noise can achieve. The ratio of peak to rms values is the *crest factor*,

$$\text{crest factor} = \frac{x_m}{x_{\text{rms}}}$$

For Gaussian noise, the probability that $|x|$ exceeds a given crest factor k_c is

$$p(|x| > k_c x_{\text{rms}}) = \text{erfc}\left(\frac{k_c}{\sqrt{2}}\right) = \sqrt{\frac{2}{\pi}} \cdot \int_{k_c}^{\infty} e^{-\left(|x|/\sqrt{2} \cdot x_{\text{rms}}\right)^2} \cdot d\left(\frac{|x|}{x_{\text{rms}}}\right)$$

For $k_c = 1$, $p = 32\%$; $k_c = 2$, $p = 4.6\%$; and for $k_c = 6.6$, $p = 0.1\%$.

Another peculiar kind of noise is *burst* or *popcorn noise*. It is caused by process-dependent wafer surface effects and is manifested as random rectangular pulses or shifts in the static level, typically below f_f, that add (algebraically) to the other noises.

Extrinsic Noise: Radiation and Crosstalk

Extrinsic noise is due to other electrical activity in the environment of the affected circuit. This noise is generally referred to as *electromagnetic interference* and can be caused by the following:

- electromagnetic radiation (far field)
- crosstalk (near field)
- conduction

Crosstalk can be either magnetic or electric in origin because in the near field it is not coupled as an electromagnetic wave.

Radiated EMI is reduced by shielding the circuit – by enclosing it with conductive material. The shield presents an impedance discontinuity to an incident wave because its characteristic impedance is much less than that of a wave in free space,

$$Z_w = \frac{E}{H} = 377\,\Omega \text{ in free space}$$

A 1 MHz wave (in air) impinging upon a copper sheet has a transmitted electric field that is 2 millionths (−114 dB) that of the incident wave; the transmitted magnetic field strength is twice the incident field. The characteristic impedance of copper is 0.37 mΩ ∠45°, or nearly zero.

The wave-induced current in the shield flows at the surface and falls off exponentially with penetration depth into the shield material. This depth is characterized by a length constant (the spatial equivalent of a time constant) called the penetration, or *skin depth*, δ:

$$\delta = \sqrt{\frac{2}{\omega \cdot \mu \cdot \sigma}}$$

where μ is permeability and σ is conductivity of the shield material. The surface resistance of the shield material is the same as the static resistance of the material with a thickness equal to the skin depth. The surface resistance is $1/\sigma \cdot \delta$.

This resistance is equal to the surface reactance. The surface impedance is analogous to the characteristic impedance of a transmission line and is

$$Z_s \cong \sqrt{\frac{j\omega \cdot \mu}{\sigma}} = \sqrt{\frac{\omega \cdot \mu}{\sigma}} \angle 45°, \quad \sigma >> \omega \cdot \varepsilon$$

where ε is the permittivity (dielectric constant). For a good conductor such as copper, steel, or aluminum, the condition of Z_s is easily satisfied. The attenuation due to reflection is then

$$R = 20 \cdot \log\left(\frac{Z_w}{4 \cdot Z_s}\right) = 20 \cdot \log\left(\frac{Z_w}{4} \cdot \sqrt{\frac{\sigma}{\omega \cdot \mu}}\right), \quad Z_w >> Z_s$$

Reflection decreases with frequency and permeability but increases with conductivity.

For poor conductors, a significant amount of the wave is transmitted through the shield and is attenuated more by absorption than by reflection. Absorption is resistive (ohmic) loss in the shield due to wave-induced eddy currents flowing in a resistive material. Absorption increases with shield thickness, frequency, permeability, and conductivity.

Shields cannot be completely closed surfaces because wires, circuit boards, and adjustment tools pass through them. These openings are also entrances for interfering waves and act as waveguide apertures. The relevant criterion is that openings have maximum lengths (in any dimension) that are much less than the wavelength of interfering radiation. Seams along case openings and metal slots and holes act as slot antennas. For a maximum slot dimension of

$$d < \frac{\lambda}{2}$$

where λ is the wavelength, the slot acts as a dipole antenna and passes frequencies above the cutoff frequency,

$$f_c = \frac{u_c}{2d}, \quad u_c = \text{speed of light} = 3.0 \times 10^8 \text{ m/s} = 30 \text{ cm/ns}$$

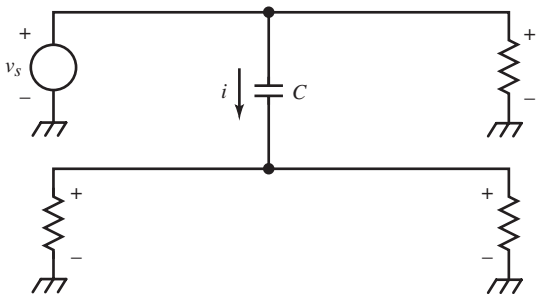
For attenuation of waves above frequency

$$f = \frac{u_c}{\lambda}$$

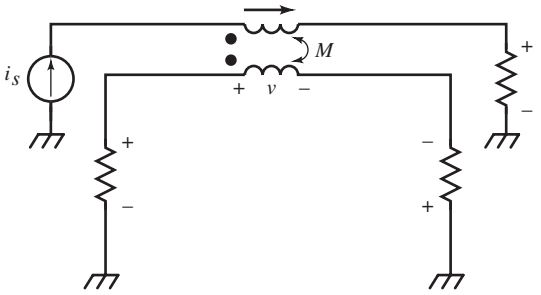
d must be much smaller than λ; $d << \lambda/100$ for 60 dB of attenuation at the highest frequency of interest. At the slot, the electric field is maximum at the center because eddy currents in the shield must go farthest around the slot from the center of its maximum length. This creates the largest voltage drop from center to center across the slot. A 0.5 m slot has a shielding effectiveness of only 5.65 dB at 300 MHz. EMI gaskets and EMI-tight enclosure construction techniques provide conductance continuity across slots, thereby maintaining shielding effectiveness.

The near field is the space less than $\lambda/2\pi$ from the radiation source. Wave impedance Z_w depends on the impedance of the source. Electric fields have high wave impedance (E is large in the expression for Z_w), and common metal shields are conductive enough to reflect them effectively. Magnetic fields have low wave impedance. Consequently, the impedance mismatch with the shield is not as great, and reflection as a shielding mechanism is not adequate. Low-frequency magnetic shielding is largely absorptive and requires high-permeability shield material to divert the field.

Shield reflection of electric fields in the near field decreases linearly on a log-log plot with frequency and distance from the source, whereas magnetic field shielding effectiveness increases linearly with frequency and source distance. At low frequencies, electric fields are well shielded by high-conductivity shields, but magnetic fields are attenuated less as frequency decreases. Therefore, low-frequency magnetic fields, such as from power-supply transformers, commonly cause the most trouble in shielding.



Near-field interference is often due to coupling between conductors in a cable or on a circuit board, as shown above. The mutual currents flow through parasitic capacitance between the two lines and cause the same polarity of voltage at each end of the line. The amount of capacitance increases with line length and decreases with spacing.



For inductive coupling, voltage across the secondary loop resistance is due to mutual inductance M between the loops. M depends on the amount of shared area of the loops and their proximity. Coupling between parallel lines causes the source end of the second loop to have the same polarity of induced voltage as the primary loop, whereas at the load end it is inverted from that of the primary.

The magnitude of coupled noise depends on the rate of change of source quantities. For inductive coupling, 1 mA/ns induces 1 mV/nH. For capacitive coupling, 1 V/ns causes 1 mA/pF of current.

To predict the amount of crosstalk, estimation of M and C are required. Analytic solutions for crosstalk in the above circuits are unwieldy. We seek an intuitive ability to estimate, and that requires simplified approximations. The length of the lines must be less than $\lambda/4$. At 100 MHz this is 75 cm. The spacing of the two coupled wires is expressed in distance between wires w and number of wire diameters d as w/d, and the separation of the lines for each loop from a ground plane is h. When separate return lines are used instead of ground plane, h is half the separation of waveform and return lines (assumed the same for both loops). Except for very close spacing, where w/d is close to one, increasing wire diameter does not appreciably increase C and affects M even less.

As length increases but remains under $\lambda/10$, both C and M increase linearly. For $w/d = 10$ and $h/d = 2$, C/length $\cong$ 1 pF/m. For $h/w = 1$, M/length $\cong$ 15 nH/m, and C increases with h approximately linearly for $w/d > 10$ and $h < 100$. M is more sensitive to h than C because increasing h increases loop area. M increases sublinearly with h/w on a log-log plot. More significantly, both C and M decrease quadratically with w/d, or at −40 dB/dec of separation. (A 10 times change in w produces a 100 times change in M or C.) Also, for length less than $\lambda/10$, the amount coupled by C or M increases linearly with frequency.

The impedance of free space, 377 Ω, sets the boundary between which kind of crosstalk dominates. High-impedance (>377 Ω) sources have dominant electric fields and dominant capacitive crosstalk. For low-impedance (<377 Ω) circuits, inductive crosstalk dominates.

Visual or geometric estimation of crosstalk when wiring a circuit or designing a circuit-board layout based on the preceding rules of thumb is often both adequate and at the practical limit of what can be reliably estimated.

Wiring and Layout Guidelines.

1. Maximize spacing between waveform lines.
2. Minimize areas between waveform lines and their return (ground) paths, running the two as close together as possible. When this is too difficult to do on a single layer, run a ground line alongside either or both of the source and receiving lines. This reduces the relative coupling about five times. Or, add a ground plane layer to the board. For long lines, use twisted-pair cable. In flat cable, ground every other conductor. On an existing board, reduction of coupling can be experimentally verified by gluing a ground line of magnet wire between coupling traces.
3. Run waveform lines perpendicular on opposite sides of the board. This reduces inductive coupling to intersecting areas formed by line pairs overlapping on opposite sides. Capacitive coupling area is reduced to the crosspoints of lines on opposite sides.
4. For high-speed circuits, confine waveforms to transmission lines or controlled-impedance environments.

Transmission lines can be as simple as twisted pairs of wires or coaxial cable. Twisted-pair cable has a characteristic impedance of

$$Z_n \cong \frac{120\,\Omega}{\sqrt{\varepsilon_r}} \cdot \ln\left(\frac{h}{d}\right)$$

where h is the distance between conductor centers, and d is the conductor diameter. The length per twist, or *pitch*, does not affect Z_n, only the propagation delay time, because the line is longer with smaller pitch. For typical wire insulation, some values of dielectric constant are given in the following table.

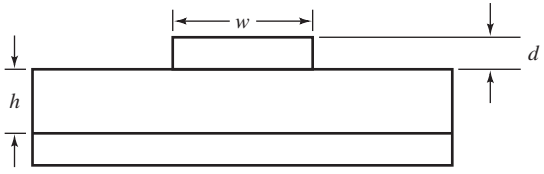
Dielectric Material	ε_r
Air	1.0
Polytetrafluoroethylene (Teflon)	2.1
Polyethylene	2.3
Polystyrene	2.5
Polyvinyl chloride (PVC)	3.5
Epoxy resin	3.6
Epoxy glass	4.7
Polyester (Mylar)	5.0
Polyurethane	7.0

Typically, Z_n is between 50 and 100 Ω.

Another form of transmission line is wires run over a ground plane. In this case,

$$\text{wire-over-ground } Z_n \cong \frac{60\,\Omega}{\sqrt{\varepsilon_r}} \cdot \ln\left(\frac{4h}{d}\right) \cong \frac{138\,\Omega}{\sqrt{\varepsilon_r}} \cdot \log\left(\frac{4h}{d}\right)$$

where ε_r is the relative permeability (dielectric constant) of the medium between wire and ground plane (usually circuit board or air), h is the distance between wire and ground plane, and d is wire diameter. For epoxy glass boards, ε_r typically is 4.7 (G-10 material), and for air it is 1. An order of magnitude increase in h/d results in an increase in Z_n of 138 Ω. A typical wire-over-ground line ($h/d \cong 1.3$) is about 100 Ω and has 3.54 nH/cm and 0.315 pF/cm.



On circuit boards, parallel-plate or *microstrip* transmission lines can be made of the board itself, as shown above. With a ground-plane width much greater than the waveform line width w and for $h < \lambda/4$,

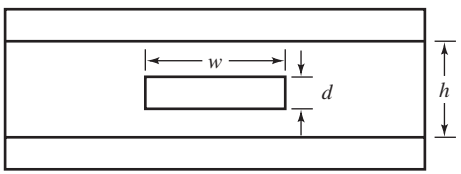
$$\text{microstrip } Z_n \cong \frac{87\,\Omega}{\sqrt{\varepsilon_r + 1.41}} \cdot \ln\left(\frac{(5.98)\cdot h}{0.8\cdot w + d}\right)$$

For 1 oz/ft^2 copper board traces, $d \cong 35.6\,\mu\text{m}$ to $38.1\,\mu\text{m}$; for 2 oz/ft^2 board, thickness is $71.1\,\mu\text{m}$ to $76.2\,\mu\text{m}$. A $50\,\Omega$ line on a 1/16 inch thick board is 2.62 mm wide (or about 0.1 in.) using 1 oz/ft^2 copper.

Propagation delay time is

$$t_{pd} \cong 3.34 \cdot \sqrt{(0.475)\cdot \varepsilon_r + 0.67}\ \text{ns/m}$$

Z_n of microstrip lines is typically about half that of wire-over-ground lines, or 50 to 100 Ω.



A symmetrical form of microstrip line, or *stripline*, can be made on a multilayer circuit board, as shown above, where the waveform conductor is embedded between two ground planes. For stripline,

$$\text{stripline } Z_n \cong \frac{60\,\Omega}{\sqrt{\varepsilon_r}} \cdot \ln\left(\frac{4\cdot h}{(0.67\pi)\cdot w\cdot(0.8 + d/w)}\right)$$

and

$$t_{pd} \cong 3.34 \cdot \sqrt{\varepsilon_r}\ \text{ns/m}$$

where $w/(h - d) < 0.35$ and $d/h < 0.25$. Striplines typically have the lowest Z_n, about half that of microstrip line.

A default case of a microstrip line is parallel, flat conductors of width w and thickness d, separated by circuit board material of thickness h. For $w >> h >> d$,

$$Z_n \cong \frac{377\,\Omega}{\sqrt{\varepsilon_r}} \cdot \left(\frac{h}{w}\right)$$

Finally, if none of these lines can be implemented, traces run side by side with thickness d, width w, and edge-to-edge spacing h, where $w >> d$, have

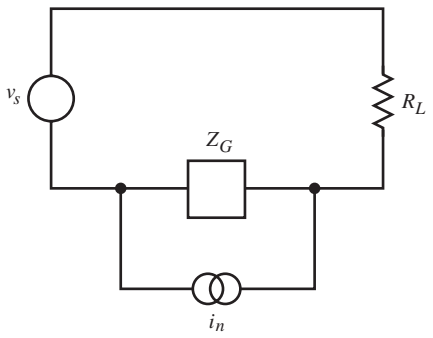
$$Z_n = \frac{120\,\Omega}{\sqrt{\varepsilon_r}} \cdot \ln\left(\frac{\pi \cdot h}{w + d}\right)$$

Flat (or ribbon) cables have parallel conductors that can be used as transmission lines. Typical flat-cable wire spacing is 0.050 inch with # 28 AWG stranded wire. One manufacturer gives the following specifications for such a cable: $Z_n = 105\ \Omega$, 41.3 pF/m, 558 nH/m, and $t_{pd} = 4.49$ ns/m. Flat cable with ground plane, 0.050 inch spacing, and # 28 AWG stranded wire has $Z_n = 65\ \Omega$, 82.0 pF/m, 558 nH/m, and $t_{pd} = 5.58$ ns/m.

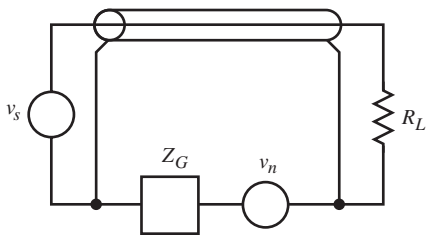
Extrinsic Noise: Conductive Interference

The third cause of EMI is conductive interference. When two circuit loops share a common path, usually a ground path, any impedance in that path develops a voltage common to both loops. Shown below, Z_G causes a noise voltage drop due to noise current i_n.

The most important general guideline for eliminating these *ground loops* is to consider the *complete* path of waveforms. The return path from the load back to the source is usually where noise enters a signal loop. This common ground return node is often distributed throughout the subsystem. If it were electrically an ideal node, its impedance Z_G would be zero. To be ideal, it must also have zero length to have zero loop area and thus no magnetic crosstalk among loops. A ground plane approximates a zero-length node.



One of the simplest and most general techniques for preventing ground loops is the *single-point ground*. Separate the return lines for each circuit loop, and run them back to their respective source grounds. Then, to connect the source grounds electrically, run separate lines to a single point where they connect, usually at the power-supply ground. This technique minimizes external signal currents in a ground return line by isolating the signal currents to individual loops.

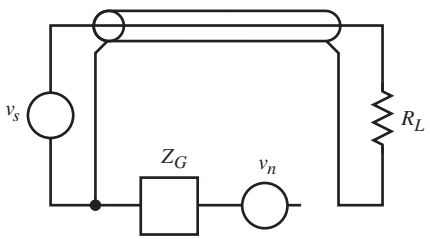


Signal return path isolation is combined with magnetic crosstalk isolation by the use of a shielded cable. Any of the transmission lines described previously can function as a shielded cable for conductive and crosstalk noise, though a constant Z_n along the cable is not required. A shielded cable provides a separate return path and minimizes loop area. For a coaxial cable, the theoretical loop area is zero, and the outer conductor (the shield) provides the return path, as shown above. Here, external ground loop noise has been Thevenized. The low shield impedance forms a current divider with Z_G and noise current flows mostly in the shield. Because of the large mutual inductance between signal and

return conductor in a shielded cable, changing source current returns mainly in the shield. For coaxial cable, $M = L_S$, the shield inductance. Above the *shield cutoff frequency*,

$$\omega_c = \frac{R_S + R_G}{L_S}$$

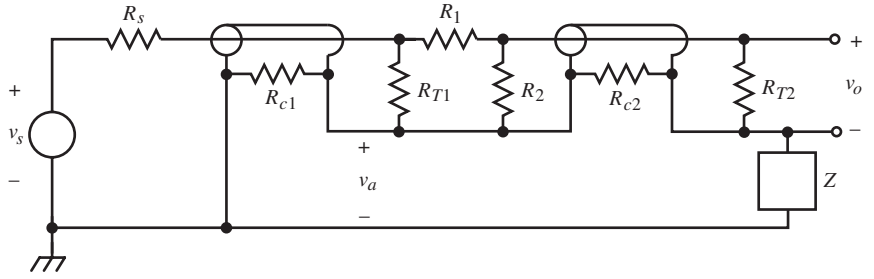
source current is in the shield; R_S is the shield resistance, and L_S the shield inductance. The fraction of shield source current falls off below ω_c to R_G/L_S, where it is then a constant, $R_G/(R_G + R_S)$. These results were obtained by solving the above circuit and using asymptotic approximations to the frequency response of the shield source-current fraction. Cutoff frequency is typically a few kilohertz.



By grounding the shield at only one end, as shown above, no external noise currents can flow in the shield. Also, no magnetically induced currents can flow in it either because it does not form a closed loop with the external ground. This is a kind of single-point ground; the source ground terminal is connected to the external ground. For applications in which grounding must occur at both ends, a small resistor (of 1 to 10 Ω) can be placed from the load return side to the external ground. This forces most of the signal current into the shield return while maintaining a relatively low-resistance path to the main ground line or plane.

All these configurations are susceptible to capacitive coupling into the shield. However, at the low impedances of ground lines ($<< 377\ \Omega$), magnetic coupling dominates.

Example: Shielded Cable Grounding with Attenuator



Shield grounding at both ends is required in equipment for safety. Instruments often require input attenuators that are connected external to the input. The circuit shows an attenuator, formed by resistors R_1 and R_2 inserted between two cables. "Barrel" attenuators with BNC connectors at each end are commonly used with 50 Ω cables that are terminated in their characteristic impedance by resistors R_{T1} and R_{T2}. The voltage v_a at the attenuator ground is the attenuated source voltage, and

$$\frac{v_a}{v_s} = \frac{R_{c1} \| (R_{c2} + Z)}{R_s + [R_{T1} \| (R_1 + R_2)] + R_{c1} \| (R_{c2} + Z)}$$

The output voltage is taken across the load cable terminator R_{T2}. $R_2 + R_{T2}$ shunt R_{c2} and form the top side of a divider with Z. The voltage across the top side of this divider is v_o, and attenuation is

$$\frac{v_o}{v_a} = \frac{(R_2 + R_{T2}) \| R_{c2}}{(R_2 + R_{T2}) \| R_{c2} + Z} \cong \frac{R_{c2}}{R_{c2} + Z}, \quad R_2 + R_{T2} >> R_{c2}$$

Now, if the load cable shield is grounded, $Z = 0$ Ω in the above equation, and v_a contributes to v_o unattenuated. With shield grounding only at the source end, Z is infinite, and attenuator error from shield resistance is zero. In effect, Z bootstraps v_o, and even small values of Z cause large improvements in attenuator accuracy.

A single ground at the source not only eliminates noise currents in the shield but preserves attenuator accuracy. Unfortunately, the chassis or earth ground of the instrument sensing v_o also must connect to the signal ground at its source to minimize its internal noise and provide a low-impedance safety ground fault path.

This problem was solved cleverly in the Hewlett-Packard model 3571A by letting Z be an inductor with a saturable core. A large power-line current through the inductor due to a grounding error would saturate it, reducing its impedance to near zero. Although attenuator error is present at 0 Hz, shield resistance is subject to the skin effect and increases with frequency, causing greatest error in the audio frequency range. In this range the inductor has enough impedance to reduce the error significantly (by 30 dB in the HP instrument). At higher frequencies, the mutual inductance of the cables decreases signal flow through Z, in effect making Z large.

Shield Effectiveness

The previously described EMI reduction techniques used a shielded cable to eliminate ground loops and inductive crosstalk, not field radiation. Shielded cables are also used in high external field environments to shield inner conductors. Within this closed shield is the signal loop. The return path is a separate conductor within the shield of a two-conductor shielded cable. Shielded cable, as a radiation shield, is used to provide a continuous shielded surface between a shielded source and shielded load. The advantages of connecting the cable shield to only the source shield apply here, but electrical isolation from the shield enclosing the load leaves an opening possibly accessible to radiation.

External fields cause noise currents in and voltages across the shield. These quantities can couple noise into the signal lines within. For sensitive applications, a second shield is placed around the first to provide additional shielding. The outer shield can be connected to other external shielding, whereas the inner shield is connected to the source ground inside.

Shielding effectiveness depends on the construction of the enclosing outer conductor, or sheath. A solid sheath is far better than braided cable but is mechanically less flexible. The tighter the braid, the better the shielding. At

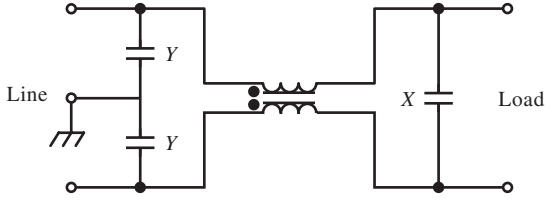
1 MHz, the shielding effectiveness of a solid sheath is about 200 dB better than a double-braided cable, which in turn is about 35 dB more effective than single-braided cable. The effectiveness of a solid sheath increases with frequency, whereas braided cable remains constant to around 1 GHz and then falls off.

Shields are connected to the enclosed circuit ground at a single point at the source. If the shield is allowed to float, large dv/dt signals can capacitively couple into it. Additional stray capacitance couples from the shield into other signal nodes. In some cases, these noise paths through the shield are amplifier feedback paths that cause instability.

Differential Transformer Common-Mode Noise Rejection

If the signal and return conductors are not magnetically coupled enough to isolate the signal path, their mutual inductance can be increased by placing the two conductors through a ferrite bead or winding several turns of both conductors together (bifilar wound) around a ferrite toroid. The high permeability of the ferrite ($\cong$ 2,000) forces the signal current to return on the other conductor. The signal current going to the load induces a voltage across the return line, via the flux linkage of the ferrite magnetic path, that causes an equal and opposite current flow in it. Differential-mode currents are passed, whereas common-mode currents are rejected.

The common-mode rejection of these ferrite transformers offers a second useful function, that of filtering. Common-mode noise currents flow in both conductors in the same direction. The high inductance due to the ferrite core forms a high-frequency filter with the load impedance for common-mode signals. The inductance for the signal currents is the transformer leakage inductance, usually 1% or less than that for common-mode currents.



Differential-mode transformers are the basis of EMI power-line filters. These filters are commonly used to keep power-line noise out of a system and keep system noise from the power line. A one-stage filter is shown above, with an *X* capacitor across the load for differential-mode filtering and two *Y* capacitors on the load side for common-mode line filtering. Another *X* capacitor on the line side improves differential-mode load-noise rejection. Attenuation (or *insertion loss*) increases with frequency to about 50 dB of rejection at 1 to 5 MHz. The break frequency increases with current rating, so low-frequency rejection is less for high-current filters.

The design objective of an EMI line filter is to pass frequencies at 50 to 60 Hz and to reject frequencies at which noise is likely to be, from 10 kHz to 100 MHz. The filter is modeled as a two-port network with source and load impedances. From the maximum power transfer theorem, the filter load port impedance should be equal to the load impedance at line frequencies and be much different at noise frequencies. This applies as well to the line port. In other words, at noise frequencies the EMI filter is a mismatching network.

The power line has a low impedance, so a high impedance input at noise frequencies is needed: a series inductor input. If the load impedance is high, load port mismatch requires a low-impedance output, which is achieved with a shunt capacitor. If load impedance is low, an additional series inductor is added. The *X* capacitor is large (0.5 μF) and shunts high-current differential-mode pulses from rectifiers and switching logic.

The mean power-line Z_n at 1 MHz is 50 Ω. For 80 % of the lines, Z_n is between 10 and 300 Ω. Below 1 MHz, line impedance falls off at about −20 dB/dec to a very low value at line frequency. Differential-mode power-line noise cannot propagate far because of wiring and transformer inductance and does not radiate far because the opposing currents cancel. The dominant power-line noise is common mode. The *Y* capacitors at the line input to the filter are intended to shunt common-mode noise to ground. *Y* capacitor size is limited for safety reasons to limit safety ground leakage current at line frequency. The transformer filters common-mode noise current from the line that the *Y* capacitors do not shunt.

Ferrite beads or toroids provide the magnetic coupling path in filter transformers; the limits of magnetic materials apply to them. As frequency increases,

signal losses in the magnetic circuit increase. As current increases to the point of magnetic saturation, permeability (and mutual inductance) decreases.

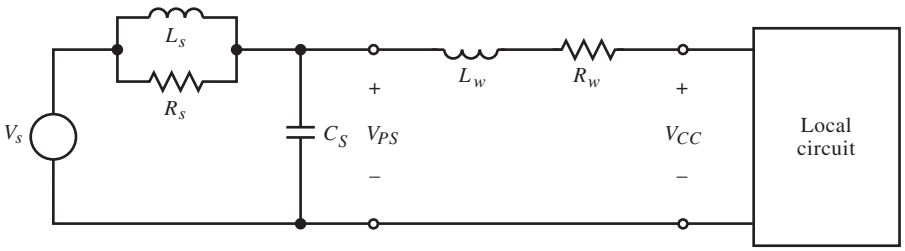
Power Distribution Noise

Now consider noise phenomena involving power distribution. Although the analytical models for circuits we have been using assume that the power supply sources are ideal voltage sources, a power-supply regulator has finite output impedance. Worse yet, the wiring required to distribute this power to the circuits is inductive and resistive. This impedance is part of the circuit to which power is delivered. The power distribution wiring is also a means for ground loops and crosstalk.

Inductive crosstalk due to varying power-supply currents can be reduced by reducing the area between the supply and ground lines. This is most easily accomplished by running them alongside each other on the same side of the circuit board or opposite each other on two sides. Circuit-board layout practice is to run signal lines perpendicular to each other on opposite sides of the board to ease interconnection and to reduce side-to-side line crosstalk. Therefore, same-side parallel power traces are usually preferred. Another approach, used commonly in digital circuits, is to run supply lines in parallel on one side, with regular spacing, connected together by a perpendicular line at the edge of the board. Then a similar pattern for ground is put on the other side and is offset relative to the power lines.

A better, but more expensive, approach is to use commercially available laminated bus bars. These are flat bars with board-mount tabs at regular spacings. They have a low, controlled impedance and are dominated by the distributed capacitance of the laminations. They not only reduce magnetic crosstalk but also eliminate parasitic line inductance. If multilayer circuit boards are feasible, a similar distribution system can be realized by making the inner two layers ground and supply planes. This has the advantages of low-impedance conduction planes, a small, enclosed area between the supply and ground currents, and the inner planes act as shields between the outer two signal layers of the board.

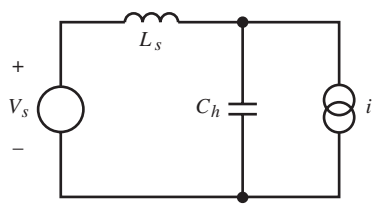
An ideal voltage source is approached locally at each circuit by reducing the Thevenin equivalent impedance of the supply terminals at the location of the circuit. This equivalent circuit is shown below. The power supply regulator has



a characteristic shunt *RL* output impedance (due to high-frequency impedance gyration) shunted by an output capacitor C_S to reduce high-frequency impedance. The distribution-wiring impedance is in series with the supply impedance. Low resistive equivalent impedance at the V_{CC} terminals is the design goal.

Local circuit activity causes current changes on the supply lines that result in voltage changes at V_{CC}. A low-impedance path for ΔI is provided by a placing a *bypass* capacitor locally across the supply. The capacitor must be a high-frequency type, typically a ceramic monolithic multilayer and usually not an electrolytic capacitor. A typical 4.7 μF aluminum electrolytic capacitor has a series resistance of about 1 Ω and a series resonance around 1 MHz. Most leaded (through-hole) ceramic capacitors have about 10 nH of inductance. Ceramic chip capacitors have less than 5 nH.

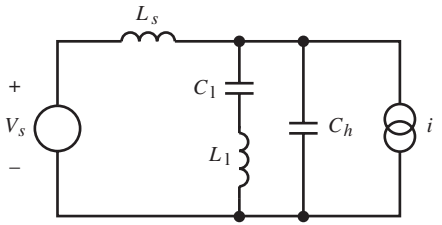
The capacitor must be large enough to present a low-impedance source over the frequency range of current. At lower frequencies, high-frequency capacitors have a capacitive reactance that is too high, due to a practical limit on their size. The problem is solved by shunting the supply with an additional large, low-frequency capacitor. The shunt combination results in a wideband low impedance.



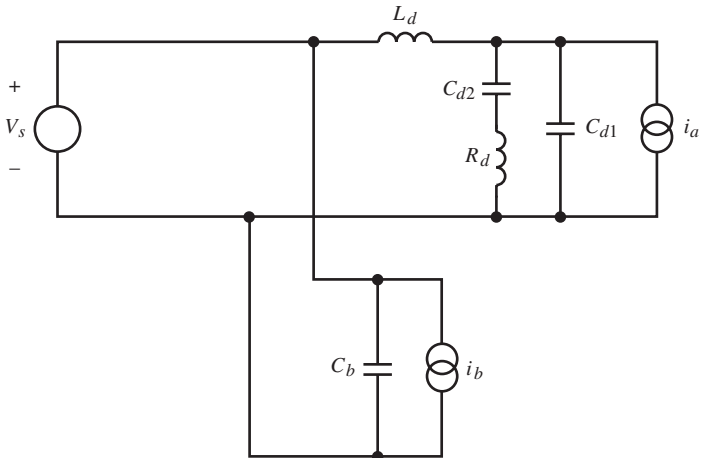
Both the inductance of the low-frequency capacitor and the noninductance of the high-frequency capacitor can cause trouble. The low-frequency capacitor inductance can resonate with the high-frequency capacitor, and the high-frequency capacitor can resonate with the line inductance. These resonant modes must be adequately damped to prevent low-level ringing on the supply line. The amplitude of this damped sinusoid is less than if no bypassing were installed, but the circuit may not be able to reject it adequately. Amplifier power-supply rejection decreases with frequency. The characteristic impedance of the resonance determines the ring amplitude of the undamped supply in the bypass circuit. For a step of current from the circuit of i, as shown above, the voltage amplitude is

$$Z_n \cdot i = \sqrt{\frac{L_s}{C_h}} \cdot i$$

This resonance can be damped by inserting resistance in series with the supply line or by increasing C_h. In the bypass circuit shown below, an electrolytic capacitor has been added.



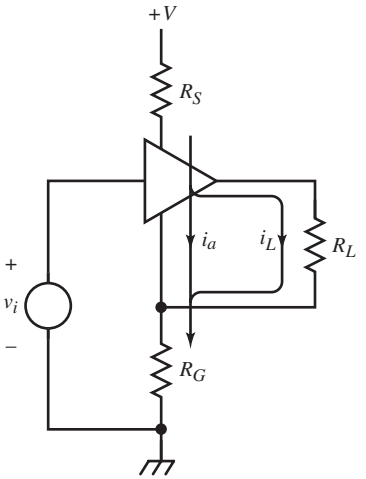
A second resonant mode is introduced by the parasitic inductance L_1, resonating with C_h. The series resistance of C_1 (not shown) is sometimes large enough to damp this resonance; otherwise, an external resistor is added in series with it. The capacitor C_1 alone, without C_h, could produce a voltage spike across L_1 on the supply with an amplitude greater than if no bypassing were present.



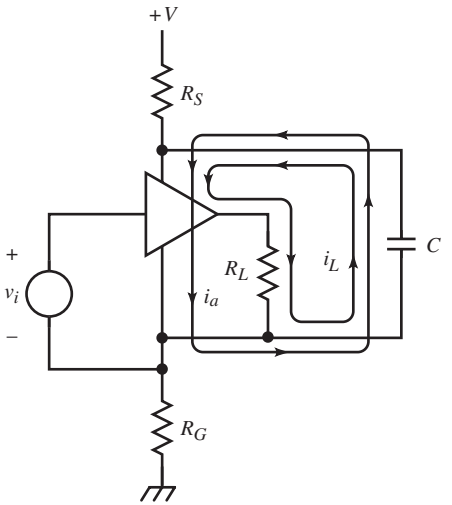
With several circuits powered from the same supply, each having different supply performance requirements, the technique of *decoupling* is sometimes used as a kind of EMI filter. A noisy circuit, one with large, fast current changes, can be isolated from the supply by the decoupling circuit shown above. An inductor or resistor in series with the line to the circuit forms a low pass filter with the capacitor shunting the circuit. It also filters the circuit noise, keeping it out of the supply. The decoupled circuit could be an output power driver, whereas the other circuits are low-level amplifiers. An alternative is to decouple the low-level circuits from a noisy supply. This may be necessary if the series impedance of a decoupling network is too large to supply adequate current to a noisy, high-current circuit. Laudie Doubrava (1978) showed that bypassing and decoupling provide distinct functions; bypassing provides a low-impedance supply source for a local circuit, and decoupling minimizes supply-coupled interactions among circuits.

Amplifier Signal Paths

Multiple signal paths occur in amplifiers and must be individually isolated. A general amplifier with ground-referenced input v_i and load R_L is connected to a voltage source, $+V$, with supply-line and ground resistances.



This circuit has two errors. First, single-point grounding has not been followed. Consequently, the voltage drop across R_G due to amplifier current i_a and load current i_L adds to v_i as noise. Second, the noise voltage, $(i_a + i_L) \cdot (R_S + R_G)$, appears across the supply terminals of the amplifier because the terminals are not bypassed.



The above diagram shows the corrected circuit. The reference terminal of v_i is returned to the input reference terminal, the negative supply terminal. Sometimes this cannot be done, and a differential amplifier (diff-amp), which has a

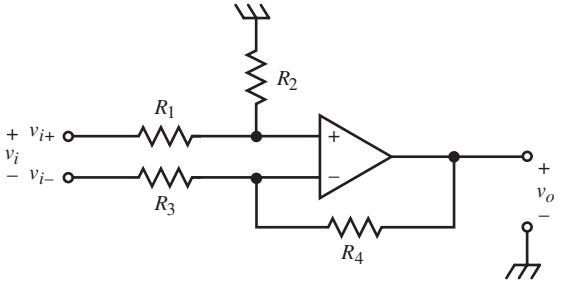
separate negative input terminal from the supply, is required to avoid ground noise at the input. The diff-amp inverting input terminal would then be connected to the negative terminal of v_i.

The second error is corrected by bypassing the amplifier supply terminals. This has two effects. It keeps signal currents of both the amplifier and load out of R_G by shunting them around the external power distribution system and directly to the positive supply terminal. And in doing so, it minimizes signal voltage changes across the supply terminals. When the negative terminal of v_i cannot be connected as shown above, the amplifier negative-supply terminal can also be bypassed to the negative terminal of v_i. This dynamically shorts the noise voltage across R_G.

This general illustration of amplifier grounding evaluation can be applied to more specific cases. In all cases, it is important to first identify the (complete) current paths involved. Then it is possible to determine the effect of these currents due to parasitic impedances in the supply paths and how to reroute current and sensing loops using isolation, bypassing, and decoupling. In some cases, no solution is possible with the existing circuitry, and a different kind of amplifier is required for isolation of signal paths.

Differential Amplifiers

The first improvement over an amplifier with a common-ground input is a differential-input amplifier. Operational amplifiers (op-amps) are an instance of diff-amps, but their gain is too large to be of use without feedback.



Design goals usually call for an amplifier with a fixed gain and nonloading differential input. These finite-gain diff-amps can be made from op-amps. A *one-op-amp diff-amp* is shown above, where

$$v_i = v_{i+} - v_{i-}$$

and

$$v_o = A_{v+} \cdot v_{i+} - A_{v-} \cdot v_{i-}$$

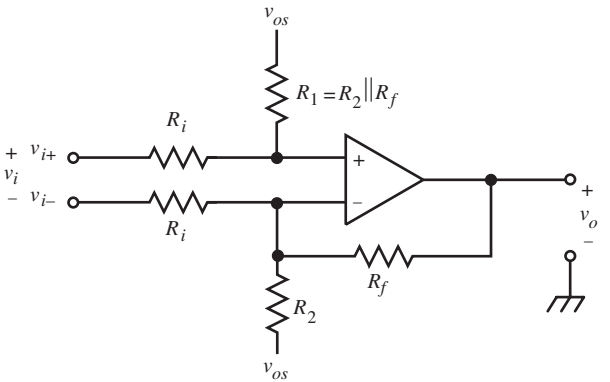
or

$$v_o = \left(\frac{R_2}{R_1 + R_2} \right) \cdot \left(\frac{R_4 + R_3}{R_3} \right) \cdot v_{i+} + \left(-\frac{R_4}{R_3} \right) \cdot v_{i-}$$

The noninverting gain path has a voltage-divider preceding the op-amp noninverting gain (in the first term). The second term represents the inverting path. For a differential amplifier, $A_{v+} = A_{v-}$. Equating gains and simplifying yields

$$\frac{R_2}{R_1} = \frac{R_4}{R_3} = A_v = \frac{v_o}{v_{i+} - v_{i-}}$$

With input resistors, this diff-amp can operate with common-mode input voltages far larger than the op-amp linear input dynamic range. Its main disadvantage is the finite input resistance of the noninverting input. In precision applications, the resistance of the ground return and input source resistance are common causes of gain error.



A more general form of the one-op-amp diff-amp is shown above, where R_2 has been added, and both R_1 and R_2 returned to an offset supply, v_{os}, to provide voltage offset of the input. Single-supply amplifiers with v_{os} can amplify bipolar input voltages with a unipolar output voltage range.

When the feedback circuit is solved, the output voltage is

$$v_o = \frac{K \cdot \left(\frac{R_1}{R_1 + R_i} \right)}{1 + K \cdot \left(\frac{R_1}{R_1 + R_i} \right) \cdot \left(\frac{R_i}{R_f} \right)} \cdot \left[v_i + \left(\frac{R_i}{R_f} \right) \cdot v_{os} \right]$$

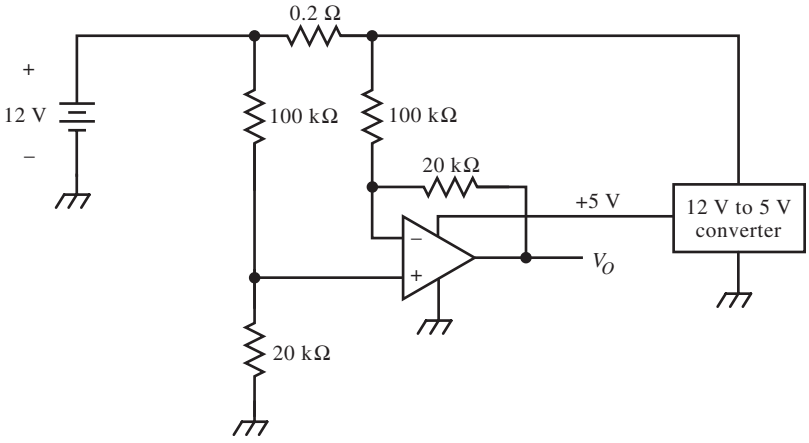
For infinite op-amp gain,

$$v_O|_{K \to \infty} = \frac{R_f}{R_i} \cdot v_i + v_{OS}$$

By increasing R_i, an increased input-voltage range can be achieved, but at the expense of increased input offset-voltage error. Multiplied by the op-amp input-voltage offset is the op-amp offset gain,

$$\frac{R_f}{R_i \| R_2} + 1$$

Example: Voltage Supply Current Sensing

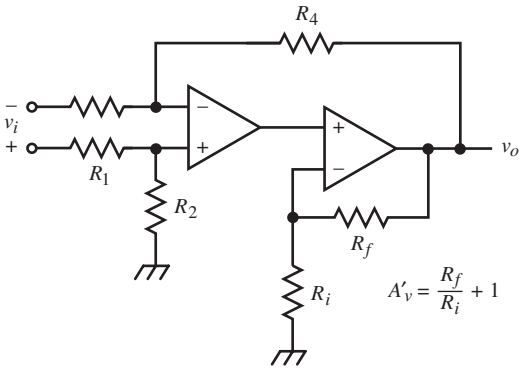


The input common-mode voltage range of the one-op-amp diff-amp allows it to be used to sense the current on the high side of a voltage supply. A single-supply op-amp operated from a 5 V converter is used to measure the current out of the 12 V battery that supplies the converter. The battery voltage must also be acquired by a digitizer requiring 2 V input for a battery voltage of 12 V.

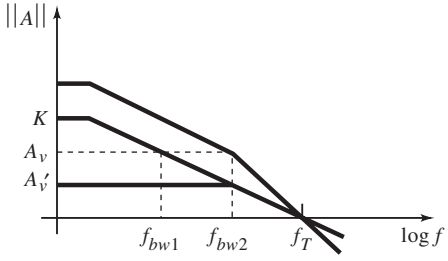
The 0.2 Ω sense resistor is insignificantly shunted by the two 100 kΩ input resistors to the op-amp. The battery voltage input to the digitizer requires a divider attenuation of 6. The noninverting input to the op-amp can be used to supply this voltage. This requires that the bottom side of the divider be 20 kΩ.

For differential current sensing, the op-amp feedback resistor must also be 20 kΩ, according to the differential constraints. Consequently, the scale of V_O to battery current is 40 mV/A and 2 V corresponds to 50 A.

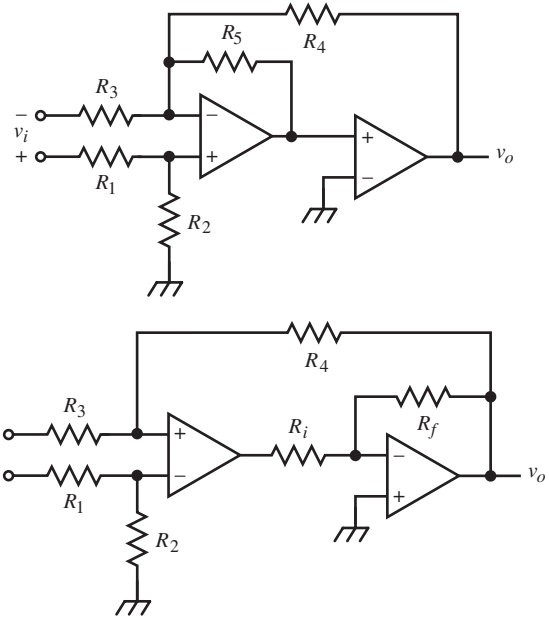
The trade-off for wide common-mode sensing range without the additional resistor on the noninverting terminal is reduced gain, and in this example, attenuation from the one-op-amp diff-amp.



The speed of the one-op-amp diff-amp can be improved by adding a second op-amp within its loop, with a closed-loop gain of A'_v. The second op-amp increases the loop gain. For op-amps with the same open-loop frequency characteristics, the Bode plots of each and their combination are shown below.

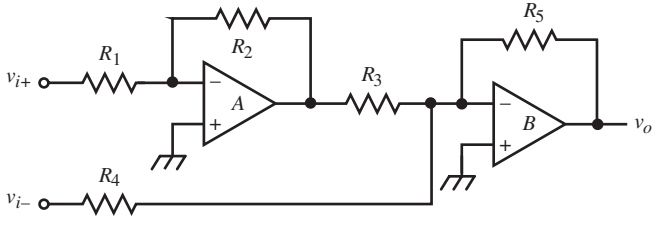


Although the combined gain-bandwidth, f_T, is not extended, more loop gain is available at higher frequencies with the additional amplifier. Variations on this theme are shown below.



An additional amplifier can also be placed in the feedback path.

A conceptually simple *two-op-amp diff-amp* is shown below.



Amplifier B is an inverting amplifier that sums inputs from v_{i-} and amplifier A, which inverts v_{i+}. The output voltage is

$$v_o = \left(\frac{R_5}{R_3}\right)\cdot\left(\frac{R_2}{R_1}\right)\cdot v_{i+} - \left(\frac{R_5}{R_4}\right)\cdot v_{i-}$$

The condition for differential inputs is

$$\frac{R_2}{R_3} = \frac{R_1}{R_4}$$

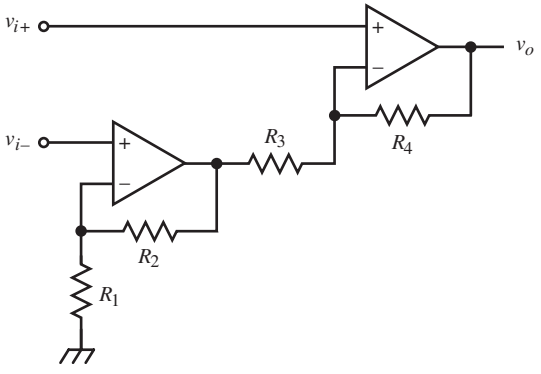
This condition is realized in two ways:

$$\begin{cases} R_2 = k\cdot R_3 \\ R_1 = k\cdot R_4 \end{cases} \quad or \quad \begin{cases} R_2 = k\cdot R_1 \\ R_3 = k\cdot R_4 \end{cases}$$

The differential gain, where $v_i = v_{i+} - v_{i-}$, is

$$\frac{v_o}{v_i} = \left(\frac{R_5}{R_4}\right)$$

This amplifier has no advantage over the basic one-op-amp diff-amp because it has comparable input impedance and involves two op-amp input error sources instead of one. But its noninverting form, shown below, has the advantage of high (ideally infinite) input impedance.



The v_{i-} input is amplified by an op-amp in the noninverting configuration and then is inverted by the output amplifier. The output voltage is

$$v_o = \left(\frac{R_4}{R_3} + 1\right) \cdot v_{i+} - \left(\frac{R_4}{R_3}\right) \cdot \left(\frac{R_2}{R_1} + 1\right) \cdot v_{i-}$$

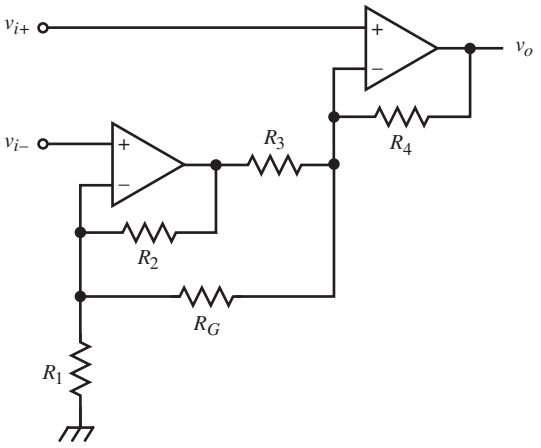
and the differential-input condition is

$$\frac{R_4}{R_3} = \frac{R_1}{R_2}$$

This results in a voltage gain of

$$\frac{v_o}{v_i} = \frac{R_4}{R_3} + 1 = \frac{R_1}{R_2} + 1$$

A useful variation on this diff-amp is to add R_G.



R_G complicates the analysis somewhat, but for ideal op-amps, it is across the (virtual) input voltage and affects only the differential gain. By straightforward analysis,

$$v_o = \left[\left(\frac{R_4}{R_3 \| R_G} + 1\right) + \left(\frac{R_4}{R_3} \cdot \frac{R_2}{R_G}\right)\right] \cdot v_{i+} - \left[\frac{R_4}{R_G} + \frac{R_4}{R_3} \cdot \left(\frac{R_2}{R_1 \| R_G} + 1\right)\right] \cdot v_{i-}$$

Solving for the differential input condition,

$$\frac{R_4}{R_3} = \frac{R_1}{R_2}$$

and the gain is

$$\frac{v_o}{v_i} = \frac{R_1 + R_4}{R_G} + \frac{R_4}{R_3} + 1$$

The gain formula derived above can be found in a simple way by superposition. When R_G is removed, the gain is that of the previous amplifier. Because v_i appears across R_G, the current through it (i_G) flows through R_2 and R_4, causing the additional gain of

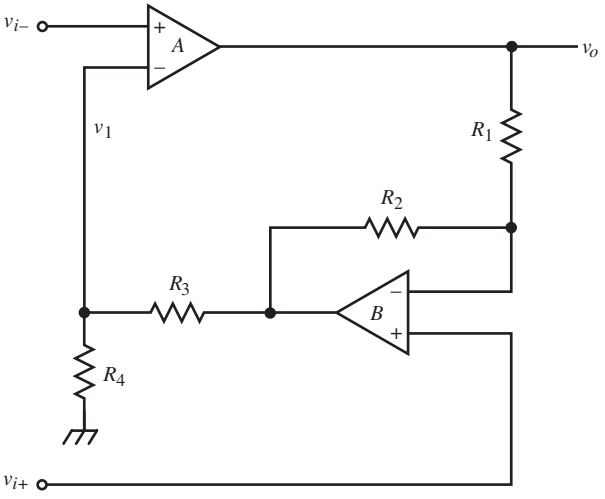
$$\left.\frac{v_o}{v_i}\right|_{R_G} = \underbrace{\frac{-R_2}{R_G}}_{\substack{i_G \text{ flows} \\ \text{through } R_2 \\ \text{out of } R_G}} \cdot \left(-\frac{R_4}{R_3}\right) + \underbrace{\frac{R_4}{R_G}}_{\substack{i_G \text{ flows} \\ \text{through } R_4 \\ \text{into } R_G}}$$

Then applying the differential condition to the first term, and by superposition of gains,

$$\frac{v_o}{v_i} = \left.\frac{v_o}{v_i}\right|_{R_G \to \infty} + \left.\frac{v_o}{v_i}\right|_{R_G}$$

which is equal to the formerly derived gain.

A *programmable-gain amplifier* (PGA) is an amplifier for which the gain can be set parametrically to given values. A differential PGA can be realized with this topology with a simple gain-setting strategy. Set the minimum gain and then program the additional gain with R_G. With R_G open, R_3 and R_4 set the minimum gain. With finite R_G and R_4 already determined, a fixed (nonprogrammable) R_1 allows R_G to change the gain in an additive manner.



The two-op-amp diff-amp shown above has the second op-amp in the feedback loop, with v_{i+} inserted there. This odd topology can be analyzed by finding v_1. It is the voltage-divider attenuation times the output of op-amp *B*, or

$$v_1 = \left(\frac{R_4}{R_3 + R_4}\right) \cdot \left[\left(\frac{R_2}{R_1} + 1\right) \cdot v_{i+} - \left(\frac{R_2}{R_1}\right) \cdot v_o\right] = v_{i-}$$

With op-amp *A* input error nulled, $v_1 = v_{i-}$ Solving for v_o,

$$v_o = \left(\frac{R_1}{R_2}\right) \cdot \left[\left(\frac{R_2}{R_1} + 1\right) \cdot v_{i+} - \left(\frac{R_3}{R_4} + 1\right) \cdot v_{i-}\right]$$

The differential-input condition is

$$\frac{R_2}{R_1} = \frac{R_3}{R_4}$$

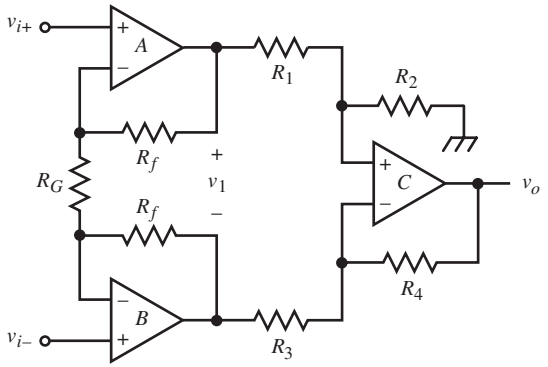
The differential gain is

$$\frac{v_o}{v_i} = \frac{R_1}{R_2} + 1$$

For these diff-amps, v_{i+} and v_{i-} are input to different op-amps. As a result, common-mode rejection (CMR) does not tend to be as good as for the one-op-amp diff-amp. This is especially true as a function of frequency because the signal paths from the two inputs are asymmetric; one input goes through two amplifiers and the other through only one.

INSTRUMENTATION AMPLIFIERS

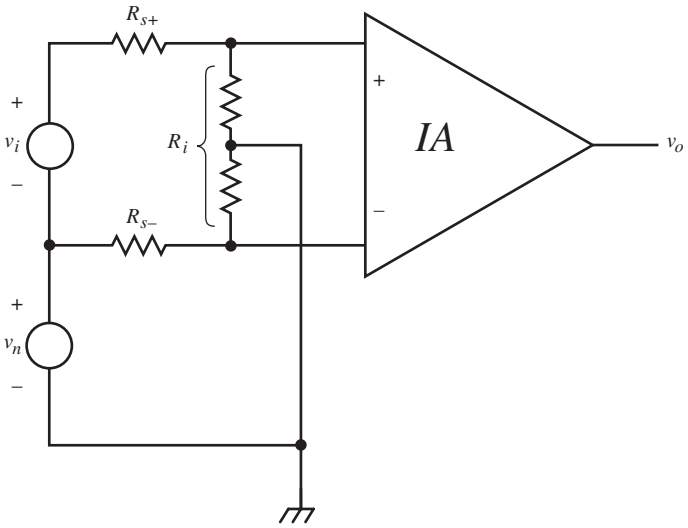
An amplifier with accurate, programmable gain, high input resistance, and high CMR is an an *instrumentation amplifier* (IA). The gain can be set to a given value, or programmed, by setting one resistor R_G. The most common instance is the *three-op-amp diff-amp.*



It has two stages. The input stage is a differential amplifier (at both input and output), and the second stage is the one-op-amp diff-amp. The differential input voltage v_i appears across R_G, creating a current flow of v_i/R_G in both feedback resistors R_f. The gain of the first stage is thus $2R_f/R_G + 1$. Because the amplifier is symmetric, the center-point of R_G is a virtual "ground" or null point, and each side behaves as though it were connected through $R_G/2$ to the center node of a split, symmetric input voltage source. For the second stage to be differential, the differential constraint applies, and the total amplifier gain is

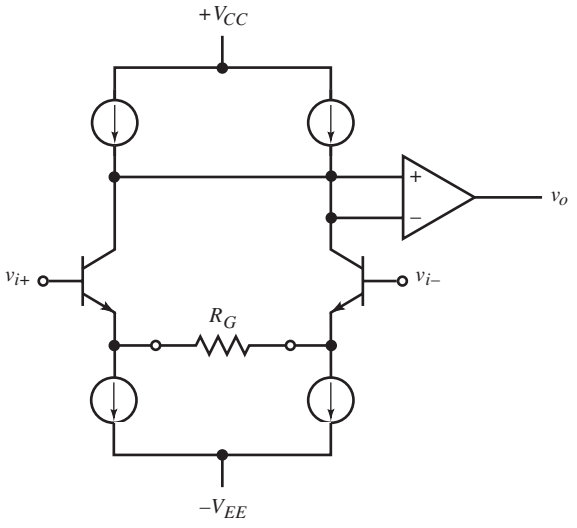
$$\frac{v_o}{v_i} = \left(\frac{R_2}{R_1}\right) \cdot \left(2\frac{R_f}{R_G} + 1\right) = \left(\frac{R_4}{R_3}\right) \cdot \left(2\frac{R_f}{R_G} + 1\right)$$

The gain of this IA is easily programmed by setting R_G, and it is amenable for use as a PGA by switching in different values of R_G.



The above topology is commonly used as an input amplifier for remotely located transducers. The above diagram shows a typical application, where the transducer voltage v_i is connected to the IA inputs via two conductors, with resistances R_{s+} and R_{s-} of generally different values. Common-mode noise v_n appears at the input, attenuated by different amounts due to R_i. A diff-amp with ideal CMR cannot reject differing amounts of noise at the two inputs. An IA has an ideally open input so that R_i is infinite and v_n is therefore unattenuated from the two conductors. In effect, v_n is retained as purely common mode.

CMR is a major parameter of the IA. Consequently, topologies with improved differential inputs have been developed. The use of two op-amps in the differential input stage involves two sets of differential inputs that must be matched for high CMR.

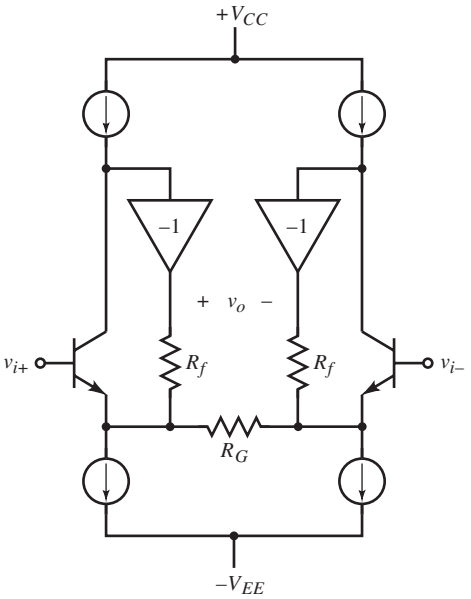


A simpler realization of IA inputs uses a single differential transistor stage, shown above. The differential input voltage is buffered by the BJT diff-amp and appears across R_G.

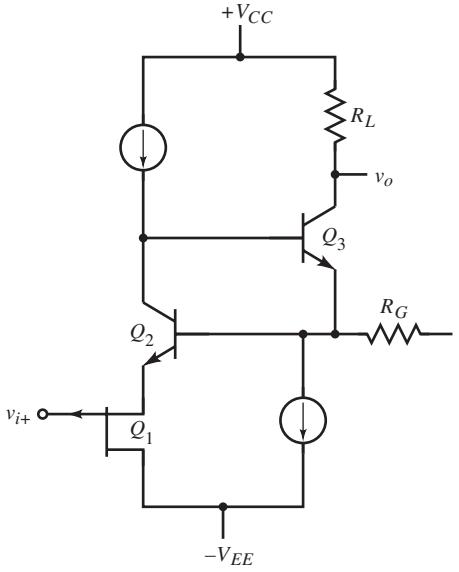
The linearity of this input stage is limited by the variation in dynamic emitter resistances with signal current. Furthermore, as gain is changed by changing R_G, the bandwidth changes since the amplifier has a fixed gain-bandwidth product. Both of these errors can be minimized by use of feedback.

An improved realization of IA inputs uses a differential transistor feedback stage, shown on the next page. The BJT collector voltages are buffered and fed back to the input BJT emitters through feedback resistors. These feedback buffers are more linear than emitter followers and more linearly transfer v_i across R_G. The output v_o is applied to the input of the second stage.

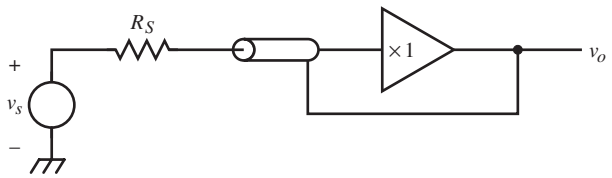
Both stage and loop gains are affected by R_G. The feedback loop gain varies inversely with R_G since it is part of the feedback path attenuator and is within the loop. The stage gain v_o/v_i is $2R_f/R_G$ and also varies inversely with R_G. These gains track with changes in R_G. When R_G is decreased, both open-loop and closed-loop gains increase the same amount. Consequently, the closed-loop dynamic response remains unchanged with gain changes.



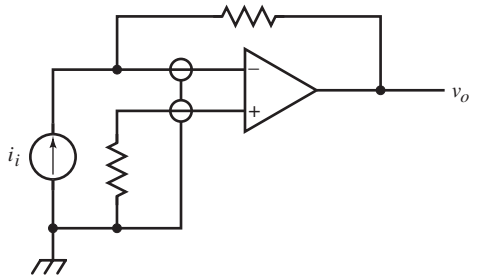
In some IAs, the input transistors are FETs. The r_m variations of FETs is greater than r_e variations of BJTs, leading to less linearity and greater advantage for a feedback input stage. A clever variation on input-stage buffering is found in the ADI AMP-05. One side of the input stage is shown below.



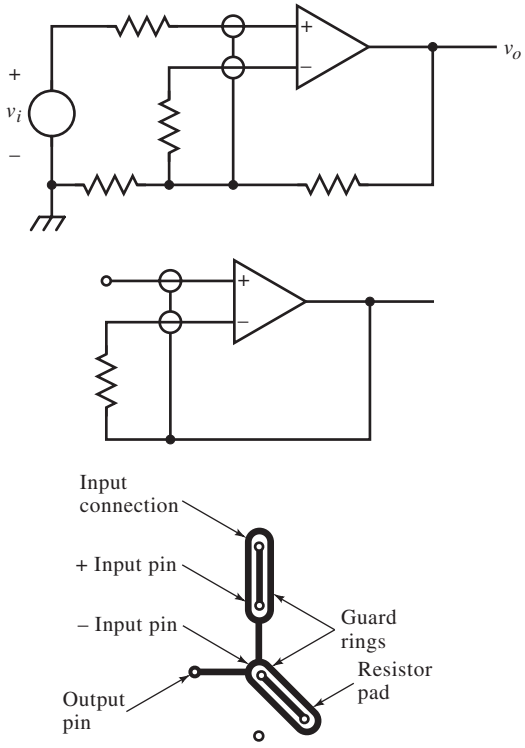
Input FET Q_1 is in series with BJT Q_2, forming a complementary differential pair. The output of Q_2 is buffered by Q_3 with no feedback resistance, fixing the closed-loop gain at one. The stage output is taken from the load resistor of Q_3.



Guarding is a technique for improving dynamic CMR by bootstrapping the input cable shield of remote sources. It is used in the buffer amplifier shown above. The shield behaves as a guard by not allowing input signal voltage variation across the cable capacitance. The situation can be generalized from the previous IA common-mode circuit drawing by letting R_i be differential input capacitance C_i. Any capacitive difference between the two sides results in asymmetrical *RC* filter circuits and different dynamic responses. Different responses cause degradation in CMR with frequency. Minimizing C_i by guarding improves CMR.

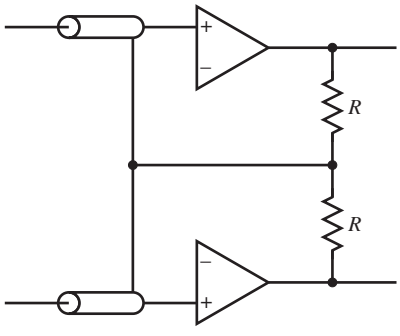


Guarding also bootstraps cable shunt resistance for high input-resistance (or low signal-current) applications, in which leakage currents are critical. Circuit-board surfaces provide stray leakage paths for current. The input nodes to the op-amp circuit above are guarded by enclosing the op-amp IC input pins, as shown below for a noninverting ×1 buffer.

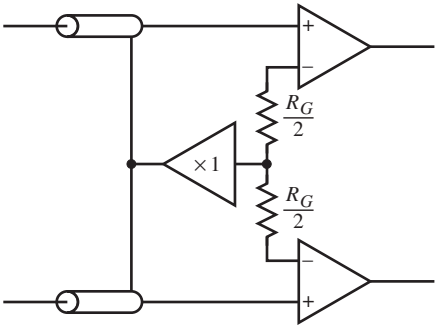


The board surface is bootstrapped at the same voltage as the input pins by the output, which easily absorbs leakage into the guard rings.

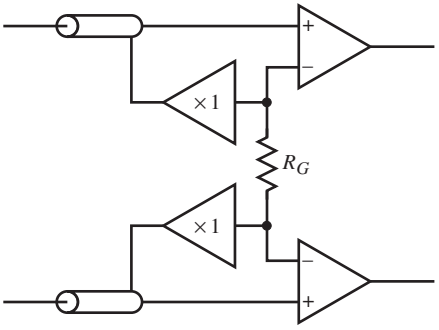
Leakage through the board, or bulk leakage, is reduced by about ten times by placing guard rings on both sides of the board. For critical applications, the through-hole connection on the board is replaced by a polyethylene standoff with feedthrough so that no conductor contact of critical nodes is made with the board.



Guarding is commonly used with instrumentation amplifiers, as shown above. The first-stage output common-mode voltage is supplied by a balanced divider.

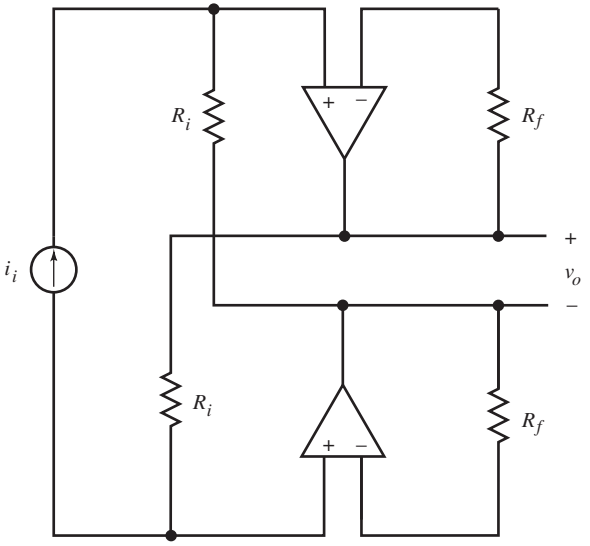


A split gain resistor also has the common-mode voltage at its center-tap, as shown above, where it is buffered. These two guard circuits reduce cable capacitance and leakage due to common-mode voltages.



In the most precise guard circuit, shown above, differential guarding bootstraps both common and differential-mode voltage variations.

The high input impedance of the IA applies well to precision voltage amplification. The IA input stage can also be made into a transresistance amplifier for current amplification, as in the circuit shown below: a differential, inverting op-amp topology. The transresistance is $2R_i$. The advantage of this circuit over



that of an IA with a resistor across its input is that it has no input voltage drop ($R_G = 0$) and is well suited for current sources with limited compliance.

Low-Level Amplification and Component Characteristics

Very small voltage or current waveforms are *low-level* waveforms, small enough so that circuit-board and cable leakage, thermal gradients, and thermocouple effects are significant. At these levels, random noise and EMI considerations are important. Most low-level waveform-processing circuits are limited in speed as a trade-off for low-frequency precision.

High precision requires that circuits use components with low temperature coefficients (TCs) to minimize drift of the bias or quiescent operating point. For example, a change in op-amp bias current can cause a change in offset voltage and output voltage error. This error can be nulled by adjusting the offset, but to remain nulled, the bias current must not change.

High precision also requires minimization of low-level static thermal noise. Two sources of this kind of noise are *thermal gradients* and *thermoelectric effects*. A thermal gradient across the two sides of a differential amplifier causes junction temperatures to be asymmetric, resulting in offset and gain error.

These gradients arise from convection currents of air across the circuit board or thermal conduction along an IC substrate. They are minimized for a discrete circuit by thermally shorting balanced components with a thermally conductive path between them. By mounting two transistors on the same metal heat sink, their static temperatures track because of the high thermal conductivity of metal. The metal behaves as the thermal equivalent of a single electrical node since it is the same temperature everywhere (or *isothermal*). The thermal-electrical analogy is

$$\text{temperature, } T \Leftrightarrow \text{voltage}$$

$$\text{thermal power, } P_\theta \Leftrightarrow \text{current}$$

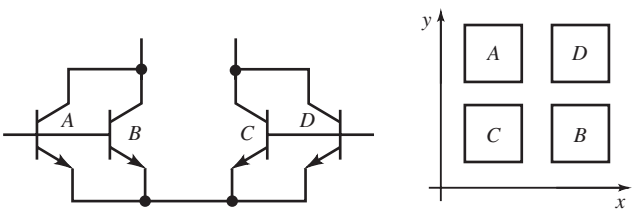
$$\text{thermal resistance, } R_\theta \Leftrightarrow \text{resistance}$$

The thermal analog of Ohm's law is then

$$T = P_\theta \cdot R_\theta$$

Approximate analogs also exist between "thermal mass" (mass × specific heat) and capacitance, leading to dynamic thermal effects, to be studied later. A heat sink has $R_\theta \cong 0$, thus minimizing ΔT between differential transistor pairs. In some cases, a box is built around sensitive circuitry to function as a thermal convection shield.

For ICs, these techniques cannot be applied. Instead, balanced components are placed across symmetric thermal gradients. This requires thermal as well as electrical consideration of the IC layout. Alternatively, circuit elements are constructed so that balanced pairs receive the same thermal stimulus. For example, a two-transistor differential pair can be constructed of two pairs of transistors in a square pattern, with parallel devices situated diagonally.



For isotherms along the x axis, A, C and B, D are at the same temperature; isotherms along the y axis heat C, B and A, D the same. In each case, thermal symmetry is preserved, and the electrical effects are canceled.

The effect of diagonal isotherms is minimized if the gradient is linear. A gradient with isotherms of slope −1 heats A and B to the same temperature, whereas each of C and D is hotter and colder than A, B. If the average temperature of C and D is the temperature of A, B, thermal effects are canceled to the extent that the thermal-to-electrical transfer function is linear. The PN junction relation shows that it is not for saturation current, $I_S(T)$. Therefore, close layout is desired to minimize temperature differences among transistors.

Thermoelectric voltage generation occurs when dissimilar metals are joined, as in thermocouples. Thermocouple voltages are generated at solder joints and connectors. IC leads are often made of the alloy kovar. A copper-kovar joint generates 40 μV/°C, whereas a copper-solder joint (with tin-lead solder) generates 1 to 3 μV/°C. A type K (chromel-alumel) thermocouple has a room-temperature TC of 39 μV/°C, for comparison. These thermocouple joints are compensated by symmetry, by having the same number of them on each side of a differential input, and by having them track with temperature.

Resistors affect performance according to their type, as the following table shows.

Resistors	Carbon Film	Metal Film	Bobbin Wirewound
accuracy, %	0.5	0.1	0.01
stability, %/kh	1	0.5	0.5
TC, ppm/°C	500	25–150	20
R range	1 Ω – 150 MΩ	10 Ω – 10 MΩ	0.05 Ω – 6 MΩ

Resistor stability is the fractional amount the resistance changes over time, usually in units of 1,000 hours. Wirewound resistors are constructed in various ways. Winding resistance wire on a bobbin results in the highest performance, as given in the table.

Variable resistors (potentiometers and rheostats) are also affected significantly by temperature and time.

Pots	Carbon	Cermet	Wirewound
Stability, %/kh	±10	±3	±2
TC, ppm/°C	400–800	100	50
R range	100 Ω – 5 MΩ	10 Ω – 2 MΩ	10 Ω – 50 kΩ

Contact resistance variation (CRV) of the wiper as it is moved along the resistive element is the main cause of adjustment noise in a variable resistor. CRV is typically 1% maximum, but the application determines the extent of the effect. As a potentiometer with no wiper current, CRV does not affect circuit operation. For rheostats, in which all the current flows in the wiper, the effect is maximum. Generally, wirewound pots have the lowest CRV, followed by conductive plastic, molded carbon, and cermet.

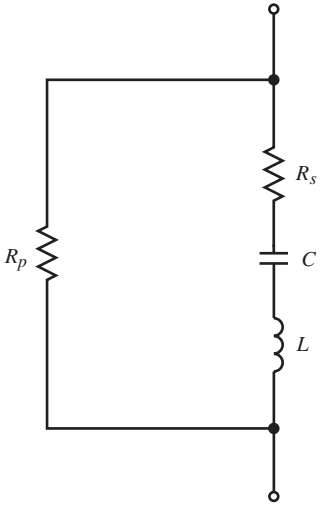
Resistors have parasitic capacitance and inductance. The shunt terminal capacitance of a 1/4 W carbon resistor is about 0.5 pF. Series inductance is more significant, especially in wirewound resistors. Noninductive winds of resistance wire (usually manganin) can reduce inductance. This is significant when power resistors are used to sense current in magnetic deflection amplifiers, current-mode switching power supplies, and motor drives.

Capacitors also affect circuit precision. Selection of the optimum capacitor type for a given application is a small specialty in itself, greatly aided by familiarity with manufacturers' specifications. A concise chart giving typical capacitor characteristics is appealing, but variations among values on existing charts suggest a more general approach. Some general facts about capacitors are helpful in design, with emphasis on plastic film capacitors.

Capacitor plates are either metal foil or a metallized deposition on the dielectric film itself. Foil capacitors are larger because the foil is thicker but have lower series resistance. They also cost less but can have higher TCs.

The figure on the next page shows a general model of a capacitor. Capacitance C has series inductance L and series resistance R_s, called *equivalent series resistance* (ESR) on data sheets. R_s is a limiting parameter in power applications, involving large ripple currents, and is characterized directly; a 1 mF, 25 V aluminum electrolytic has a typical series resistance of 50 mΩ.

In small-signal applications, R_s is characterized by its *dissipation factor* (DF) the ratio of R_s/X_c, where X_c is capacitive reactance. Alternative characterizations



abound. The arctangent of the ratio is the *loss angle*, and its reciprocal is the *quality factor*. The *power factor* is the sine of the loss angle. Dissipation factor is a function of both temperature and frequency.

Current leakage, both through the dielectric and across the body of the capacitor, is modeled by the parallel resistance R_p. The dielectric resistance is expressed by the geometric resistance formula,

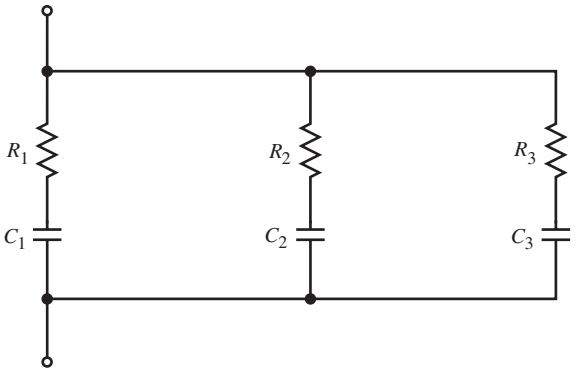
$$R = \rho \cdot l / A$$

where ρ is resistivity, l is the dielectric thickness, and A the plate area. The geometric capacitance is $C = \varepsilon \cdot l/A$. R_p is proportional to C and is usually given in units of ohm-farads (or megohm-microfarads). Insulation resistance decreases with temperature for all film capacitors. Foil capacitors have about five times higher insulation resistance than metallized types. For low-leakage applications such as slow ramp generators or long-interval timers, the outer surface of the capacitor must be clean. A capacitor with fingerprints shows a noticeable increase of leakage, by orders of magnitude. Moisture also degrades insulation and creates current paths.

For bypassing, a minimum L is desired. Electrolytic capacitors have the largest L, with series-resonant frequencies typically between 100 kHz and 5 MHz. Other

capacitors are usually not limited in their frequency by L but by dissipation factor. The best general high-frequency capacitors are ceramic.

Dielectric materials tend to become polarized by an electric field that remains after the field is removed. If a capacitor is shorted for a while so that its terminal voltage is zero and then the short is removed, the capacitor exhibits a nonzero voltage! This phenomenon is due to the partial polarization of the dielectric material into an *electret*, the electric equivalent of a permanent magnet. It can be modeled by several series RC elements in parallel.



This characteristic is quantified as *dielectric absorption*, measured as the fraction of the applied voltage that the capacitor exhibits after being shorted for a fixed time, usually 5 seconds. This effect can be exploited to make transducers but causes anomalies in circuit response. For high-performance integrators or sample-and-hold circuits, dielectric absorption must be minimized.

Plastic-film capacitors are made by winding together alternate layers of foil and dielectric film, or layers of metallized film, with opposing plates offset to opposite sides of the roll. In "noninductive" capacitors, the extra foil coming out of each end is smashed against the end and soldered to a lead. The plates are electrically paralleled; this reduces both L and R_s. In an "inductive" construction, the foils are brought together at the ends of the wrap. One lead is placed in the center and the other at the perimeter. Both leads run the width of the capacitor. The length of the foil or film wind is usually much greater than its width, and L and R_s are greater. But this method of construction is simpler and is the only way to build metallized film capacitors.

The plastic films used as dielectrics in capacitors number about a half dozen. The common name for *polyester* is Mylar. For general use, where high performance is not required, polyester is preferred because it is low cost and has good volumetric efficiency (high charge density). Capacitor size depends mainly on dielectric constant and dielectric strength, the maximum electric field before breakdown occurs. Polyester has a dielectric constant of 3.2. Its TC is large at temperature extremes. It monotonically increases from –5%/°C at –55 °C, with an inflection point at 25 °C where the TC levels off. At its maximum temperature of 125 °C, the TC is a huge +14%/°C. But from 0° to 50 °C, the TC is about 1%/°C, and then only at the extremes. The typical DF of polyester is 0.5%, dielectric absorption is 0.2%, and foil insulation resistance is the worst of the films, at a resistance-capacitance product of 10^5 ΩF at 25 °C.

Polycarbonate capacitors, now no longer commercially produced, are slightly larger than polyester capacitors of the same capacitance and voltage rating, with a similar shape of TC versus temperature, increasing with temperature and flat with zero TC at 25 °C. The TC extremes are –1.5%/°C at –55 °C and +1.5%/°C at 125 °C. Dissipation factor is about half that of polyester, and foil insulation resistance is twice as high. Polyester and polycarbonate insulation resistance decreases superlinearly with temperature; the better capacitors decrease linearly. Dielectric absorption is typically 0.08%. Both decrease superlinearly with temperature. It is a medium-grade capacitor and a likely choice when polyester is not quite good enough. Its biggest weakness is its moisture sensitivity; it is the worst of the film capacitors. Although they are not abundant and manufacturing of them has also ceased, *polysulfone* capacitors are similar to polycarbonates but have less moisture sensitivity and DF, have a flatter TC, and can operate up to 150 °C.

The most commonly used high-performance film capacitors are *polypropylene* and *polystyrene*. Polypropylenes are comparable in price to polyesters, are only slightly larger, and have a maximum temperature of 105 °C. Polystyrene is the better of the two, electrically, but costs more, is three times the size of polyester, and operates to only 85 °C. For polypropylene, typical values are DF $\cong$ 0.02% over temperature, foil insulation resistance $\cong 8 \times 10^5$ ΩF at 25 °C, dielectric absorption = 0.02% with negligible moisture sensitivity. The TC is a linear –250 ppm/°C. Polystyrene is similar but has half the DF and maintains insulation resistance at higher temperatures much better. Its TC is also

linear but varies with material. The standard TC is −120 ppm/°C ± 50 ppm/°C. Polypropylene capacitors are a good choice for high-frequency power applications due to their low DF.

The common name for *polytetrafluoroethylene* (PTFE) is Teflon. It is the highest-performance dielectric and highest in cost. PTFE is comparable to polystyrene in most properties and is twice the size of polyester. It has twice the foil insulation resistance of polystyrene, the best available. It has a linear negative TC of −200 ppm/°C and is, along with polystyrene, the most stable in capacitance over time, at 0.1%/year. The worst is metallized polyester, at 0.5%/year, with the others around 0.2%/year. DF is flat with temperature. It is the highest in operating temperature: 200 °C.

One approach to zero TC is to make a hybrid dielectric of materials with opposite TCs that cancel. One such hybrid capacitor uses films of polyester and polypropylene with a TC of zero ±100 ppm/°C. It retains some of the worst properties of polyester, however: DF = 0.5% and dielectric absorption = 0.15%. Two newer dielectrics, *polyphenylene sulfide* (PPS) and *polyvinylidene fluoride* (PVDF), may be used to make the most stable capacitors. Polyphenylene sulfide has a DF comparable to that of polypropylene.

In summary, and with some simplification, there are three performance classes of film capacitors: polyester is low performance, polycarbonate and polysulfone are medium performance, and polypropylene, PPS, polystyrene, and PTFE (in order of improvement) are high performance. Except for frequency characteristics, they are better than ceramic capacitors except for "zero" TC NP0 ceramics. Mica capacitors give high performance at a high price and accuracy, with a range of TC down to ±70 ppm/°C. The ultimate capacitor, except for size and cost, has a vacuum dielectric. Glass and air approach it in stability and low TC but have extremely low charge density.

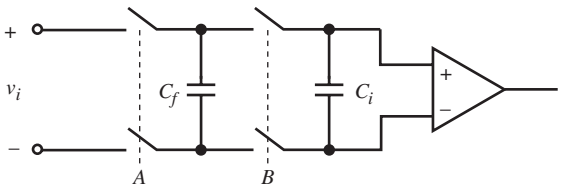
Another important kind of capacitor is that made from glass epoxy circuit-board material. These capacitors are sometimes intentional but usually are parasitic. The most common board materials, G-10 and fire-retardant FR-4, have a dielectric constant of about 4.8 and a DF (at 1 MHz) of 0.02%, comparable to that of polypropylene. The volume resistivity of G-10 is 5×10^8 MΩ-cm, five times that of FR-4. The surface resistivity of G-10 is 4×10^8 MΩ, whereas for FR-4 it is an order of magnitude less. Both materials can be used up to 130 °C. Low-cost, low-performance phenolic boards have a dielectric constant of 4.1, a 1 MHz DF

of 0.03%, volume resistivity of 5×10^6 MΩ-cm, surface resistivity of 5×10^4 MΩ, and a maximum temperature of 125 °C.

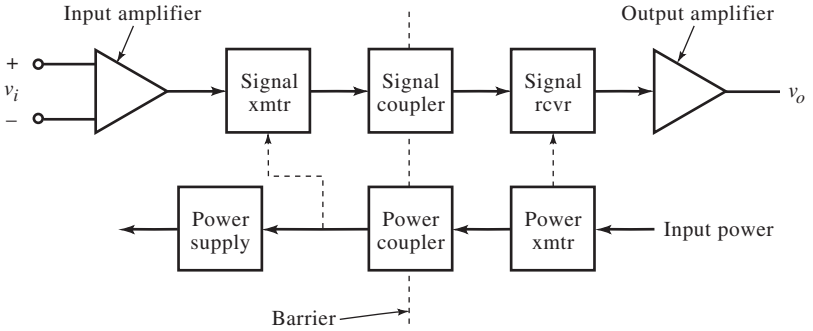
Boards that have not been cleaned after assembly contain solder flux that can be a cause of excessive leakage. When even clean board leakage is excessive and guarding techniques inadequate, critical high-resistance nodes can be constructed with Teflon standoffs or mechanically stable parts connected in midair.

Isolation Amplifiers

Some applications require extreme isolation of amplifier input from power supply and output. These include patient monitoring, floating measurements involving high-voltage circuits, digital voltmeters, and circuits in environments with severe ground loops.



A long-used isolation technique is *flying capacitor* isolation, shown above. Switches activated by *A* are closed, charging capacitor C_f to the input voltage. The *A* switches are then opened, and the *B* switches are closed, charging the amplifier stray input capacitance C_i. For negligible loss of voltage, $C_f >> C_i$. By using reed relays for the switches, large maximum voltage ratings and low switch resistance can be achieved. Limitations are finite switch-cycle life, contact bounce, contact noise, relatively large size and power requirements, and slow switching speed. A solid-state version overcomes these problems but suffers from lower voltage ratings; the LTC1043 "switched-capacitor building block" contains two flying-capacitor circuits and an oscillator to drive them, realized in monolithic silicon.



A more general approach is shown above. An input amplifier drives a kind of signal transmitter that drives a signal coupler. This coupler has no static (dc) conductive (or *galvanic*) path between input and output sides and preserves an isolation barrier. Energy transmission occurs by some means other than electrical conduction, such as optical coupling using optoisolators. Magnetic coupling is common using a transformer. Since static amplification is desired, the signal transmitter in this case is a square-wave amplitude modulator and the receiver on the output side, a demodulator and filter. The form of transmission is *pulse amplitude modulation* (PAM).

Because the circuits on the input side require power, a separate isolated power coupler is required with a power oscillator driving it and a power supply on the input side. If the power port is also isolated from both input and output sides and supplies isolated power to both, then the system has three isolated ports, the most general case of isolation. A two-port isolation amplifier, with power and output ports not isolated, is often adequate.

The dotted paths in the diagram above indicate that the switched power waveform can be supplied to the modulator and demodulator for synchronous demodulation and a doubling of the signal-to-noise ratio.

Other forms of modulation can be used. Instead of PAM, voltage can be encoded in frequency with a voltage-to-frequency converter (VFC). Pulses coupled across the barrier are converted back to a voltage by a FVC, or their frequency is measured digitally. With pulse-width modulation (PWM), the voltage is encoded as a duty-ratio so that the coupled pulse frequency is not related to voltage accuracy. For any modulation technique, the signal bandwidth is limited by the carrier (or modulating) frequency. Optical coupling is static,

but light-emitting diodes (LEDs) and phototransistors are nonlinear, and their scaling is difficult to control. Consequently, matched optical paths are used. One is the input amplifier feedback loop, and the other is on the output side. Other statically-responding devices, such as Hall-effect devices, have similar difficulties and have not become commonplace for such coupling.

A common industrial instrumentation interface is the *20 mA current loop*. Signals are encoded as analog currents with 4 mA as zero scale (zs) and 20 mA as full scale (fs). Several channels can be remotely connected to the inputs of isolation amplifiers using a cable containing a twisted pair of wires for each channel. At the input end of the cable, transducers with 20 mA current-loop interfaces send back sensor signals with little voltage variation, causing minimal capacitive coupling between twisted pairs in a cable. The transducers are required to operate on a maximum of 4 mA. Then the two-wire interface supplies power as a voltage, and power-supply current is measured as the transducer output signal.

When transformers are used to isolate circuits – not only in isolation amplifiers but in systems with EMI in general – parasitic capacitance between primary and secondary windings provides an electrical path for noise or fluctuating leakage current and degrades isolation. Interwinding capacitance is minimized by placing the windings on opposite sides of a toroidal core or separating them on opposite ends of an E core, using a split bobbin. Capacitive imbalances arise from winding distribution, in which interwinding capacitance on one end of a winding is larger than on the other end. These differential imbalances cause differential outputs from common-mode inputs. Interwinding capacitance can be minimized by placing a nonshorting turn of insulated copper sheet between concentric transformer windings as an electrical or *Faraday shield*. This shield is connected to the side generating the shield current, providing a return to the generating source.

AUTOCALIBRATION

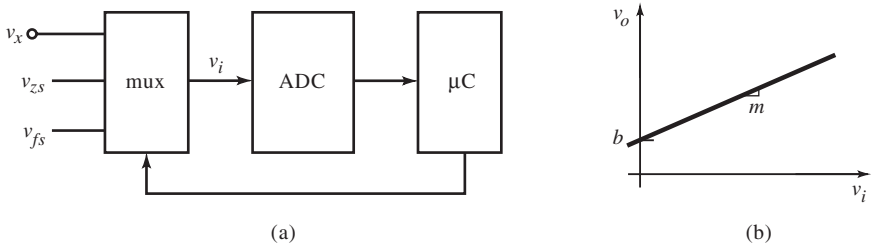
Precise signal processing is achieved by making the components precise. But component improvement is limited and compensation methods are required to improve performance. Several techniques are based on a general idea, called *autocalibration*. Instead of trying to null circuit errors with more circuitry, circuit

behavior is characterized using known inputs. The deviation of the resulting outputs from the ideal are measured and can be used to correct the unknown signal output.

A general autocalibration scheme is shown in (a) below for an analog-to-digital converter (ADC). An analog multiplexer switches zs and fs reference voltages into the ADC, and the output is sent to a microcomputer (μC) that is controlling the multiplexer sequencing. If the ADC is linear, a linear model is assumed for its transfer function, shown in (b). With two measurements based on known inputs, the slope and offset are computed. Then measurements of V_x are corrected according to these parameters. Autocalibration thus requires two modes, calibration and measurement.

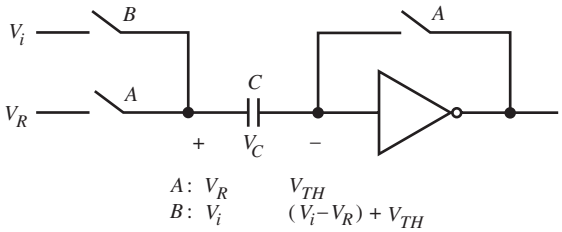
A general autocalibration equation is based on two arbitrary reference voltages V_{R1} and V_{R2} and their measured values V_{R1m} and V_{R2m}. Then the unknown input V_X is calculated from measurement values V_{Xm}:

$$V_X = \left(\frac{V_{Xm} - V_{R1m}}{V_{R2m} - V_{R1m}} \right) \cdot (V_{R2} - V_{R1}) + V_{R1}$$



Integrating ADCs often have an *autozero* feature that corrects for offset error as follows. Without a microcomputer, a clock oscillator drives switches that first ground (or null) the ADC input and store the error on a capacitor. This offset error is then switched in during measurement so that it subtracts from the input signal.

Autozeroing is used in the autobalanced CMOS comparator. A CMOS inverter is connected in series with capacitor C and switches. When the A switches are closed on the first half-cycle of a clock, the inverter input and output are forced



to be equal at threshold, V_{TH}. The input side of C is at the comparator reference voltage V_R. On the second half-cycle of the clock, the A switches are opened, and B is closed, connecting the input voltage. The inverter input voltage is then

$$V_{in} = V_i + V_C = V_i - (V_R - V_{TH}) = V_i - V_R + V_{TH}$$

Because the inverter input is offset by V_{TH}, the difference between input and reference is also offset by V_{TH}, nulling the offset. The inverter output consequently responds to $V_i - V_R$ and performs the comparator function. A large number of these comparators have been used to implement parallel ADCs and other mixed analog and digital systems with digital IC processes. A similar input scheme is used in the LTC1040 dual micropower comparator.

Amplifier input offset voltage error can be reduced with autozeroing by nulling the amplifier inputs and charging a capacitor to the input error voltage. Then the waveform is switched in, and the offset error is nulled. A fault in this approach is that by switching between ground and waveform, the switching waveform becomes part of the output. Also, the input waveform is interrupted. Such amplifiers are called *chopping* or *chopper-stabilized* amplifiers. Multiple-path amplifier topologies have been devised to deal with these problems.

The chopping or switching of the input waveform makes autocalibration techniques discontinuous, or discrete, in time. Systems with continuous values and discrete time functions are called *sampled-data systems* and exhibit a range of behavior investigated in *Designing Waveform-Processing Circuits*. The bandwidth of these systems is limited by the chopping or *sampling rate*; the input waveform bandwidth must be much less than the sampling frequency. In practice, autocalibrating or switched-capacitor circuits have a limited bandwidth (10 Hz to

100 kHz) but excellent input offset voltages (1–10 μV). They are well suited as thermocouple and vacuum-system ion gage input circuits.

DISTORTION

Besides noise, another major cause of degradation in precision is distortion. Its two main causes are device nonlinearity and thermal effects.

Active devices are inherently nonlinear. Nonlinearity can be minimized by setting the operating point in the most linear region. Circuit techniques that minimize nonlinear behavior can be used, such as the minimization of BJT base-width modulation for the common-emitter (CE) in a cascode stage. These attempts at linearization are often limited by other constraints. We now examine distortion due to nonlinearity and develop a simple way of estimating the amount of it.

Bruce Hofer long ago developed a technique for estimating *harmonic distortion* in amplifiers. He began with the truncated series expansion of an amplifier output:

$$v_O = \sum_{k=0}^{3} a_k v_I^k = V_{os} + a_1 v_I + a_2 v_I^2 + a_3 v_I^3$$

The technique is based on measurement of the amplifier gain at the center and positive and negative peaks of an input sinusoid. The incremental gain is

$$A = \frac{dv_O}{dv_I} = a_1 + 2a_2 v_I + 3a_3 v_I^2$$

Solve for a_k at $v_I = 0$, V, and $-V$. Substituting the three values,

$$A(0) = a_1$$

$$A(V) = a_1 + 2a_2 V + 3a_3 V^2$$

$$A(-V) = a_1 - 2a_2 V + 3a_3 V^2$$

These equations are solved for a_k by subtracting the third from the second and by adding the second and third. The solutions are

$$a_2 = \frac{A(V) - A(-V)}{4V}$$

$$a_2 = \frac{A(V) + A(-V) - 2A(0)}{6V^2}$$

For a sinusoidal input of

$$v_I = V \cdot \sin(\omega \cdot t)$$

the output is

$$v_O = V_{os} + a_1 V \sin \omega t + a_2 V^2 \sin^2 \omega t + a_3 V^3 \sin^3 \omega t$$

Applying trigonometric identities,

$$\sin^2 \omega t = \frac{1}{2} - \frac{1}{2} \cdot \cos 2\omega t, \quad \sin^3 \omega t = \frac{3}{4} \cdot \sin \omega t - \frac{1}{4} \cdot \sin 3\omega t$$

to v_O,

$$v_O = \underbrace{\left(v_{os} + \frac{1}{2} a_2 V^2 \right)}_{\text{static term}} + \left(a_1 + \frac{3}{4} a_3 V \right) V \sin \omega t - \frac{1}{2} a_2 V^2 \cos 2\omega t - \frac{1}{4} a_3 V^3 \sin 3\omega t$$

The distortion due to the nth harmonic is defined as

$$nHD \equiv \frac{\|X_n\|}{\|X_1\|} = \frac{\text{amplitude of } n\text{th harmonic}}{\text{amplitude of fundamental}}$$

The second-harmonic distortion is

$$2HD = \left| \frac{\frac{1}{2} a_2 V^2}{a_1 V + \frac{3}{4} a_3 V^2} \right| \cong \frac{V}{2} \left| \frac{a_2}{a_1} \right| = \left| \frac{A(V) - A(-V)}{8A(0)} \right|, \quad \frac{3}{4} |a_3| V << |a_1|$$

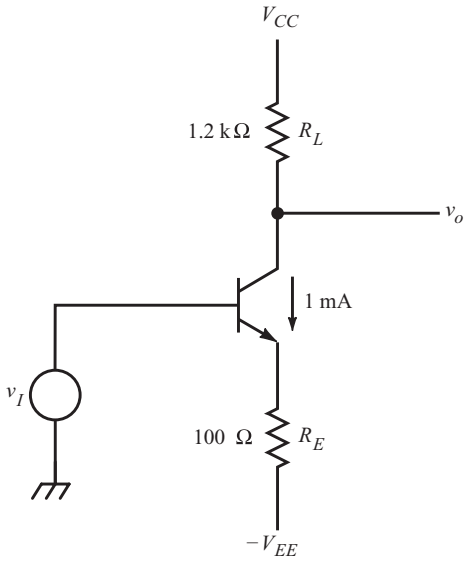
The condition of $2HD$ is that the distortion be small. Similarly, for third-harmonic distortion,

$$3HD = \left| \frac{\frac{1}{4}a_3V^3}{a_1V + \frac{3}{4}a_3V^2} \right| \cong \frac{V^2}{4}\left|\frac{a_3}{a_1}\right| = \left|\frac{A(V) + A(-V) - 2A(0)}{24A(0)}\right|, \quad \frac{3}{4}|a_3|V << |a_1|$$

The total harmonic distortion can be estimated as the rms sum of the nHD.

Example: Estimation of CE Harmonic Distortion

The CE amplifier stage is analyzed for second and third harmonic distortion for an input sinusoid with $V = 10$ mV amplitude. The BJT is biased with 1 mA of emitter current. $\alpha \cong 1$.



The distortion is due to variation in r_e due to i_e. First find the gain at the three input levels: −10 mV, 0 V, +10 mV. The gain in general is

$$A = -\frac{R_L}{R_E + r_e}$$

Then substitute values

$$A(0) = -\frac{1.2\,\text{k}\Omega}{100\,\Omega + 26\,\Omega} = -9.52$$

At $V_I = 10$ mV, $\Delta I_E \cong 10$ mV/126 Ω ≅ 80 μA. Then $r_e \cong 24\ \Omega$. The value of $A(V)$ is then

$$A(10\,\text{mV}) = -\frac{1.2\,\text{k}\Omega}{100\,\Omega + 24\,\Omega} = -9.68$$

Similarly,

$$A(-10\,\text{mV}) = -\frac{1.2\,\text{k}\Omega}{100\,\Omega + 28\,\Omega} = -9.38$$

With the calculated gains, solve for the distortion values

$$2HD \cong \left|\frac{(-9.68) - (-9.38)}{8 \cdot (-9.52)}\right| = 0.39\%$$

$$3HD \cong \left|\frac{(-9.68) + (-9.38) - 2 \cdot (-9.52)}{24 \cdot (-9.52)}\right| = 0.0086\%$$

If the input amplitude is increased to 25 mV,

$$2HD \cong 4.2\%$$

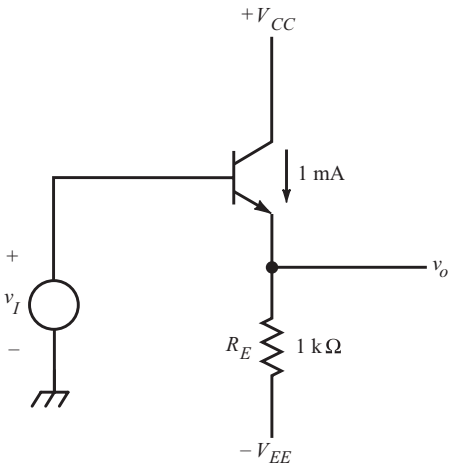
and

$$3HD \cong 0.055\%$$

Example: Estimation of CC Harmonic Distortion

The common collector (CC) amplifier has a sinusoidal input of 0.25 V in amplitude and is biased at an emitter current of 1 mA. The calculations of gain are similar to those of the previous example, with voltage gain,

$$A = \frac{R_E}{R_E + r_e}$$



The three gains are

$$A(0) = \frac{1\,\text{k}\Omega}{1\,\text{k}\Omega + 26\,\Omega} = 0.975$$

At $V_I = 0.25$ V, the emitter current changes by approximately

$$\frac{0.25\,\text{V}}{1.026\,\text{k}\Omega} = 0.244\,\text{mA}$$

Then $I_E \cong 1\text{ mA} + 0.244\text{ mA} = 1.24\text{ mA}$, $r_e \cong 21\ \Omega$, and

$$A(0.25\,\text{V}) = \frac{1\,\text{k}\Omega}{1\,\text{k}\Omega + 21\,\Omega} = 0.980$$

And similarly,

$$A(-0.25\,\text{V}) = \frac{1\,\text{k}\Omega}{1\,\text{k}\Omega + 34\,\Omega} = 0.967$$

The distortion is

$$2HD \cong \left|\frac{(0.980) - (0.967)}{8 \cdot (0.975)}\right| = 0.17\%$$

$$3HD \cong \left|\frac{(0.980) + (0.967) - 2 \cdot (0.975)}{24 \cdot (0.975)}\right| = 0.013\%$$

When analyzed by this technique, the two-transistor diff-amp is found to have no $2HD$ and dominant $3HD$. In general, odd transfer functions, for which $v_O(v_I) = -v_O(-v_I)$, have harmonics that are only odd-integer multiples of the fundamental, or *odd harmonics*, whereas even functions, for which $v_O(v_I) = v_O(-v_I)$ have only even harmonics.

Harmonic distortion consists of frequencies that are harmonically related (integer multiples of) the fundamental. Nonlinearity also produces interactions between frequency components of the input signal when it is not a sinusoid. This form of distortion is *intermodulation* (IM) *distortion*. It consists of sum and difference frequencies as a modulator or multiplier produces.

IM distortion can be demonstrated by a system with only a quadratic term that produces only distortion terms in the output:

$$v_O = v_I^2$$

Instead of a single sinusoid, v_I is set to be the sum of two sinusoids at different frequencies:

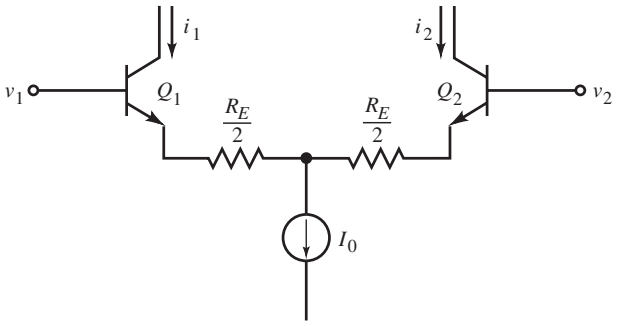
$$v_I = V_1 \cdot \sin \omega_1 t + V_2 \cdot \sin \omega_2 t$$

Substituting v_I gives the output

$$v_O = \underbrace{\frac{V_1^2 + V_2^2}{2}}_{\substack{\uparrow \\ \text{offset} \\ \text{error}}} - \underbrace{\frac{V_1^2}{2}\cos 2\omega_1 t - \frac{V_2^2}{2}\cos 2\omega_2 t}_{2HD} + \underbrace{V_1 V_2[\cos(\omega_1 - \omega_2)t - \cos(\omega_1 + \omega_1)t]}_{\text{IM}}$$

Because v_o is purely nonlinear, all terms of the above expression for it are distortion terms. Besides second-harmonic distortion of each input frequency, sum and difference frequencies are also produced. These are the IM terms.

TRANSCONDUCTANCE LINEARITY OF BJT DIFF-AMP



The BJT differential amplifier stage, or *emitter-coupled pair*, is of interest because of its frequent use in precision circuits. As the external emitter resistance R_E is increased, the input dynamic range is extended, and the effect of r_e decreases because of the much larger series R_E. Thus Δr_e is less significant and linearity increases. To gain greater insight into the BJT diff-amp, we derive its differential transconductance. Let the output current be

$$i_O = i_2 - i_1$$

and let the emitter bias current be

$$I_0 = i_1 + i_2$$

The differential input is

$$v_I = v_2 - v_1$$

Assume $\alpha = 1$ and matched junctions: $I_{S1} = I_{S2} = I_S$. Then Kirchhoff's voltage law (KVL) is applied around the input loop:

$$v_I = V_T \ln\left(\frac{i_2}{I_S}\right) + i_2\left(\frac{R_E}{2}\right) - i_1\left(\frac{R_E}{2}\right) - V_T \ln\left(\frac{i_1}{I_S}\right) = V_T \ln\left(\frac{i_2}{i_1}\right) + i_O\left(\frac{R_E}{2}\right)$$

The first term can be expressed as

$$V_T \cdot \ln\left(\frac{1 + i_O/I_0}{1 - i_O/I_0}\right) = 2V_T \cdot \tanh^{-1}\left(\frac{i_O}{I_0}\right)$$

The incremental transconductance is

$$G_m = \frac{di_O}{dv_I}$$

G_m is found by implicitly differentiating v_I and solving;

$$G_m = \frac{1}{\dfrac{2V_T}{I_0\left[1 - \left(i_O/I_0\right)^2\right]} + \dfrac{R_E}{2}}$$

G_m depends on i_O and is not linear. When $v_I = 0$, the amplifier is balanced, and $i_O = 0$ A. Then G_m is maximum and is

$$G_m(0) = \frac{1}{2V_T/I_0 + R_E/2} = \frac{1}{r_e + R_E/2}$$

This result, when linearized, is consistent with the transresistance method in *Designing Amplifier Circuits*.

Instead of calculating distortion, define the error in G_m as

$$\varepsilon = -\frac{G_m - G_m(0)}{G_m(0)} = \frac{\left(i_O/I_0\right)^2}{\left(R_E I_0/4V_T\right)\left[1-\left(i_O/I_0\right)^2\right]+1}$$

With $i_2 = 0.75$ mA and $i_1 = 0.25$ mA, then $i_O/I_0 = 0.5$. Assuming $R_E = 104\ \Omega$, $\varepsilon = 14.3\%$. With $R_E = 0\ \Omega$, ε is simply $(i_O/I_0)^2$, or 25%.

When $R_E = 0\ \Omega$, v_I can be solved for i_O. First,

$$\frac{i_2}{i_1} = e^{(v_I/V_T)}$$

Applying I_0, i_1 can be expressed as

$$i_1 = \frac{I_0}{e^{(v_I/V_T)}+1}$$

By definition,

$$\tanh x \equiv \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

and

$$\frac{1}{2}\cdot[1-\tanh x] = \frac{1}{e^{2x}+1}$$

By a change of variable, let $x \to x/2$. Then,

$$\frac{1}{2}\cdot\left[1-\tanh\frac{x}{2}\right] = \frac{1}{e^x+1}$$

Using this relationship with i_1,

$$i_1 = \frac{1}{2} \cdot I_0 \cdot \left(1 - \tanh \frac{v_I}{2V_T}\right)$$

The differential output current is

$$i_O = i_2 - i_1 = I_0 - 2i_1 = I_0 - I_0 \cdot \left(1 - \tanh \frac{v_I}{2V_T}\right)$$

or

$$i_O = I_0 \cdot \tanh \frac{v_I}{2V_T}$$

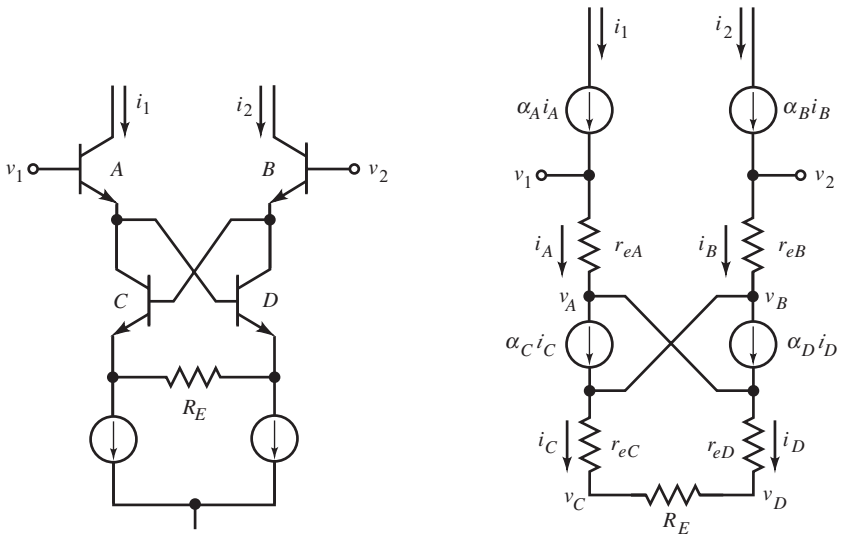
The hyperbolic tangent, an odd function, is linear around the origin and flattens out to ±1 for large inputs. At room temperature, $2V_T \cong 52$ mV. The deviation of i_O/I_0 from a line tangent at the origin for a few values of v_I demonstrates the useful input range of v_I.

v_I, mV	$v_I/2V_T$	i_O/I_0	i_O error, %
10.4	0.20	0.197	–1.31
17.1	0.33	0.319	–3.48
25.9	0.50	0.462	–7.58
38.8	0.75	0.635	–15.3
51.7	1.00	0.762	–23.8
77.6	1.50	0.905	–39.7
103.5	2.00	0.964	–51.8
258.7	5.00	1.000	–80.0

About 100 mV of input is the practical limit of the dynamic range. Above this, severe compression of i_O results. With zero R_E, the transconductance is

$$G_m|_{R_E=0} = \frac{I_0}{2V_T} \cdot \left[1 - \left(\frac{i_O}{I_0}\right)^2\right]$$

The effect of R_E is to linearize the hyperbolic tangent curve around the origin and add a linear term to G_m.



Various schemes have been devised to improve the linearity of the BJT diff-amp. The *cross-quad* circuit (a) is one of them. The input voltage is applied across four *b-e* junctions, two per side, and the common emitter resistance R_E. The voltage drop around the input loop is

$$\begin{aligned} v_2 - v_1 &= v_B + v_C + v_{RE} - v_D - v_A \\ &= V_T \ln\left(\frac{i_B i_C}{I_S^2}\right) + i_{RE} \cdot R_E - V_T \ln\left(\frac{i_D i_A}{I_S^2}\right) \end{aligned}$$

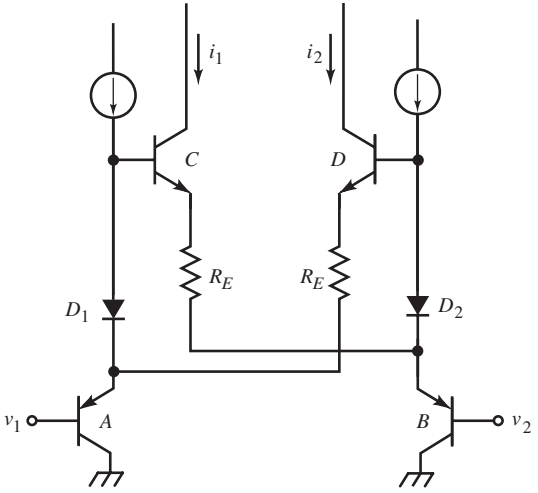
where, for instance, the base-emitter junction drop of transistor A is v_A. This equation can be further simplified for matched transistors to

$$v_2 - v_1 = V_T \ln\left(\frac{i_B i_C}{i_D i_A}\right) + (i_A - i_B) \cdot R_E$$

For $\alpha = 1$, $i_B = i_D$ and $i_C = i_A$. This reduces the log term to zero, leaving only the linear term. The large-signal transconductance is

$$G_M = \frac{i_O}{v_I} \cong \frac{i_B - i_A}{v_2 - v_1} = -\frac{1}{R_E}$$

The input loop is linearized by summing voltages across junctions that are conducting currents from both sides.



A variation on this idea, using complementary pairs of BJTs and diodes for biasing, is shown above. The diodes bias the NPN pair to approximately the same currents as the PNP pair at zero input voltage.

Incremental analysis of the cross-quad circuit around $v_I = 0\,\text{V}$ can assume equal static current in the four BJTs so that r_e is the same for all of them. Assuming matched transistors with equal β gives the inverted small-signal transconductance,

$$\frac{v_i}{i_o} = -\frac{1}{\alpha}\cdot\left[\left(\frac{2}{\beta-1}\right)\cdot r_e + \left(\frac{\beta+1}{\beta-1}\right)\cdot\frac{R_E}{2}\right] \cong -\left[\frac{2r_e}{\beta} + \frac{R_E}{2}\right], \quad \beta >> 1$$

Without the cross-BJTs, $2r_e$ would not be divided by β. In effect, the nonlinearity is reduced by β. The cross-quad amplifier is inverting because input voltage v_2 mainly drives BJT C, which generates i_1.

BJT and FET Diff-Amp Temperature Characteristics

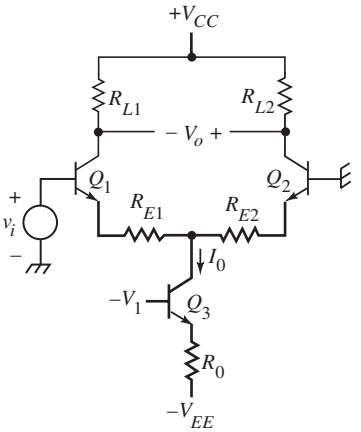
Temperature affects the gain of emitter-coupled BJT pairs (diff-amps) by changing dynamic emitter resistance:

$$r_e = \frac{V_T}{|I_E|}$$

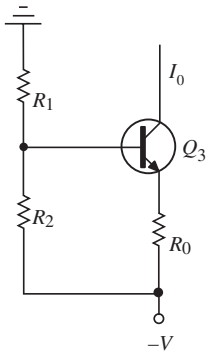
If emitter current is held constant over temperature, the thermal voltage,

$$V_T = \frac{kT}{q_e}$$

causes r_e to vary in proportion to absolute temperature.



A simple compensation is to make the emitter current source, I_0, of the differential BJT pair also proportional to T. Then the TCs of V_T and I_E cancel, and gain variation due to $r_e(T)$ is removed. Such an I_0 supply can be implemented as a CE stage without temperature compensation, as shown above. The TC of Q_3 compensates Q_1 and Q_2 for temperature-independent gain.



Consider a voltage-divider circuit supplying Q_3 base voltage, as shown. This implementation of I_0 is more versatile and more common in occurrence than the previous scheme. The base divider provides extra freedom for setting TC%(I_0), which, for ignored TC(β), is now

$$\begin{aligned}
\mathrm{TC\%}(I_0) &= \frac{1}{I_0}\cdot\frac{dI_0}{dT} = \frac{1}{I_0}\cdot\frac{d}{dT}\left(\frac{\left(\frac{R_2}{R_1+R_2}\right)\cdot V - V_{BE}}{R_0+(R_1\|R_2)/(\beta+1)}\right)\\
&= \frac{1}{I_0}\cdot\frac{\left(-\frac{dV_{BE}}{dT}\right)}{R_0+(R_1\|R_2)/(\beta+1)}\\
&= \frac{\left(-\frac{dV_{BE}}{dT}\right)}{\left(\frac{R_2}{R_1+R_2}\right)\cdot V - V_{BE}}\\
&= \frac{\frac{1}{V_{BE}}\cdot\left(-\frac{dV_{BE}}{dT}\right)}{\left(\frac{R_2}{R_1+R_2}\right)\cdot\frac{V}{V_{BE}}-1} = \frac{-\mathrm{TC\%}(V_{BE})}{\left(\frac{R_2}{R_1+R_2}\right)\cdot\frac{V}{V_{BE}}-1}
\end{aligned}$$

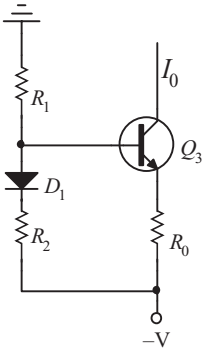
The divider ratio that gives the correct compensation can now be found. When TC%(I_0) is set equal to TC%(V_T),

$$\begin{aligned}
\left(\frac{R_2}{R_1+R_2}\right)\cdot\frac{V}{V_{BE}} &= \frac{-\mathrm{TC\%}(V_{BE})}{\mathrm{TC\%}(V_T)}+1\\
&= \frac{+2\,\mathrm{mV/^\circ C/}V_{BE}}{0.33\%/^\circ\mathrm{C}}+1 = \frac{0.6\,V+V_{BE}}{V_{BE}}
\end{aligned}$$

or

$$\frac{R_2}{R_1+R_2} = \frac{0.6\,\mathrm{V}+V_{BE}}{V} \cong \frac{1.25\,\mathrm{V}}{V}$$

for $V_{BE} = 0.65$ V. This result is interesting: Whatever the value of V, the unloaded divider voltage must be 1.25 V for gain compensation. This is also the voltage of bandgap references, as well it should be. Bandgap circuits use the negative TC(V_{BE}) and scale it to cancel the positive TC(V_T). When this is done, the resulting bandgap voltage always comes out to be close to 1.25 V and varies slightly with BJT doping levels.



Another current-source variation that is often used to provide rough temperature compensation is to insert a diode in series with R_2, as shown above. The usual explanation is that the TC of the diode compensates for the TC of the BJT *b-e* junction, resulting in a more stable I_0. A typical example is to use a 1N4152 diode to compensate a PN3904. The diode and BJT *b-e* junctions are quite different, however. The diode doping levels are far less than the BJT base, to achieve a higher breakdown voltage. The emitter minority carrier concentration is made intentionally large for good emitter injection efficiency into the base at the expense of V_{BE} reverse breakdown, which is typically around 7 V, much below the 40 V of the diode. Although both junctions are silicon, they are rather unmatched.

Suppose, however, that a similar BJT is used as a diode, with base connected to collector. Then the junction matching is better (though not as good as adjacent integrated BJTs), and allowing for $\alpha \cong 1$, then applying KVL around the BJT input loop,

$$V_T \cdot \ln\left(\frac{I_0}{I_D}\right) = I_D \cdot R_2 - I_0 \cdot R_0 \quad \text{or} \quad I_0 = \frac{I_D \cdot R_2 - V_T \cdot \ln\left(I_0 / I_D\right)}{R_0}$$

where I_D is the diode current. If the junction currents are equal, the TC due to V_T is removed and the $\mathrm{TC}\%(I_0) \cong 0\%/^\circ\mathrm{C}$. This is useful for applications where a stable current source is needed, but it does not compensate r_e of the diff-amp. The currents must deliberately be set unequal to achieve the desired TC, and for a compensating polarity of TC, it must be positive. Consequently, we must have that $I_D > I_0$.

The $\mathrm{TC}\%(I_0)$ is found through implicit differentiation of I_0 in the above equation:

$$\begin{aligned}\mathrm{TC}\%(I_0) &= \frac{1}{I_0}\cdot\frac{dI_0}{dT} = \frac{1}{I_0\cdot R_0}\cdot\left(-V_T\cdot\frac{d}{dT}\ln\left(\frac{I_0}{I_D}\right) - \ln\left(\frac{I_0}{I_D}\right)\cdot\frac{dV_T}{dT}\right)\\ &= -\frac{V_T}{I_0\cdot R_0}\cdot\left(\mathrm{TC}\%(I_0) + \frac{1}{T}\cdot\ln\left(\frac{I_0}{I_D}\right)\right)\end{aligned}$$

With additional algebraic manipulation,

$$\mathrm{TC}\%(I_0) = -\frac{(V_T/I_0\cdot R_0)\cdot\ln(I_0/I_D)}{T\cdot(1+V_T/I_0\cdot R_0)}$$

Then to compensate, set $\mathrm{TC}\%(I_0) = \mathrm{TC}\%(V_T) = 1/T$ and solve

$$\frac{I_D}{I_0} = \exp\left(1+\frac{I_0\cdot R_0}{V_T}\right)$$

Because of the exponential function, practical current ratios require that the voltage across R_0 be not much larger than V_T. For $I_0 = 2$ mA, $R_0 = 22\ \Omega$, and $V_T = 26$ mV, then the voltage across R_0 is 44 mV, or $1.69\cdot V_T$, and $I_D = 14.77\cdot I_0 = 29.5$ mA, larger than is desired in most designs. Such small values of R_0 are required to keep R_0 from dominating the emitter circuit so that the TC of V_{BE} can be expressed. Yet in many designs, R_0 is relatively large, and its voltage drop far exceeds V_T. As a consequence, the $\mathrm{TC}\%(V_T)$ of r_e is not correctly compensated, and a TC drift in gain exists.

The previous scheme, which omitted the base diode, was only slightly better in allowing for larger R_0 voltage. Perhaps we should go in the opposite direction and add a diode or two in the emitter. The TC of the combined junctions would

be multiplied by the number of them, and that would allow R_E to be made proportionally larger. It is usually not desirable to add a large number of series diodes because the static stability of I_0 is not benefited.

Consequently, use of the diff-amp current source to temperature-compensate r_e results in a circuit requiring careful static design. I_0 is then made sensitive to junction parameters, and these parameters, such as I_S, have a somewhat wide tolerance among discrete transistors, even of the same part number. Expect as much as 50 mV of variation among PN3904 BJTs at the same current and temperature. This compensation method is best suited for monolithic integration.

Ambient temperature variations can cause changes in offset voltage of a diff-amp. This *offset voltage drift* is derived by applying KVL to the input loop of a BJT diff-amp. The input offset voltage is

$$V_{os} = V_2 - V_1 = V_T \ln\left(\frac{I_{E2}}{I_{S2}}\right) - V_T \ln\left(\frac{I_{E1}}{I_{S1}}\right) = V_T \ln\left(\frac{I_{E2}}{I_{E1}}\right) - V_T \ln\left(\frac{I_{S1}}{I_{S2}}\right)$$

The thermal voltage, $V_T = kT/q_e$, is in both terms, making them both temperature dependent. The first term has drift due to emitter-current mismatch; the second is due to mismatched transistor I_S. BJT matching, especially monolithic matching, minimizes the second term. The first term requires that the static currents of the differential pair be equal. Its drift TC is

$$\frac{dV_{os}}{dT} = \frac{k}{q_e} \cdot \ln\left(\frac{I_{E2}}{I_{E1}}\right) \cong (198\,\mu\text{V}/^\circ\text{C}) \cdot \log\left(\frac{I_{E2}}{I_{E1}}\right)$$

A 1 μV/°C TC requires a 1% match of emitter currents. For perfectly matched BJTs, zero TC occurs at zero offset voltage. Slight mismatches cause the second term of V_{os} to be significant. The matched-BJT TC has nonzero offset when the emitter currents are equal. In addition, ohmic emitter resistance variations and surface leakage introduce offset error.

Another useful quantity is the drift TC per offset voltage, for matched BJTs:

$$\frac{dV_{os}/dT}{dV_{os}} \cong 3.3 \frac{\mu\text{V}/^\circ\text{C}}{\text{mV}}, \quad T = 300\,\text{K}$$

A 1 mV offset produces a drift of 3.3 μV/°C.

Two adjustments are required to precisely balance a practical diff-amp because offset voltage drift is not nulled at zero offset voltage. Offset voltages are also caused by load resistor and r_o mismatch. The emitter currents are adjusted first for zero TC. This adjustment introduces its own output offset voltage, which is nulled by adjusting the load resistor balance.

For junction field-effect transistor (JFET) diff-amps, the drift mechanisms are carrier mobility μ and gate-channel junction barrier voltage ϕ. With a constant applied V_{GS}, the terminal V_{GS} is the effective $V_{GS} + \phi$. Then $dV_{GS} = d\phi$. The V_{GS} TC is thus

$$\frac{dV_{GS}}{dT} = \frac{\partial V_{GS}}{\partial \phi} \cdot \frac{d\phi}{dT} + \frac{\partial V_{GS}}{\partial I_D} \cdot \frac{\partial I_D}{\partial \mu} \cdot \frac{d\mu}{dT}$$

We need the TCs of ϕ and μ. The TC of ϕ is that of a *pn* junction, or about –2 mV/°C at room temperature. For an *n*-channel JFET, $d\phi/dT > 0$. Mobility varies with doping concentration and type, and only an average approximation can be given. Its fractional TC is about –0.5%/°C. The drain current of a JFET in the current saturation region is

$$i_D = I_{DSS} \cdot \left(1 - \frac{v_{GS}}{V_P}\right)^2$$

where I_{DSS} is the drain current when $V_{GS} = 0$ V and V_P is the pinch-off voltage. $V_P < 0$ V for an *n*-channel JFET. The saturation region is defined by

$$V_{DS} \geq V_{GS} - V_P$$

I_D is a function of device geometry and electrical parameters. It is directly dependent on channel mobility, so that

$$\frac{dI_D}{I_D} = \frac{d\mu}{\mu}$$

The equation for dV_{GS}/dT can now be simplified upon noting that

$$\frac{dI_D}{dV_{GS}} = g_m = -\frac{2I_{DSS}}{V_P} \cdot \left(1 - \frac{V_{GS}}{V_P}\right)$$

Substituting into dV_{GS}/dT,

$$\frac{dV_{GS}}{dT} \cong (2\,\text{mV}/^{\circ}\text{C}) + (-0.5\%/^{\circ}\text{C}) \cdot \left(\frac{I_D}{g_m}\right)$$

For zero voltage offset TC,

$$\left.\frac{I_D}{g_m}\right|_{\text{TC}(V_{GS})=0} \cong 0.4\,\text{V}$$

Divide I_D from $i_D(v_{GS})$ by $g_m = dI_D/dV_{GS}$, resulting in

$$\frac{I_D}{g_m} = \frac{1}{2} \cdot (V_{GS} - V_P)$$

Set this equal to I_D/g_m at zero TC and solve for the zero-TC V_{GS}:

$$V_{GSZ} = V_{GS}|_{\text{TC}=0} \cong V_P + 0.8\,\text{V}$$

In other words, the zero-drift V_{GS} is about 0.8 V above pinch-off. The typical range of V_{GSZ} is −2 V to −4 V. The corresponding zero-TC I_D is

$$I_{DZ} = I_D|_{V_{GS}\text{TC}=0} = I_{DSS}\left(-\frac{0.8\,\text{V}}{V_P}\right)^2$$

I_{DZ} is typically several hundred microamps. The dV_{GS}/dT equation can be written in terms of V_{GSZ} by substituting for V_P from V_{GSZ}:

$$\frac{dV_{GS}}{dT} \cong (-0.25\%/^{\circ}\text{C}) \cdot (V_{GS} - V_{GSZ})$$

This formula is compared to $(dV_{os}/dT)/dV_{os}$ for a BJT. The TC per volt of offset from V_{GSZ} is 0.25%/°C or 2.5 μV/°C per millivolt of offset.

To find the effect of I_D on TC(V_{GS}), substitute I_{DSS} of I_{DZ} into i_D, retaining the sign within the square:

$$I_D = I_{DZ} \cdot \left(\frac{V_P - V_{GS}}{-0.8\,\text{V}}\right)^2$$

Then replace V_P from V_{GSZ}, solve for $V_{GS} - V_{GSZ}$, and substitute it into dV_{GS}/dT. The result is

$$\frac{dV_{GS}}{dT} = (-2\,\text{mV}/^{\circ}\text{C}) \cdot \left(1 - \sqrt{\frac{I_D}{I_{DZ}}}\right)$$

The sign can be checked by noting that when $I_D > I_{DZ}$, the TC is positive. The TC(V_{GS}) is opposite in sign to TC(I_D). Below I_{DZ}, TC(I_D) is positive because it is dominated by TC(ϕ). As $|\phi|$ decreases with increasing temperature, the effective V_{GS} increases, and more current flows. If I_D is forced to be constant, then V_{GS} must decrease. Thus, below V_{GSZ}, TC(V_{GS}) < 0. Above I_{DZ}, the TC(I_D) < 0 and is dominated by TC(μ) < 0. Here, $V_{GS} > V_{GSZ}$ and TC(V_{GS}) > 0. For constant I_D, V_{GS} must increase with increasing temperature.

These results apply to a single JFET. For a matched differential pair with currents I_{D1} and I_{D2},

$$V_{os} = V_{GS2} - V_{GS1}$$

and

$$\frac{dV_{os}}{dT} = \frac{dV_{GS2}}{dT} - \frac{dV_{GS1}}{dT} = (-2\,\text{mV}/^{\circ}\text{C}) \cdot \left(\sqrt{\frac{I_{D1}}{I_{DZ}}} - \sqrt{\frac{I_{D2}}{I_{DZ}}}\right)$$

The diff-amp is compensated when each drain current is I_{DZ}, and

$$I_{D1} + I_{D2} = 2 \cdot I_{DZ}$$

The radicals can be written in the form

$$\frac{I_{D1}}{I_{DZ}} = \frac{2I_{D1}}{2I_{DZ}} = \frac{(I_{D1} + I_{D2}) - (I_{D2} - I_{D1})}{2I_{DZ}} = 1 - \frac{I_O}{2I_{DZ}}$$

$$\frac{I_{D2}}{I_{DZ}} = \frac{(I_{D1} + I_{D2}) + (I_{D2} - I_{D1})}{2I_{DZ}} = 1 + \frac{I_O}{2I_{DZ}}$$

dV_{os}/dT can now be expressed as

$$\frac{dV_{os}}{dT} = (2\,\text{mV}/^{\circ}\text{C})\cdot\left(\sqrt{1+\frac{I_O}{2I_{DZ}}} - \sqrt{1-\frac{I_O}{2I_{DZ}}}\right)$$

where $I_O = I_{D2} - I_{D1}$ is the output current offset. This error is usually small relative to I_{DZ}, so that the binomial approximations,

$$(1 \pm x)^{1/2} \cong 1 \pm \frac{x}{2} - \frac{x^2}{8}, \quad |x| << 1$$

can be applied. Then the TC reduces to

$$\frac{dV_{os}}{dT} = (2\,\text{mV}/^{\circ}\text{C})\cdot\left(\frac{I_{D2} - I_{D1}}{2I_{DZ}}\right)$$

Comparing this result to dV_{os}/dT for a BJT diff-amp, we find that a TC of 1 μV/°C occurs when the currents are mismatched by 0.1%. For the same TC, current matching must be an order of magnitude better for FETs than BJTs, which is why the input offset specification of FET-input op-amps is generally worse than that of their BJT counterparts.

The preceding derivations were for JFETs, but the results, including i_D, hold for MOSFETs too. The difference is in I_{DSS}:

$$I_{DSS}(\text{JFET}) = \frac{\mu \cdot C}{2 \cdot L^2}$$

$$I_{DSS}(\text{MOSFET}) = \frac{\mu \cdot \varepsilon \cdot W}{2 \cdot L \cdot T} = \frac{\mu}{2 \cdot L^2} \cdot \frac{\varepsilon \cdot L \cdot W}{T} = \frac{\mu}{2 \cdot L^2} \cdot C_G$$

where μ is mobility, L is channel length, W is channel width, T is gate oxide thickness, and ε is the gate-silicon dielectric constant. As shown by this equation, even the form of I_{DSS} is the same. The difference is in how the gate and channel capacitances relate to the different device structures. Also, the MOSFET analog of pinch-off voltage V_P is threshold voltage V_T. The difference here is largely semantic.

Thermal Distortion

As the voltages and currents of electronic components vary, their power dissipation varies also. For components with parameters that are significantly affected by temperature, this self-heating can be regarded as a temperature signal. The electrical response is a kind of dynamic thermal distortion or noise.

Semiconductor devices are strongly sensitive to temperature variation. A PN junction has a voltage TC of typically −2 mV/°C. As a diode is heated, its *v-i* curve moves toward the vertical (current) axis, as seen on a curve tracer. The β of BJTs is sensitive to temperature but is hard to derive theoretically. It is typically about +1%/°C for silicon BJTs. The most significant parameter for gain variation with temperature is r_e, which is directly proportional to the thermal voltage,

$$V_T = \frac{kT}{q_e}$$

At an ambient temperature of $T = 300$ K, the fractional TC(V_T) is

$$\mathrm{TC}\%(V_T) = \frac{dV_T/dT}{V_T} = \frac{1}{T} = \frac{1}{300\,\mathrm{K}} \cong 0.33\%/\mathrm{K} = 0.33\%/^{\circ}\mathrm{C}$$

As a BJT heats on a curve tracer, not only does the *b-e* junction *v-i* curve decrease in voltage, but its slope at a given current also becomes less steep; that is, r_e increases. These three thermal effects must be considered in circuit design.

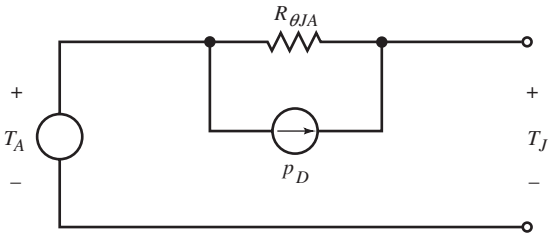
A well-designed amplifier is not very β-sensitive; α appears in front of most gain expressions and depends on β as

$$\frac{d\alpha}{dT} = \left(\frac{1}{\beta+1}\right)^2 \cdot \frac{d\beta}{dT}$$

For $\beta = 100$, TC(α) ≅ 1 ppm and can be ignored.

Good amplifier design also minimizes the dependence of gain on r_e. The remaining effect is the TC of v_{BE}. BJT *b-e* junctions are highly doped and have a TC of about −2.2 mV/°C; diodes are lightly doped to increase breakdown voltage and have a TC closer to −1.8 mV/°C, similar to the *b-c* junction of BJTs

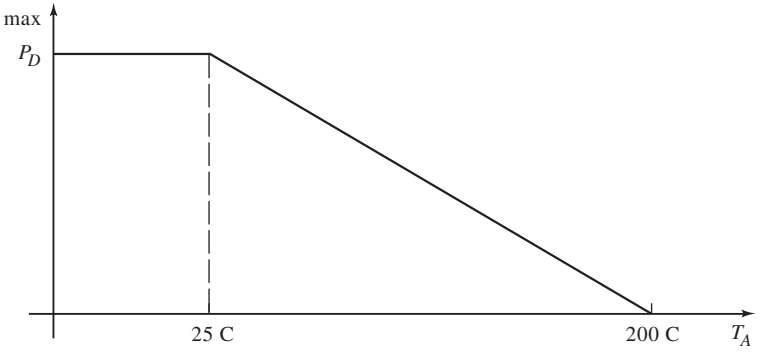
when operated in the inverse mode. Thermal effects are mainly the result of a *b-e* bias shift with temperature change. This shift appears as a dynamic *b-e* signal and is modeled as a voltage source in series with a fixed V_{BE}.



Junction temperature T_J can be derived from the thermal model shown above. Ambient temperature T_A is modeled by a voltage source. In series with it is the thermal resistance from junction to ambient. Dissipated in this resistance is power p_D, modeled as a current. The junction temperature is

$$T_J = p_D \cdot R_{\theta JA} + T_A$$

The maximum power that a transistor can dissipate is limited by the maximum junction temperature. For silicon, this is about 200 °C. The thermal model leads to the power derating curve.



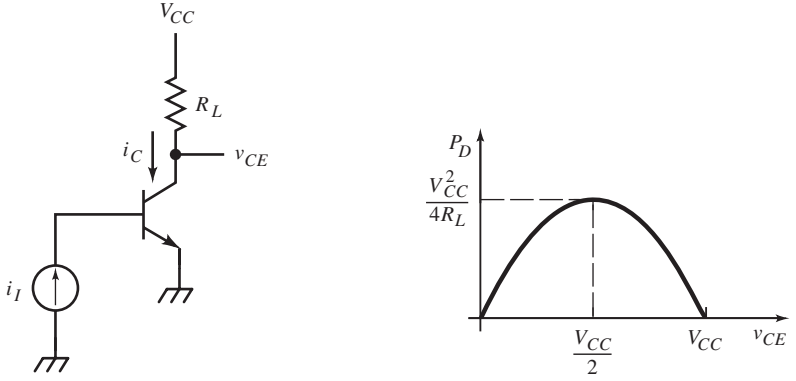
Above $T_A = 25$ °C, maximum power decreases linearly up to the maximum junction temperature. The slope of the derating curve is the thermal conductance

in W/°C. A 16-pin dual-in-line (DIP) IC package has a thermal resistance of about 100 °C/W and a TO-92 transistor package, about 300 °C/W.

Minimization of dynamic thermal effects requires minimization of either $R_{\theta JA}$ or ΔP_D. Semiconductor packaging often limits the minimum $R_{\theta JA}$, though the addition of heat sinks can significantly reduce it. We shall seek ways of minimizing Δp_D instead. Direct reduction in p_D comes from reducing P_D, the static power dissipation. Waveform swings around this quiescent power are scaled down accordingly. As it is, power is often related to waveform quantities which also scale down relative to thermal noise. The approach taken here is to find the operating point at which waveform excursion causes minimal change in power dissipation.

The simple CE circuit (a) below illustrates the basic idea of thermal compensation. BJT power dissipation, by Watt's law, is

$$p_D = v_{CE} \cdot i_E \cong v_{CE} \cdot \left(\frac{V_{CC} - v_{CE}}{R_L} \right) = \left(\frac{V_{CC}}{R_L} \right) \cdot v_{CE} - \frac{v_{CE}^2}{R_L}$$



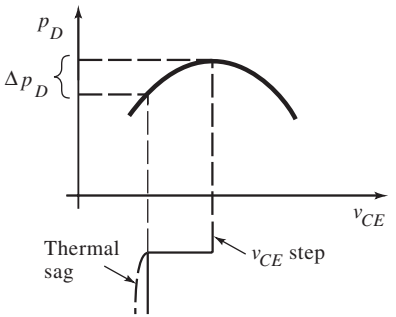
(a) (b)

The curve of p_D is parabolic, (b). Maximum p_D occurs at $V_{CC}/2$ and is

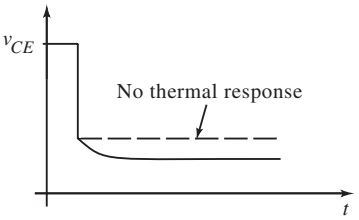
$$\max p_D = \frac{V_{CC}^2}{4R_L}, \quad v_{CE} = \frac{V_{CC}}{2}$$

The slope of dp_D/dv_{CE} is minimum at maximum power. By biasing the transistor at maximum power, waveform excursions from quiescence will cause minimal changes in power.

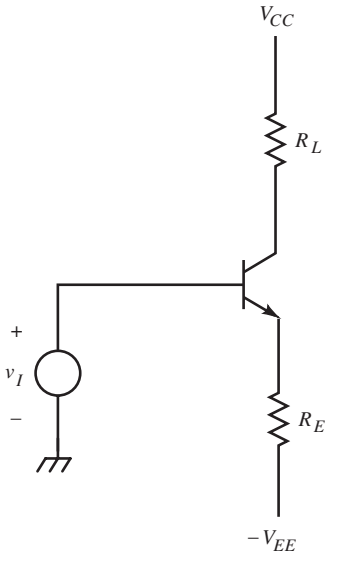
The amplifier gain is highly β and r_e dependent but is not sensitive to Δv_{BE}. When a step of base current is applied, i_C increases causing v_{CE} to decrease, and the power steps down the curve from the peak. The transistor cools. As it does, the junction temperature decreases with time, with a cooling rate determined by the thermal mass between the junction and ambient and the specific heat of this mass. As T_J decreases, v_{BE} increases, and V_T and r_e decrease, causing the gain to increase. This further decreases v_{CE}, and it contributes a thermal "sag" due to increasing gain.



The dominant effect is due to β-dependency. With cooling, β decreases, causing i_C to decrease more than it increases from Δr_e. The two effects tend to cancel but β dominates. This circuit is not typical of good design, although sometimes β and r_e thermal effects are unavoidable. The response is shown below.



A better amplifier circuit is the CE shown below. It is relatively β and r_e insensitive but has a voltage-source input and amplifies Δv_{BE}. Assuming $\alpha = 1$ and $R_E >> r_e$, we get the BJT power dissipation:



$$p_D \cong v_{CE} \cdot i_C$$
$$= \left[V_{CC} - \left(\frac{v_I - V_{EE}}{R_E} \right) \cdot (R_L + R_E) - (-V_{EE}) \right] \cdot \left(\frac{v_I - V_{EE}}{R_E} \right)$$

Differentiating p_D with respect to v_I and solving for v_I,

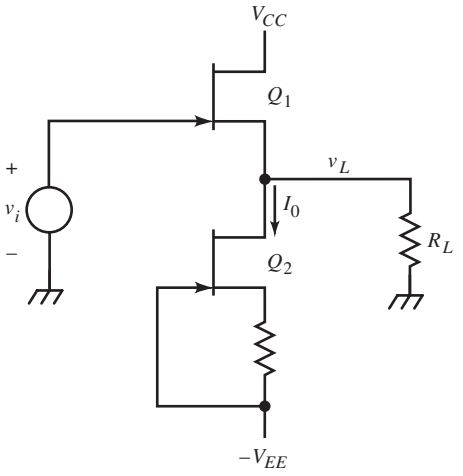
$$v_I|_{\max p_D} = \left(\frac{V_{CC} + V_{EE}}{2} \right) \cdot \left(\frac{R_E}{R_L + R_E} \right) + V_{EE}$$

Usually, v_I is determined by the static or quiescent input level V_I. The supply voltage, $-V_{EE}$, can be adjusted instead. Solving v_I for V_{EE}, then

$$V_{EE}|_{\max p_D} = \frac{2 \cdot V_I \cdot \left(\frac{R_L + R_E}{R_E} \right) - V_{CC}}{\left(\frac{R_L + R_E}{R_E} \right) + 1} > 0$$

When the supply voltages are given, V_{EE} can be made a Thevenin equivalent voltage by the addition of an emitter resistor to ground.

Example: CS Buffer Thermal Compensation



The FET ×1 buffer amplifier uses a matched FET as the current source. Its source resistor is chosen to set V_{GS} to V_{GSZ} so that $I_D = I_{DZ}$, and the current has zero TC. (See "BJT and FET Diff-Amp Temperature Characteristics.") The current from this source is I_0. Our goal is to choose V_{CC} so that thermal balance is achieved when $R_L = 1\ \text{k}\Omega$ and $I_0 = 5\ \text{mA}$.

In this case, two devices are involved. We want the power difference between the FETs to change minimally since a power change in either of them relative to the other introduces thermal distortion. The power difference between them is

$$\Delta p_D = p_1 - p_2 = \left[\left(V_{CC} - v_L\right)\cdot\left(I_0 + \frac{v_L}{R_L}\right)\right] - \left[\left(v_L + V_{EE}\right)\cdot I_0\right]$$
$$= -\frac{v_L^2}{R_L} + \left(\frac{V_{CC}}{R_L} - 2I_0\right)\cdot v_L + \left(V_{CC} - V_{EE}\right)\cdot I_0$$

The maximum power difference occurs at

$$v_L\big|_{\max \Delta p_D} = \frac{V_{CC}}{2} - I_0\cdot R_L \Rightarrow V_{CC}\big|_{\max p_D} = 2\cdot\left(v_L + I_0 R_L\right)$$

where v_L is the quiescent value. Substituting and solving,

$$V_{CC} = 10\,\text{V}$$

If the available supply is not 10 V, a Thevenin equivalent supply can be constructed from a higher-voltage supply. If a series resistor R_C is inserted in the drain of Q_1, then

$$R_C = \frac{\frac{V_{CC}}{2} - (I_0 \cdot R_L + v_L)}{\frac{v_L}{R_L} + I_0}$$

For $V_{CC} = 12$ V and $V_L = 0$ V at quiescence, $R_C = 200\ \Omega$. To avoid the Miller effect, the drain of Q_1 is bypassed with a capacitor to load ground.

For applications in which static offset must be minimized, the addition of the zero-TC bias resistor in the source of the lower FET must be matched by a resistor of equal value between the source of the upper FET and the output. This resistor drops the same voltage as the V_{GS} of the lower FET, thus compensating for the same V_{GS} in the top FET when the same current flows through it. The output loading also must be minimal so that I_0 flows through both FETs.

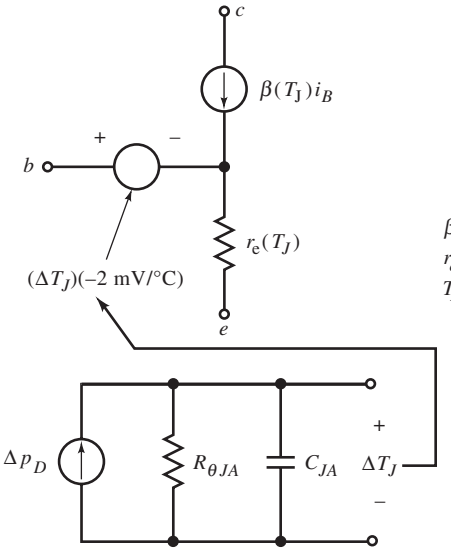
A change in temperature at the *b-e* junction of a BJT is a thermally generated noise. From the power formula for a BJT, the change in power can be calculated, given the input signal change. The resulting power change Δp_D results in a change in v_{BE}. This Δv_{BE} is in series with the base and adds to v_I (in the CE diagram above). Knowing the gain, the effect on the output of Δv_{BE} is found, given

$$\frac{dv_{BE}}{dp_D} = \frac{dv_{BE}}{dT_J} \cdot \frac{dT_J}{dp_D} \cong (-2\,\text{mV}/^\circ\text{C}) \cdot R_{\theta JA}$$

For a 16-pin DIP package,

$$\frac{dv_{BE}}{dT} \cong (-2\,\text{mV}/^\circ\text{C}) \cdot (100^\circ\text{C}/\text{W}) = -200\,\text{mV}/\text{W}$$

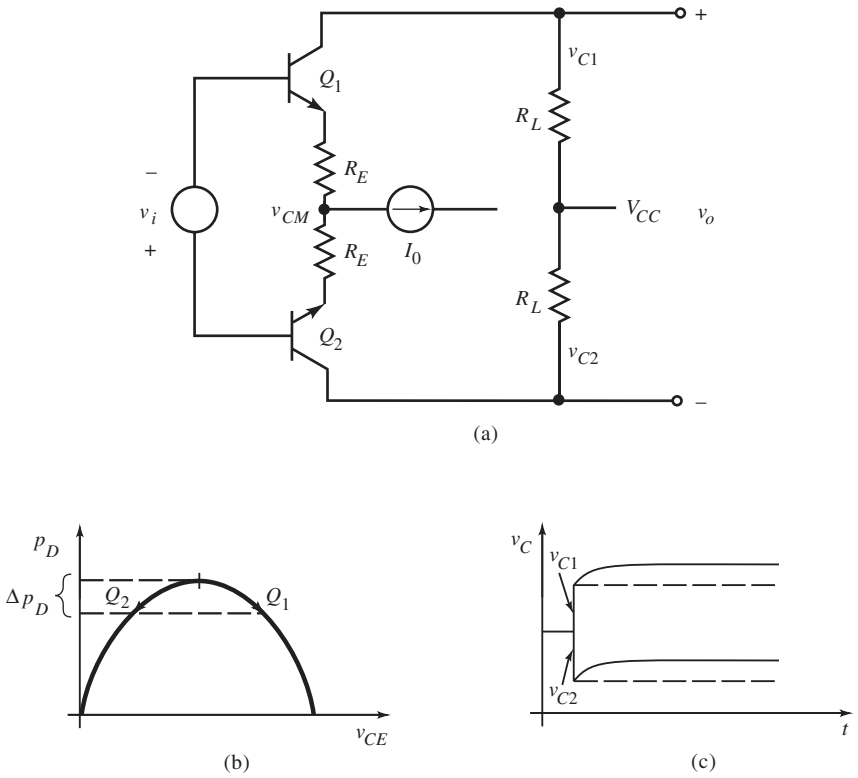
For dynamic analysis, $R_{\theta JA}$ is a thermal impedance, but here only the resistive component is considered. The reactive component, a thermal "capacitance", approximates the response of the thermal effect. The actual response is the solution of Laplace's equation and is only approximated by a first-order capacitive model. In practice, compensation networks use several time constants to approximate the thermal response. The complete BJT model, with thermal effects included, is shown below.



$\beta(T_J) = \beta(T_A)[1 + (0.01)(T_J - T_A)]$
$r_e(T_J) = r_e(T_A)[1 + (0.0033)(T_J - T_A)]$
$T_A \cong 27°C$

Differential amplifiers reject thermal noise as a common-mode voltage when electrically and thermally balanced. With transistors on each side operating at maximum p_D at quiescence and operating on the same power curve, a positive step input to the diff-amp shown in (a) below causes the behavior shown in (b) and (c).

Both transistors move away from their maximum power point. Q_2 conducts more current and its v_{CE} decreases, whereas Q_1 conducts less with increased v_{CE}. Each move down the power curve the same amount, symmetrically. The effect is that the same thermal Δv_{BE} occurs for both, as a small common-mode input.



To derive an approximate formula for the step error, begin with the expressions for p_D for each side. Each transistor is conducting I_C at V_{CE}. An input step v_i perturbs the amplifier from quiescence. Q_1 now conducts $i_{C1} = I_C - i_c$ at $v_{CE1} = V_{CE} + v_{ce}$. For Q_2, $i_{C2} = I_C + i_c$ at $v_{CE2} = V_{CE} - v_{ce}$. The power expressions are

$$p_1 = i_{C1} \cdot v_{CE1} = (I_C - i_c) \cdot (V_{CE} + v_{ce})$$

$$p_2 = i_{C2} \cdot v_{CE2} = (I_C + i_c) \cdot (V_{CE} - v_{ce})$$

Thermal error results from a change in the differential power,

$$\Delta p_D = p_2 - p_1 = 2 \cdot (V_{CE} i_c - I_C v_{ce})$$

The thermal response of v_{BE} due to Δp_D is

$$\text{thermal } \Delta v_{BE} \cong \frac{dv_{BE}}{dp_D} \cdot \Delta p_D$$

Express the incremental variables i_c and v_{ce} in v_i. Neglecting α and r_e, and applying the transresistance method,

$$i_c \cong \frac{v_i}{2R_E}$$

$$v_{ce} = i_c \cdot (R_L + R_E) = \frac{R_L + R_E}{2R_E} \cdot v_i$$

where the polarity of v_{ce} is already in the equations for p. Substituting these expressions and Δp_D into thermal Δv_{BE}, we get the expression for Δv_{BE}. The thermal error, referred to the input, is

$$\frac{\text{thermal } \Delta v_{BE}}{\text{signal } v_i} \cong -\frac{V_{CE} - I_C \cdot (R_L + R_E)}{R_E} \cdot \left(\frac{dv_{BE}}{dp_D} \right)$$

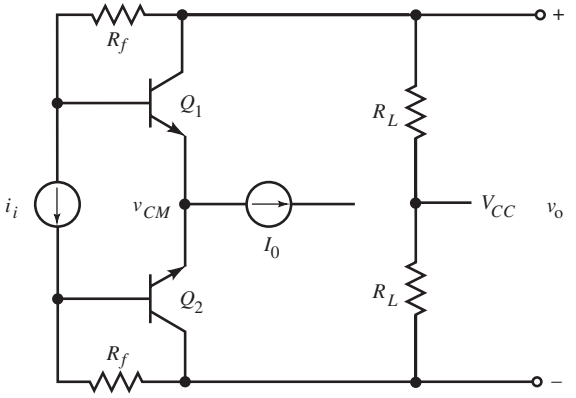
When V_{CE} and the drop across $R_L + R_E$ are equal, they are both half of the supply, and the thermal error is zero.

Δp_D is applicable for differential stages in general. For a particular stage, two incremental, single-ended gains, v_{ce}/v_i and i_c/v_i, are required to compute the thermal error. The general formula for fractional input-referred thermal error is

$$\text{thermal error} = -\left(V_{CE} \cdot \frac{i_c}{v_i} - I_C \cdot \frac{v_{ce}}{v_i} \right) \cdot \left(\frac{dv_{BE}}{dp_D} \right)$$

The general condition for thermal compensation is that thermal error be zero. Then

$$\text{zero thermal error} \rightarrow \frac{v_{ce}}{i_c} = \frac{V_{CE}}{I_C}$$



The figure above shows a differential shunt-feedback amplifier. (Refer to *Designing Amplifier Circuits*, "Shunt-Feedback Amplifier Feedback Analysis" and "Shunt-Feedback Amplifier Substitution Theorem Analysis" for shunt-feedback theory.) The previous equation for thermal error can be expressed as

$$\text{thermal error} = -\left(V_{CE} - I_C \cdot \frac{v_{ce}}{i_c}\right) \cdot \left(\frac{i_c}{v_i}\right) \cdot \left(\frac{dv_{BE}}{dp_D}\right)$$

The output resistance v_{ce}/i_c is found by assuming an ideal transconductance amplifier for the BJT. Then

$$\frac{v_{ce}}{i_c} = \frac{v_{ce}}{i_i} \cdot \frac{i_i}{i_c} \cong R_f \cdot \left(\frac{R_L}{R_f + R_L}\right) = R_f \| R_L$$

The first factor is the ideal shunt-feedback transresistance. The second factor is a current divider for i_c. Ideally, $v_i = 0$. Thus, R_f and R_L are in parallel at the collector node. The second gain is

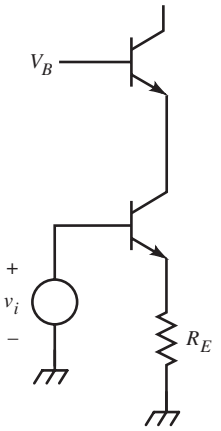
$$\frac{i_c}{v_i} = \frac{1}{R_f}$$

A voltage in series with the base causes current that can only flow in R_f. The ideal amplifier responds by sinking the current as i_c. This causes a drop across

R_f of v_i, keeping the input a virtual ground. The current is determined by R_f. When these expressions are substituted, the thermal error for a differential shunt-feedback amplifier is

$$\text{shunt-feedback thermal error} = -\frac{V_{CE} - I_C(R_f \| R_L)}{R_f} \cdot \left(\frac{dv_{BE}}{dp_D} \right)$$

When $V_{CE}/I_C = R_f \| R_L$, thermal compensation is achieved.

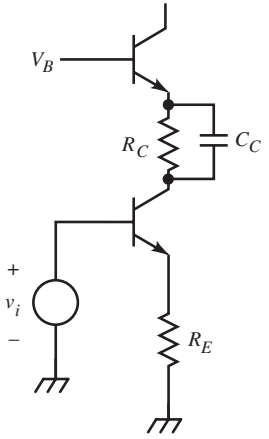


Finally, the CE of the cascode amplifier shown above requires thermal compensation. Its collector is at the fixed voltage of the CB base. The thermal error is found by noting that $v_{ce} = v_i$ and $i_c/v_i = 1/R_E$. Then $v_{ce}/i_c = R_E$. Substituting into the thermal error gives:

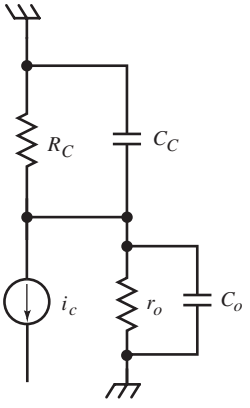
$$\text{cascode thermal error} = -\frac{V_{CE} - I_C \cdot R_E}{R_E} \cdot \left(\frac{dv_{BE}}{dp_D} \right)$$

Thermal compensation requires $V_{CE}/I_C = R_E$. This is often not possible because V_{CE} is determined partly by V_B. To compensate, add a collector series resistor R_C. Then the CE thermal error is expressed by the thermal voltage ratio. The addition of R_C has dynamic response consequences and can be bypassed with capacitor C_C.

$R_C C_C$ forms a time constant, and it is reasonable to wonder by what criterion its value should be chosen. First, C_C must be large enough to have negligible



reactance in the hf region of the CB transistor. But this still leaves a wide range for C_C. Hofer noted that the output capacitance of the CE forms a shunt RC time constant with r_o of the CE as shown below.



By setting

$$R_C \cdot C_C = r_o \cdot C_o$$

a compensated current divider for the CE collector current is formed, and i_e of the CB is

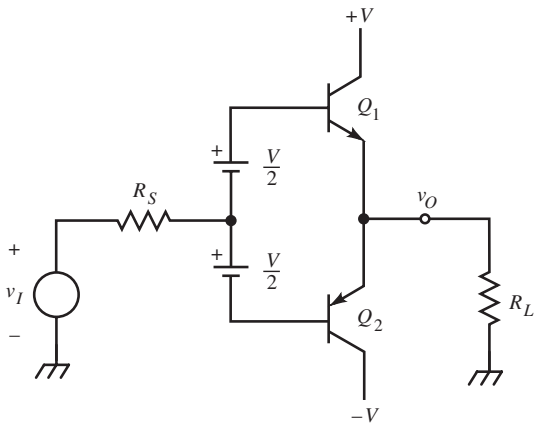
$$\text{CB } i_e = \left(\frac{r_o}{r_o + R_C} \right) \cdot i_c$$

In single-ended or unbalanced differential amplifiers, after Δp_D is minimized, thermal effects remain. At this point, electrical compensation is required. A series *RC* network is commonly used in the emitter circuit to cancel thermal error. Furthermore, the simple single-pole thermal model we have used is only a dominant-pole approximation to most heat-transfer temperature functions. Thus, it is not unusual to find that several series *RC* networks are required to achieve acceptable response.

Complementary Emitter-Follower Output Amplifier

A common need in circuit design is for a bipolar voltage buffer with current-drive capability. Such a circuit is often used as the output stage of a power amplifier, to drive transmission lines and cables with large capacitance, transducers, or power devices. The common requirement is that the driver have large-signal dynamic range relative to its transistors and that the output resistance be constant, usually 50 Ω for transmission lines used to interconnect electronic laboratory equipment.

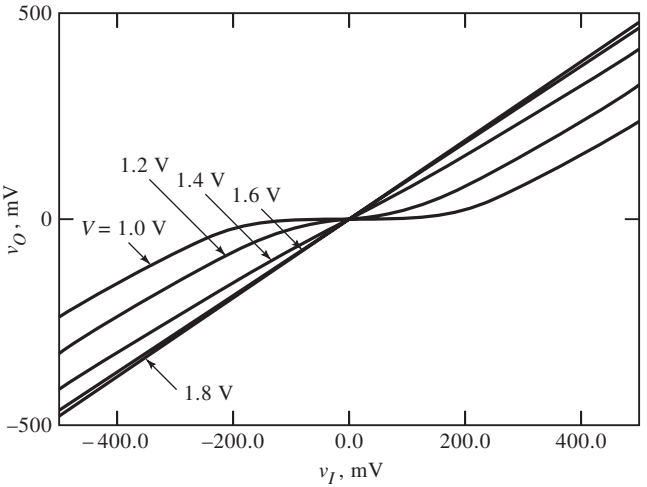
One of the simplest drivers is the CC configuration. Its near-unity voltage gain and a current gain of β make it an attractive output buffer stage. Its disadvantage is that it can provide current in only one direction; it is inherently unipolar. To achieve bipolar drive, a CC of the opposite polarity (a *complementary* CC), is connected in parallel.



Without the constant voltage sources (depicted as batteries), Q_1 conducts when v_I is positive and Q_2 when negative. Because of $V_{BE} \cong 0.6$ V for silicon, there is a deadband of reduced gain around zero output voltage where neither transistor conducts. The gain around zero is much less than the gain for $|v_I| > V_{BE}$. This gain variation results in a kind of nonlinearity, called *crossover distortion*, which can occur whenever the output drive "crosses over" from one transistor to another. The nonlinearity is due to the change in output resistance with output voltage. The voltage gain is

$$A_v = \frac{R_L}{R_L + r_{out}}$$

where r_{out} is the incremental (small-signal) output resistance. As v_O varies, r_{out} varies, and so does A_v.



The deadzone caused by V_{BE} can be narrowed and even eliminated by adding the base offset voltages. The transfer characteristics of $v_O(v_I)$ are shown above for V of 1.0 –1.8 V in 0.2 V steps and $R_L = 1$ kΩ. At $V = 1.8$ V, nonlinearity is not discernible from the graph. This base bias causes conduction of current I_0 through the two transistors. As v_O increases with v_I, some of I_0 is diverted into the load R_L. With further increase, all of I_0 is diverted from Q_2, and it approaches

cutoff, leaving Q_1 to drive the load alone. The emitter currents of Q_1 and Q_2 are found by applying KVL to the input loop:

$$V = V_T \ln\left(\frac{i_{E1}}{I_{S1}}\right) + V_T \ln\left(\frac{i_{E2}}{I_{S2}}\right) = V_T \ln\left(\frac{i_{E1} i_{E2}}{I_{S1} I_{S2}}\right)$$

Assuming $I_{S1} = I_{S2} = I_S$ and solving for the emitter current product,

$$i_{E1} \cdot i_{E2} = I_S^2 \cdot e^{(V/V_T)} \neq 0$$

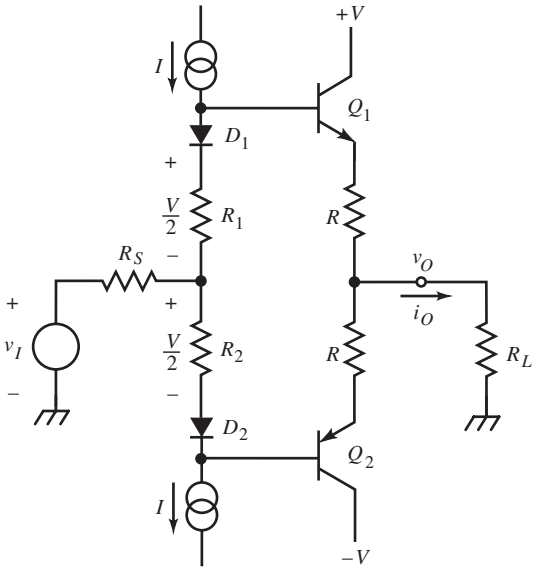
For a fixed V/V_T and I_S, neither emitter current goes to zero but only approaches it in the limit. This desirable property keeps both transistors on to reduce Δr_e and r_{out}. However, the ratio of emitter currents has such a wide range that r_{e1} and r_{e2} do also, and Δr_{out} is large.

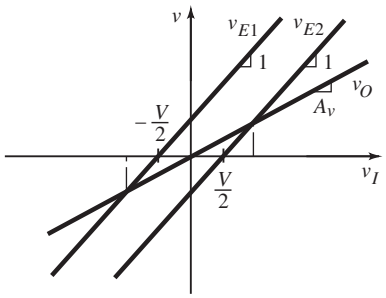
Another problem with this circuit is thermal instability. The i_E product varies with I_S and V_T. Both are temperature dependent. The dominant effect is the exponential variation of current with temperature due to V_T that can cause excessive I_0. This circuit is thermally unstable because increased I_0 causes increased junction heating leading to further increases in current. This phenomenon in BJTs is called *thermal runaway* and requires stabilization.

The improved circuit (below) shows the addition of emitter resistors R and the replacement of constant voltage sources with a more practical biasing network. The current sources cause voltage drops across the series resistors and diodes that bias the CCs. The diodes provide thermal tracking of the CC *b-e* junctions, and the resistors R_1 and R_2 are set to each drop $V/2$. With ideal junction compensation, diode voltages equal the CC v_{BE}, and $V/2$ appears stably across R in the emitters.

When I_0 is large and Δr_e is negligible relative to R, the cutoff points of the transistors are at the values of v_I where the voltage across R is zero. On the graph below, v_O is plotted along with the emitter voltages v_{E1} and v_{E2}. As v_I increases, v_{E2} increases at the same rate. Because $A_V < 1$, v_O increases at a more gradual rate, and the voltage across R of Q_2 decreases until it is zero; Q_2 is cut off. A similar argument applies to decreasing v_I and Q_1 cutoff. The input crossover voltages are

$$\text{crossover } v_I = \pm \frac{V/2}{1 - A_V}$$

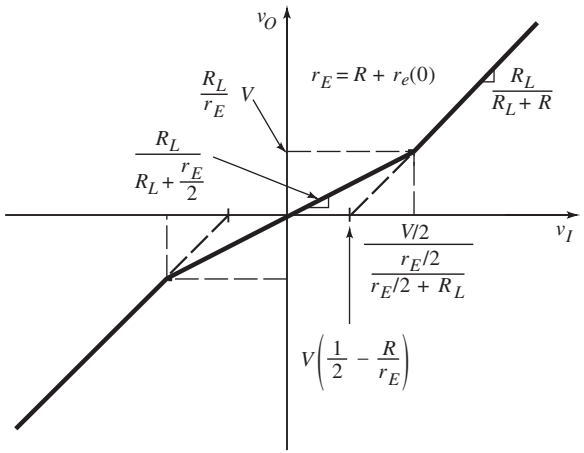




and

$$\text{crossover } v_O = \pm\left(\frac{A_v}{1-A_v}\right)\cdot\left(\frac{V}{2}\right)$$

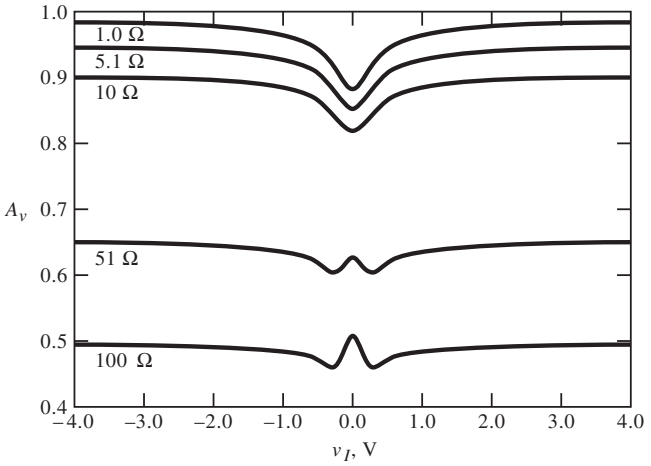
The transfer characteristic of the above plot is approached at relatively large values of I_0, where excessive power is dissipated due to biasing. At a reduced I_0, r_e variation is significant and the gain is not constant over the output range. The resulting transfer function is asymptotically approximated as shown below.



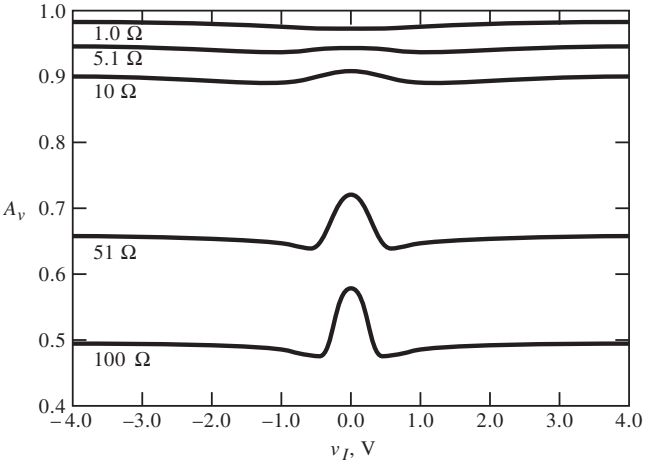
In the crossover region, the gain has been linearly approximated as the gain at the origin; the circuit is symmetrical there, with equal dynamic emitter resistances $r_e(0)$. Then r_{out} consists of two shunt paths, each with a dynamic resistance of $R + r_e$. Base resistance is referred to the emitter and included in R. Outside the crossover region, r_{out} is $R + r_e$ and approaches R with increasing v_O. The gain is asymptotic with

$$A_v \cong \begin{cases} \dfrac{R_L}{R_L + [R + r_e(0)]/2}, & \text{crossover region} \\ \dfrac{R_L}{R_L + R}, & \text{outside region} \end{cases}$$

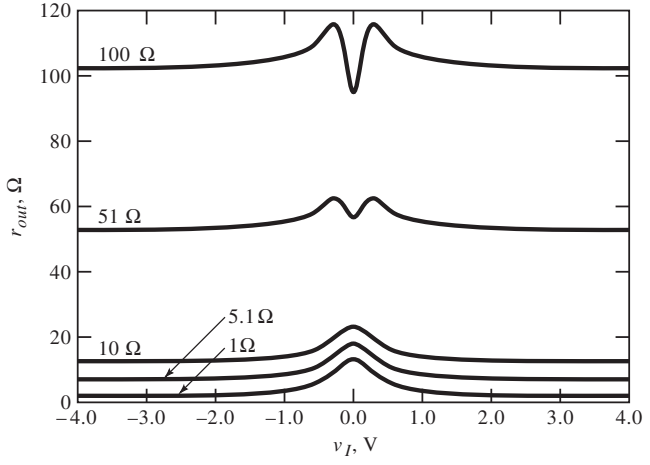
The effect of Δr_e on incremental gain over the amplifier range is shown below for $V = 0$ V.



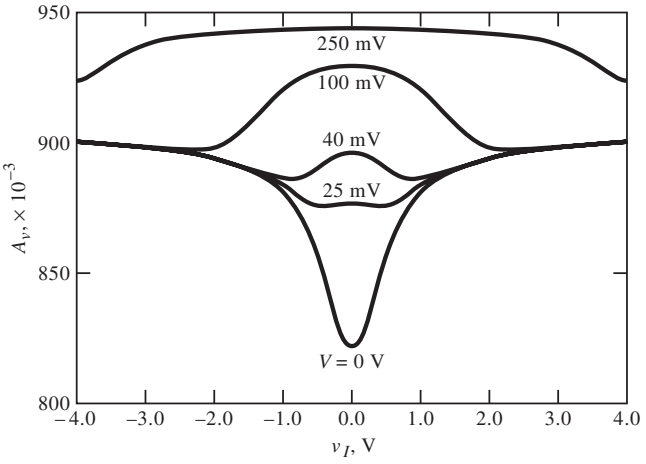
The smaller values of R show a decrease in gain around zero input due to the large increase in r_e . This is caused by a low value of I_0 resulting from inadequate V. For larger R, the gain begins to dip but then peaks at zero input. In these cases, r_e contributes a smaller fraction of the total r_{out} per side. When both sides conduct equally around zero, their shunt resistance is lowest and the gain peaks.



In the plot above, $V = 100$ mV, and this peak around zero is more pronounced since the shunting effect is significant over the wider crossover range. With greater I_0, r_e does not increase as much before the crossover region is entered.



The zero-bias output resistance is shown above. It affects the gain and varies inversely with it, as A_v predicts.



The plot above shows the gain with $R = 10\,\Omega$ and V as parameter. As V increases, the dip in gain around zero due to increased r_e begins to show the effect of the two sides shunting, the central peak. As shunting grows in dominance, the dip disappears and only the peak remains. The peak then broadens as the crossover range broadens with V.

The endpoints of the crossover region and the transfer curve offsets are symmetrical about the v_O axis and are found by a total-variable analysis. Given

$$i_{E1} = \frac{(v_I + V/2) - v_O}{R}, \quad i_{E2} = \frac{v_O - (v_I - V/2)}{R}$$

the endpoints are found by setting the emitter currents to zero, substituting the crossover-region gain for v_O and solving for v_I. The asymptotic crossover endpoints are

$$\text{crossover extremum of } v_I = \pm\left(\frac{R_L + r_E/2}{r_E/2}\right)\cdot\left(\frac{V}{2}\right), \quad r_E = R + r_e(0)$$

These correspond to asymptotic output values of

$$v_{Ox} = \text{crossover extremum of } v_O = \pm\left(\frac{R_L}{r_E}\right)\cdot V$$

The transfer function outside the crossover region has an input offset of

$$v_{Ix} = \pm\left(\frac{r_e(0) - R}{r_e(0) + R}\right)\cdot\left(\frac{V}{2}\right)$$

Because

$$r_e(0) = \frac{V_T}{I_0} = \frac{V_T}{((V/2)/R)} = \left(\frac{V_T}{V}\right)\cdot 2\cdot R$$

all of the values indicated in the above crossover plot can be expressed in R, R_L, V_T, and V. For example, from the previous equation for v_{Ix},

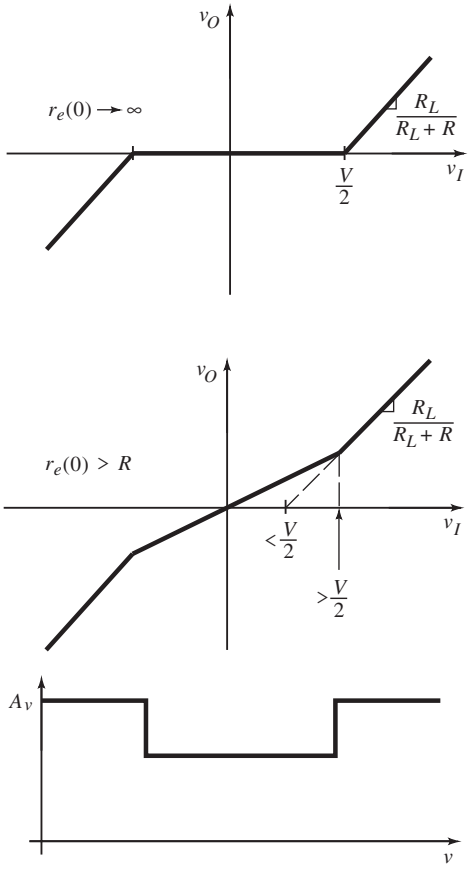
$$v_{Ix} = \pm\left(\frac{V}{2}\right)\cdot\left(\frac{2V_T - V}{2V_T + V}\right)$$

From the numerator, $v_{Ix} = 0\,\text{V}$ when $V = 2V_T$. This gives the most linear design:

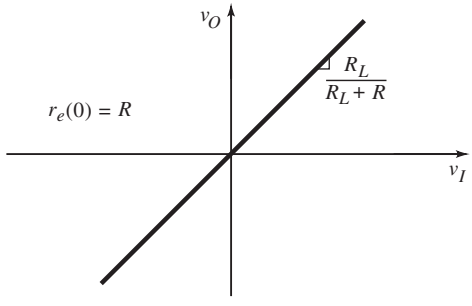
$$\text{optimum } V = 2V_T$$

V_T varies with temperature, and optimum performance over temperature requires that V track V_T.

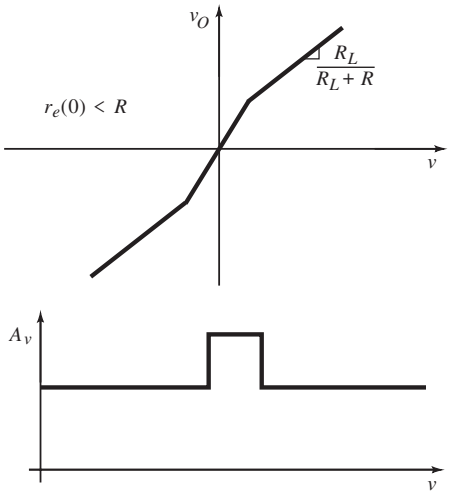
The effect of r_e (0) on v_O and gain is derived from the asymptotic crossover plot for several interesting cases. When $r_e(0)$ is very large, as shown on the top of the next page, a deadzone of V exists. As $r_e(0)$ decreases but remains larger than R, the effect of increasing r_e in the crossover region is shown with $A_v(v_I)$ on the second plot on the next page.



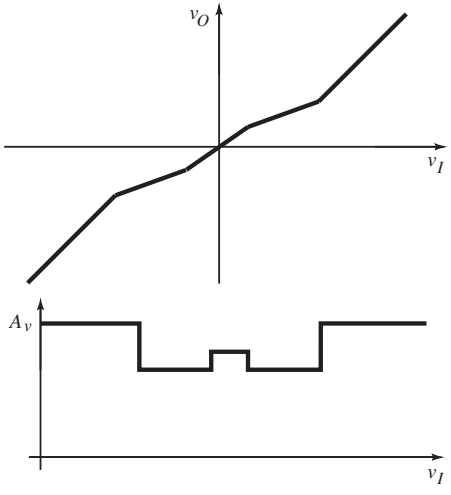
The dip in A_v is evident in previous figures when $r_e(0)$ is large relative to R. This occurs when V is small. As $r_e(0)$ increases, the optimum value is reached when $V = 2V_T$, shown below.



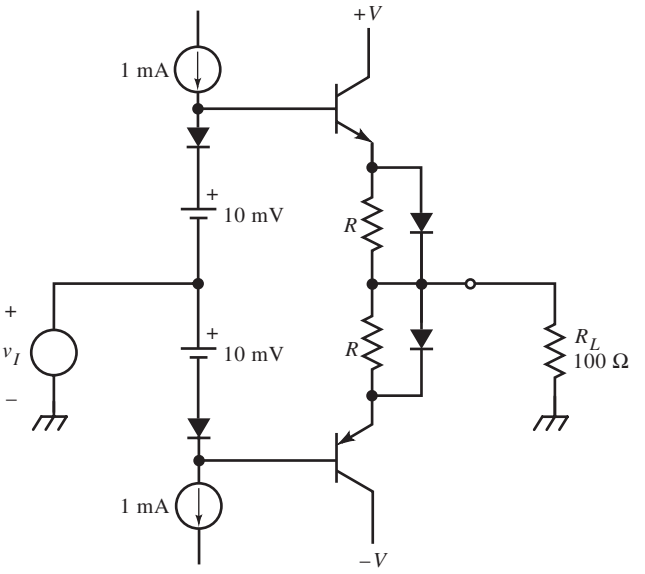
As $r_e(0)$ continues to decrease, the crossover gain is larger than that of the outside regions as shown on the top plot below. The plot of A_v below it shows the characteristic peak of previous gain plots due to the shunting effect. When $r_e(0)$ is zero, R dominates and $v_{Ix} = \pm V/2$.



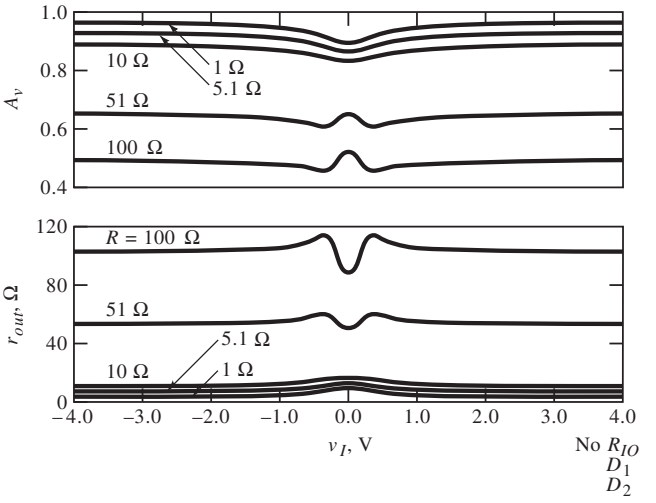
The small peak in a large dip is common in the $A_v(v_I)$ plots. The following plots are a piecewise-linear approximation of this phenomenon. Increasing r_e reduces gain in the outer part of the crossover region. At its center, shunting increases the gain and causes the peak.



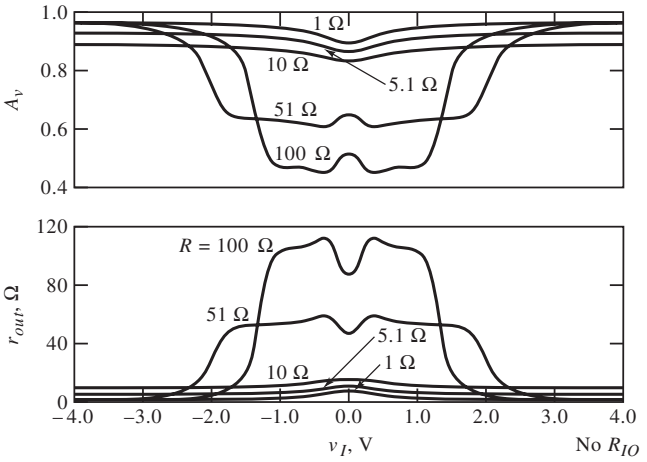
For high-current buffers, power dissipation in R can be reduced by diode shunts. As $|v_I|$ increases, the voltage across R increases until the diode conducts most of the current. In the crossover region, R dominates, but when the diodes dominate, the effective R decreases (when $r_d << R$) and the gain increases. The circuit shown below illustrates this.



Without the diodes, the plots below result.



With shunt emitter diodes, the plots are as shown below.

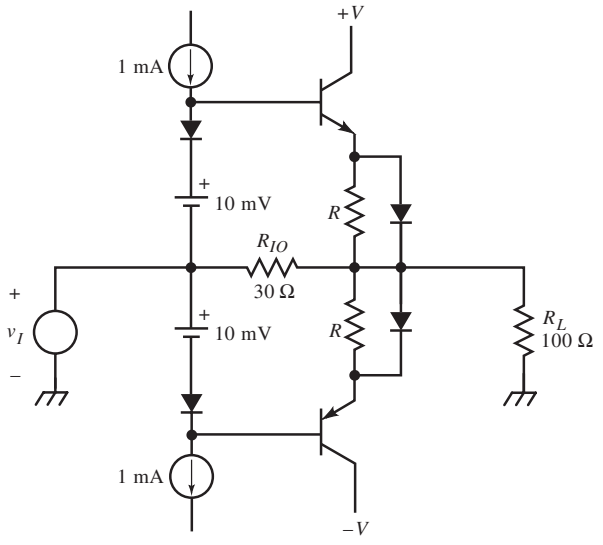


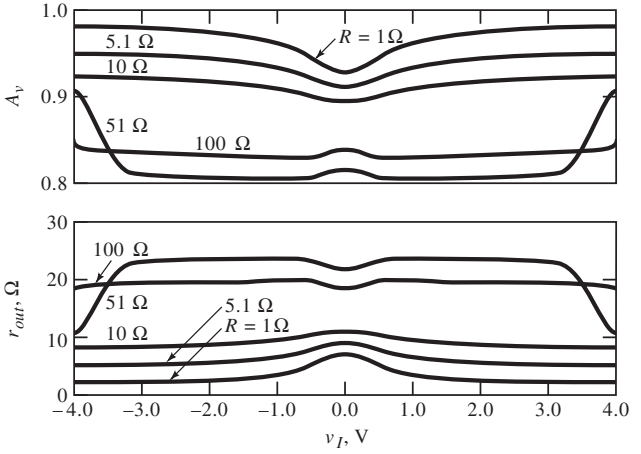
For the larger R, the diodes cause the gain to increase when they dominate conduction. Their dynamic resistances, r_d, are much less than R, and gain increases appreciably. The diodes thus improve gain accuracy outside the crossover region.

A second circuit modification is the addition of R_{IO}, as shown in the circuit on the next page. It reduces Δr_{out} when the input to the stage has a low resistance. In the crossover region, R_{IO} supplies load and bias current to the output from the buffer input source. Then $r_{out} < R_{IO}$ for the entire v_I range. This dramatic reduction in r_{out} is evident by comparing the above plot (no R_{IO}) with the one for the circuit with R_{IO}, on the next page.

When the complementary CC buffer is driven by a high-resistance source (such as the circuit of the example in *Designing Dynamic Circuit Response*, "Emitter-Follower High-Frequency Compensation" without R_B, shown below on page 193), the output resistance is also very high, from the high base resistance. In this case, the buffer behaves as a current amplifier, and the ratio of gains of the two sides is the ratio of the βs of their transistors. Matched-β transistors are then important for linearity.

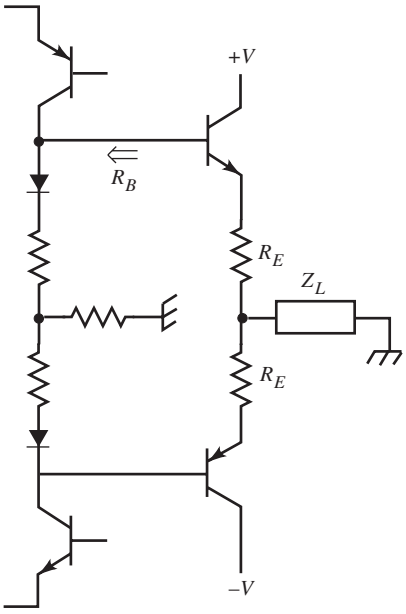
Feedback reduces both β-mismatch and crossover distortions by $1 + GH$. Because the nonlinear stage is a part of the loop, GH is affected by it. For the





minimum gain of the stage, the loop gain is minimal and distortion reduced least.

For the pathological case of $r_e(0)$ approaching infinity, the gain is zero in the deadband, and no amount of feedback improves linearity there. Too wide of a gain variation due to nonlinearity can also produce feedback instability in high-gain regions of the dynamic range. Reduction of loop gain required for stability



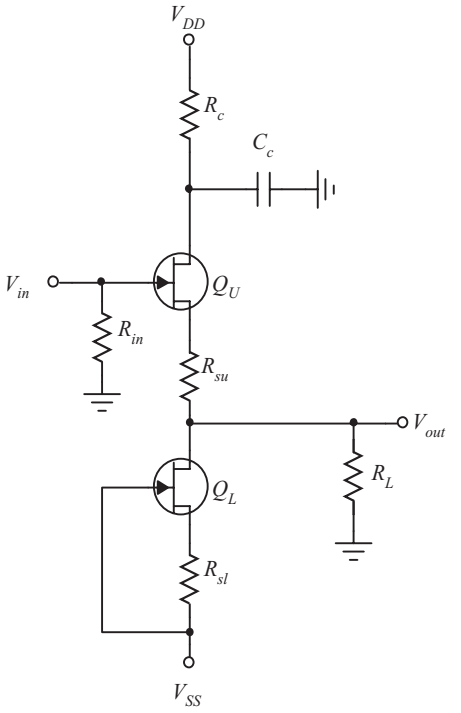
then compromises distortion reduction in the low-gain regions. Therefore, a good general strategy is to make the open-loop stages as linear as possible before closing the loop.

Buffer Amplifier Design

The ×1 voltage amplifier, or "buffer," is a standard building block of analog design. When high precision and minimal space are not worth the extra cost, a dual-JFET amplifier having low parts cost can deliver multiple hundreds of megahertz of bandwidth with an offset error of 10 mV or less and an offset drift of 10 μV/°C or less.

The design goal of the ×1 voltage amplifier is to achieve the ideal voltage amplifier: infinite input impedance, zero output impedance, and linearity. To achieve high input impedance, a JFET instead of a BJT is used, as shown in the buffer circuit below.

For low cost, discrete JFETs are chosen, although the higher-cost dual FETs, being monolithic, have better thermal tracking. Typical discrete n-channel JFET alternatives for the buffer are the 2N5484 through 2N5486. The 2N5485 has a



drain current selected for a nominal design value at midrange of I_{DSS} = 7 mA (I_{DSS} is I_D @ V_{GS} = 0 V) in a 4 to 10 mA specified range. Furthermore, the two JFETs are chosen to match using a curve tracer. This takes less than one minute per pair, by sorting them into matching I_{DSS} bins.

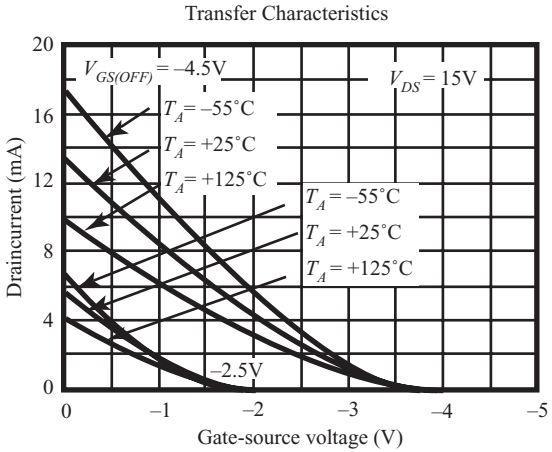
Next, supply voltages are chosen: V_{DD} = +12 V and V_{SS} = −5 V. These voltages are common in both desktop computers and instrumentation.

Offset Voltage

The first design feature of matched JFETs is the static (dc) tracking of the matched transistors. If the gate of the lower transistor, Q_L, is connected to its source, then V_{GS} = 0 V, and its drain current will be I_{DSS}. If this same current flows (with open load) through Q_U, then because it is matched, its V_{GS} is also zero and there is no voltage offset from input to output.

This nifty technique can be improved by setting the JFET operating point to the zero TC point instead, where thermal drift of V_{GS} with a given I_D is minimal

over temperature. For JFETs, the zero-drift V_{GS} is about 0.8 V above the pinch-off voltage. The value of this V_{GSZ} is where the I_D lines for various temperatures intersect. For the 2N5485 this is at about −1.2 V, about 0.8 V higher in value than the pinch-off voltage of around −2 V. The curves for the 2N5485 (Siliconix) are shown below.



Using these values, $R_{sl} = R_{su} = R_z = 1.2\text{ V}/5\text{ mA} = 240\ \Omega$, a 5% value. The voltage drop across R_{sl} is compensated in the signal path by a similar drop across the matched resistor, R_{su}. For better matching, these resistors can instead be 1% tolerance or better.

Thermal Distortion

With a varying input voltage, the power dissipation of the two JFETs will also vary. A change in power causes a change in silicon temperature, which results in thermally induced electrical noise, or "thermals," in the amplifier response. This "noise" is waveform-related and can better be regarded as thermal distortion. It can be minimized by setting the static operating conditions (or bias) for maximum power dissipation in the JFETs with no input (that is, at *quiescence*). The change in power (which we want to minimize) is least around the peak power value.

Let the quiescent bias current (or "standing current") of the JFETs be I_0. Then the power dissipated by upper and lower transistors is

$$p_u = \left(V_{DD} - v_L - (R_c + R_z)\cdot\left(I_0 + \frac{v_L}{R_L}\right)\right)\cdot\left(I_0 + \frac{v_L}{R_L}\right)$$
$$= V_{DD}\cdot I_0 + \left(\frac{V_{DD}}{R_L} - I_0\right)\cdot v_L - \frac{v_L^2}{R_L}$$

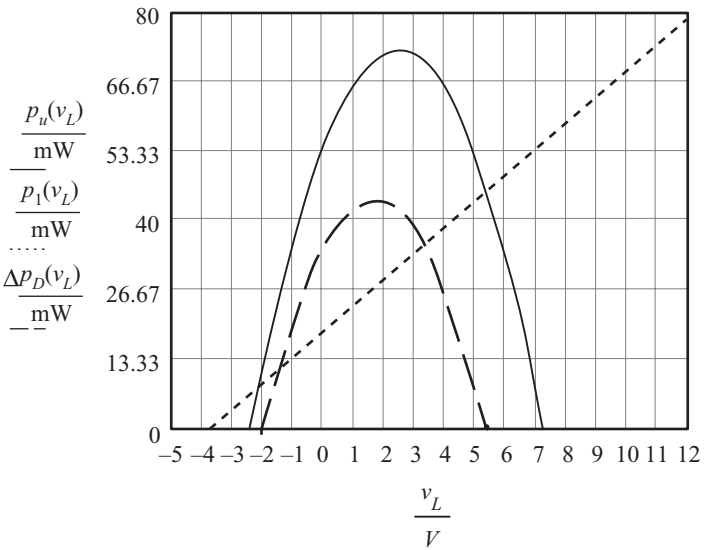
and

$$p_l = (v_L - V_{SS} - R_z\cdot I_0)\cdot I_0$$

where v_L is the load voltage (across R_L). The difference in power dissipation is

$$\Delta p_D = p_u - p_l$$

The power graphs are shown below.



At maximum Δp_D, the change in power with v_L is minimum, which is desired to minimize thermals. Differential power is maximum at

$$v_L(\max) = \frac{R_L}{(R_c + R_z) + R_L} \cdot \frac{V_{DD}}{2} - I_0 \cdot R_L$$

Another voltage of passing interest is where p_u and p_l are equal. Solving for v_L at $\Delta p_D = 0$ W,

$$v_{L0} = v_L(\max) \pm \sqrt{v_L(\max)^2 + \left(\frac{R_L}{R_c + R_z + R_L}\right) \cdot I_0 \cdot R_L \cdot (V_{DD} + V_{SS} - R_c \cdot I_0)}$$

On the plot, $v_{L0} = 5.3$ V. Though the power dissipation is matched at this output voltage, any change around this value causes a larger change in Δp_D than the same change in v_L around $v_L(\max)$. Consequently, the preferred bias point is at $v_L(\max)$.

On the plot, $v_L(\max) = 1.62$ V, but the given circuit parameters result in a quiescent v_L of 0 V instead. To adjust the quiescent voltages across the JFETs, an additional series resistor, R_c is added. For the general case, let the quiescent output voltage be V_L. Then to set the vertex of the differential-power parabola at V_L,

$$v_L(\max) = \frac{V_{DD}}{2} - R_L \cdot I_0 = V_L$$

and solve for the value of V_{DD} that will satisfy the desired condition:

$$V_{DD}(\max) = 2 \cdot (V_L + R_L \cdot I_0)$$

Then substitute V_L and the available supply V_{DD} into

$$R_c = \left(\frac{V_{DD}}{2 \cdot (V_L + R_L \cdot I_0)} - 1\right) \cdot R_L - R_z$$

For this design, $R_c = 490\ \Omega$. C_c bypasses R_C so that no appreciable voltage change occurs at the drain.

Matched-BJT Buffer Amplifier

JFETs are superior to BJTs in that they have high input resistance and low input bias current. However, for the same TC, current matching must be an order of magnitude better for FETs than BJTs. This is why the input offset specification of FET-input opamps is generally worse than their BJT counterparts. Put simply, BJTs match better than JFETs.

If a buffer does not call for extremely high input resistance, use BJTs instead. The biasing will have to be done somewhat differently, using a fixed base voltage for Q_L. This makes Q_L a current source, which drifts with temperature due to $V_{BE}(T)$. But a matched Q_U drifts similarly with the same bias current, and with a dynamic emitter resistance of

$$r_e = \frac{V_T}{|I_E|} = \frac{kT/q_e}{|I_E|} \cong \frac{26\,\text{mV}}{|I_E|}, 300\,\text{K}$$

It is kept constant with temperature as I_{CL} (= I_{EU}) varies with temperature. As temperature increases, V_{BEL} decreases and I_{CL} increases. At the same time, r_{eU} increases with thermal voltage, but increased emitter current compensates by decreasing r_{eU}. The TC of current from Q_L compensates for changes in r_e, which affects the buffer voltage gain.

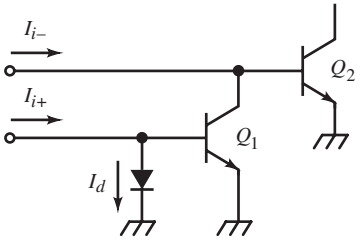
3

High-Performance Amplification

The basic conflict between speed and precision in amplifier design has led to the development of techniques that are both fast and precise; this high-performance amplification is applied as analog preprocessing for analog-to-digital converters (ADCs), as postprocessing for digital-to-analog converters (DACs), and for amplifiers in test instruments. We here examine novel subsystem-level amplifier topologies involving multiple signal paths first and then single-stage amplifiers. These topologies are applied to buffer amplifier design, continued from the previous chapter. Finally, the versatility of controlling or *programming* amplifier gain leads to a discussion of multipliers and programmable-gain amplifiers.

Current-Input and Current-Feedback Amplifiers

The transistor differential amplifier (diff-amp) amplifies differential input voltages, not currents. Because of the Miller effect and the dominance of stray capacitance over stray inductance, changing voltages are often more easily degraded than changing currents. It is therefore desirable to have op-amp topologies that sum currents instead of voltages at their inputs.



A current-differencing amplifier input stage, shown above, is commercially implemented in the National Semiconductor LM3900 (called a *Norton amplifier*). The noninverting input current I_{i+} flows mainly through the diode and develops a voltage across the *b-e* junction of Q_1. If the diode junction is matched with Q_1, then

$$I_d = I_{e1}$$

The diode current is

$$I_d = I_{i+} - I_{b1}$$

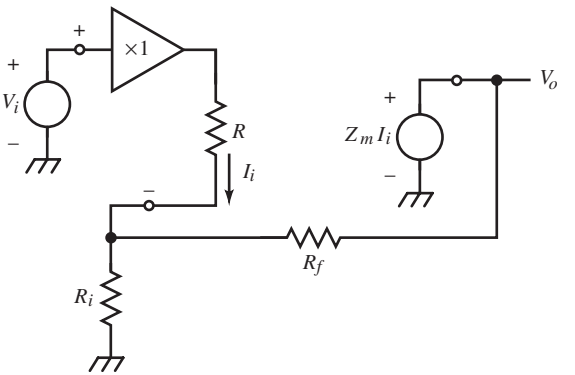
Noting that $I_{c1} = \alpha \cdot I_{e1}$, then

$$I_{c1} = \alpha \cdot I_d = \left(\frac{\beta}{\beta + 2} \right) \cdot I_{i+} \cong \alpha \cdot I_{i+}$$

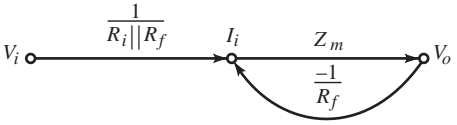
The differential input current is

$$I_i = I_{i+} - I_{i-} \cong -I_{b2} = I_{c1} - I_{i-}$$

Current differencing accuracy is limited by β of Q_1 and is compensated by decreasing the area of the diode relative to the Q_1 *b-e* junctions or by a better current-mirror topology.



The feedback amplifier above has a noninverting high-impedance voltage input and an inverting low-impedance current input. This strange combination has some advantages over feedback amplifiers with voltage-differencing inputs. The inverting input is, ideally, a voltage source $(R = 0)$. Its voltage follows the noninverting input. This eliminates the possibility of error-voltage summation at this node; summation of feedback current results in the error current I_i. For this reason, this topology is called a *current-feedback op-amp*. The forward-path transmittance is a transimpedance amplifier with input I_i and output V_o. From the op-amp terminals, the topology is identical to that of a conventional voltage-gain, noninverting op-amp, yet the dynamic response characteristics, both small- and large-signal, are quite different.



The amplifier flow graph is shown above. The input voltage, from the input buffer amplifier, causes error current I_i to flow through the Thevenin resistance $R_i \| R_f$ at the inverting terminal. Then I_i is amplified by the transimpedance Z_m to produce V_o. The feedback current is the current through R_f generated across the voltage difference $V_o - V_i$. Feedback analysis of the closed-loop voltage gain yields

$$A_v = T_i \cdot \frac{G}{1+GH} = \left(\frac{1}{R_i \| R_f}\right) \cdot \frac{Z_m}{1 + Z_m\left(1/R_f\right)} = \left(\frac{R_f}{R_i} + 1\right) \cdot \frac{Z_m}{Z_m + R_f}$$

The first factor is the conventional noninverting op-amp closed-loop gain. For $Z_m \gg R_f$, it is also the current-feedback amplifier closed-loop gain.

The frequency response is derived from A_v by assuming, as was done for the voltage-gain op-amp, a single-pole roll-off. Let the transimpedance be

$$G = Z_m = R_{mo} \cdot \frac{1}{s\tau_{bw} + 1}$$

Substituting into A_v, the closed-loop gain is

$$A_v(s) = \left(\frac{R_f}{R_i} + 1\right) \cdot \left(\frac{R_{mo}}{R_{mo} + R_f}\right) \cdot \frac{1}{s\left[R_f / \left(R_{mo} + R_f\right)\right] \cdot \tau_{bw} + 1}$$

This result is quite interesting. Unlike its conventional op-amp counterpart, the closed-loop bandwidth of this amplifier does not depend significantly on the closed-loop gain. (Compare with the single-pole A_v in "Feedback Circuit Response . . ." in *Designing Dynamic Circuit Response*. Also compare with ω_{bw}.) In effect, there is no gain-bandwidth product. Both gain and bandwidth can be set independently if the bandwidth is first set with R_f and then the gain with R_i. Ideally, for infinite R_{mo}, the gain is exactly that of the conventional noninverting op-amp formula, and the closed-loop bandwidth is infinite. Bandwidth actually increases with increasing forward-path gain. In practice, τ_{bw} also depends on R_{mo} because the transimpedance amplifier itself has a gain-bandwidth product.

The bandwidth is independent of the gain only when the input buffer is ideal. With output resistance R, the inverting input is not constrained to be V_i, and a voltage-feedback interpretation could be given. Keeping with current feedback, we can modify the transmittances to account for R:

$$T_i = \frac{1}{R + R_i \| R_f}, \quad H = -\left(\frac{R_i}{R_f + R_i}\right) \cdot \left(\frac{1}{R + R_i \| R_f}\right)$$

G remains the same. These transmittances are found by Nortonizing the external feedback circuit to the inverting input and then solving for the current dividers in each direction. The resulting closed-loop gain is

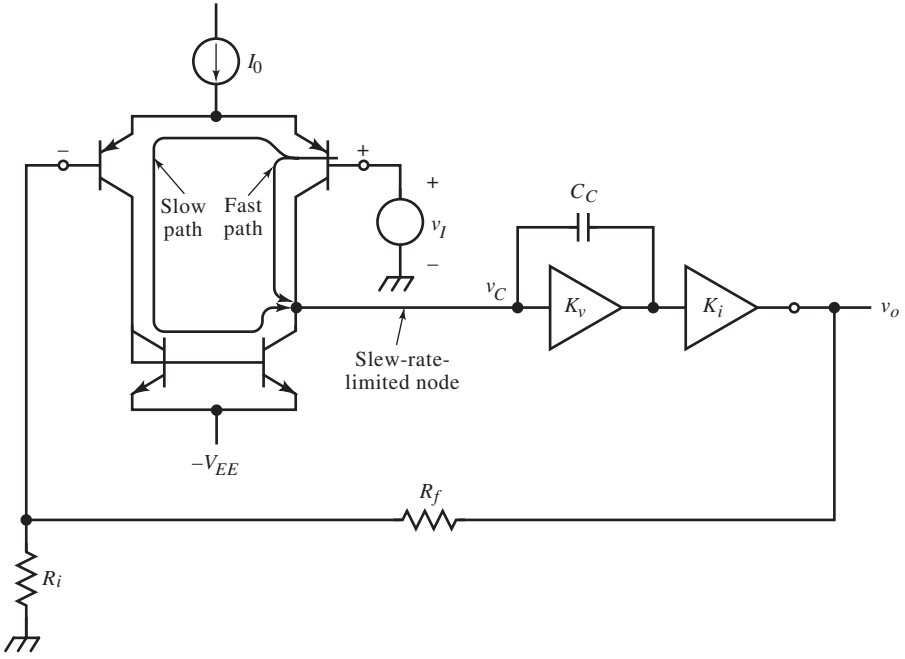
$$A_v = A_{vc} \cdot \left(\frac{R_{mo}}{R_{mo} + R_f + R \cdot A_{vc}}\right) \cdot \frac{1}{s\tau_{bw}\left(R_f + R \cdot A_{vc}\right) / \left(R_{mo} + R_f + R \cdot A_{vc}\right) + 1}$$

where

$$A_{vc} = \frac{R_f + R_i}{R_i}$$

R is effectively increased by the closed-loop gain A_{vc} and also has the effect of adding to R_f. As A_{vc} increases, the time constant in A_v approaches the open-loop

value of τ_{bw}. In good design, $R_{mo} \gg R_f$; therefore, the effective increase in R_f due to RA_{vc} only slightly decreases bandwidth. The above gain equation can be used to determine more precise values of R_f and R_i than the more ideal $A_v(s)$ equation, given R.



Current-feedback amplifiers also have a large-signal advantage over voltage op-amps. To show the contrast, first examine the conventional op-amp, shown above, with three stages. (See "Two-Pole Compensation" in *Designing Dynamic Circuit Response* for background.) The first stage has a transconductance of $G_m = 1/2r_e$ when balanced, loaded by the high-gain second stage with compensation capacitor C_C. This small on-chip capacitance is multiplied by the Miller effect and effects dominant single-pole compensation. The transconductance stage can supply, at most, I_0 to charge the second-stage input capacitance C_i. For large K_v,

$$C_i = (K_v + 1) \cdot C_C \cong K_v \cdot C_C, \quad K_v \gg 1$$

With a step input, the voltage, Δv_C, becomes slew-rate limited as Δv_I increases. The output current of the first stage is

$$i_O = G_m \cdot v_l$$

A maximum i_O of I_0 causes the output voltage to change at the maximum rate of

$$\text{slew rate of } v_O = \frac{dv_O}{dt} = K_v \cdot \frac{dv_C}{dt} = K_v \cdot \frac{I_c}{C_i} \cong \frac{I_0}{C_C}$$

The unity-gain frequency f_T corresponds to a time constant τ_T determined by the compensation. The small-signal exponential response to an input step has a maximum slope of

$$\max \frac{dv_O}{dt} = \frac{\Delta v_O}{\tau_{bw}} = \frac{\Delta v_I}{\tau_T}$$

At the onset of slewing, the maximum incremental step is

$$\max \Delta v_I = \frac{I_0}{C_C} \cdot \tau_T$$

The small-signal dynamic range of the input stage is limited to

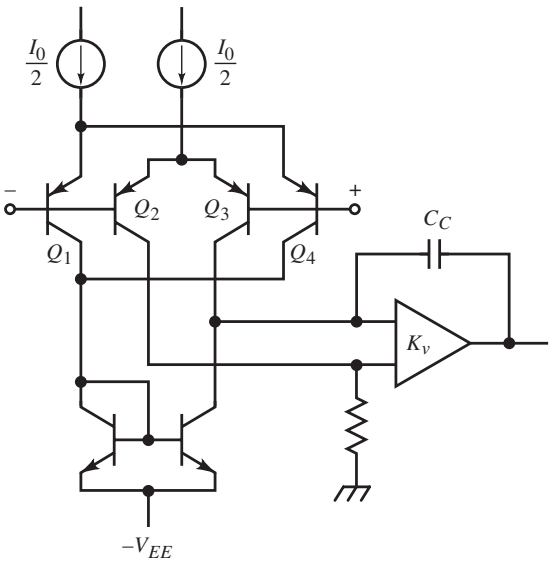
$$\max \Delta v_I = \frac{I_0}{G_m} = I_0 \cdot (2 \cdot r_e) = I_0 \cdot \left(2 \cdot \frac{V_T}{I_0/2} \right) = 4V_T \cong 100 \text{ mV}$$

Then τ_T is determined by equating these two equations and solving

$$\tau_T = \frac{4 \cdot V_T \cdot C_C}{I_0} \Rightarrow f_T = \frac{I_0}{8\pi \cdot V_T \cdot C_C}$$

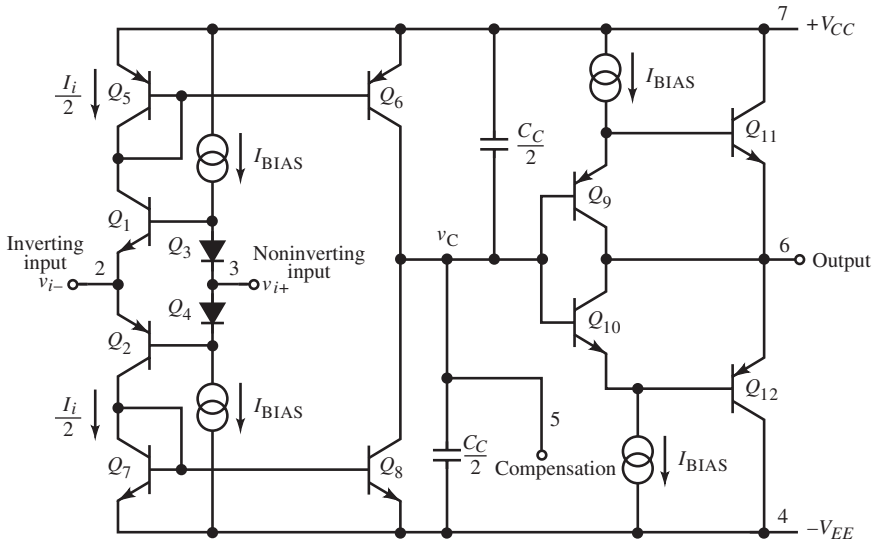
The small-signal dynamic range at the input can be increased by decreasing G_m. This can be achieved by adding R_E to the emitter circuit. In addition, the small-signal response of the input stage is complicated by its two signal paths.

In the noninverting configuration shown above, v_i travels directly through the input transistor to the second stage. This is the direct path and is faster than the path through the other diff-amp transistor and the current mirror. The C_{be} of the mirror transistors adds an additional pole to this route, leading to a combined response in v_c.



This split-path response is eliminated in the above circuit. The additional shunt diff-amp at the input provides differential drive to the second stage. It also drives the current mirror with a common-mode static current, the effect of canceling inverted waveforms. The resulting current biases the mirror. No gain is lost due to elimination of the slow path because the second stage is driven differentially. The LF400 junction field-effect transistor (JFET)-input amplifier uses this technique.

Another "transconductance spoiler" uses diff-amp transistors with split collectors. The large-area collectors are connected to the supply and shunt most of the dynamic current. In effect, the G_m is reduced by collector current-divider behavior. An example of the use of this technique is the National Semiconductor LM346 programmable op-amp.



The large-signal behavior of current-feedback amplifiers does not have this slew-rate limitation on input dynamic range. The figure above shows simplified circuitry of a typical current-feedback amplifier, the Analog Devices AD846. The inverting input connects to the output of a buffered complementary common-collector (CC) stage driven by the noninverting input. The CC collector currents are a differential output of the input current difference at v_{I-}. The first stage of the transimpedance amplifier is a differential current mirror – one current mirror per side – that drives another complementary CC buffer. The dominant pole is determined at the high-resistance input node of this buffer, with dominant-pole capacitance C_C split between the power-supply "rails."

Slewing does not occur at v_C for a wide range of input voltages because a larger Δv_I produces a larger current into the transimpedance amplifier. This current is generated by Δv_I across the resistance driven by the inverting-input buffer,

$$r_{in-} = R + R_i \| R_f$$

The output slew rate is

$$\max \frac{dv_O}{dt} = \frac{i_C}{C_C} = \frac{(\Delta v_I / r_{in-})}{C_C} = \frac{\Delta v_I}{r_{in-} C_C} = \frac{\Delta v_O}{K_v r_{in-} C_C}$$

For $R = 0\ \Omega$, this reduces to

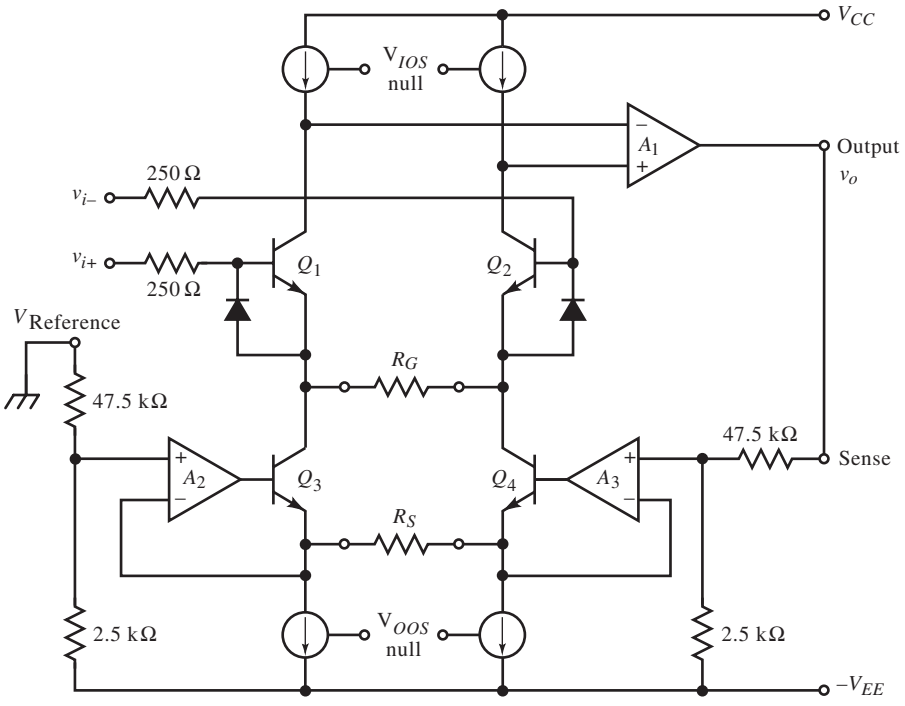
$$\max \left. \frac{dv_O}{dt} \right|_{R=0} = \frac{\Delta v_O}{R_f C_C}$$

and

$$\tau_{bw} = R_f \cdot C_C$$

In contrast to the conventional amplifier τ_T, no dependence on large-signal parameters appears; the current-feedback amplifier is free of the slew-rate limitations of conventional op-amps.

Although the inverting input of a current-feedback amplifier has low open-loop impedance, the error quantity – the terminal current I_i – is nulled by feedback to a low value. Consequently, current-feedback op-amps can be used with the same external circuits as voltage-feedback op-amps.



The current-feedback concept also applies to instrumentation amplifiers. The figure above shows the Precision Monolithics AMP-01 simplified topology, a typical current-feedback instrumentation amplifier (IA). This topology is different from the current-feedback op-amp because it has a voltage diff-amp input and a voltage-divider feedback path. The divider is followed by a voltage-to-current (V/I) converter consisting of a BJT diff-amp with emitter resistor R_S. The error current is generated by the cascoded diff-amps and in feedback nomenclature is

$$i_E = i_i - i_B = \frac{v_i}{R_G} - \left(\frac{R_i}{R_f + R_i}\right) \cdot \frac{v_o}{R_S}$$

The output voltage is

$$v_o = R_m \cdot i_E$$

where R_m is the transmittance of the transimpedance amplifier. Combining these equations and solving for the closed-loop voltage gain,

$$A_v = \left(\frac{R_f + R_i}{R_i}\right) \cdot \left(\frac{R_S}{R_G}\right) \cdot \frac{1}{1 + \left[(R_f + R_i)/R_i\right](R_S/R_{in})} = A_{vo} \cdot \left(\frac{R_S}{R_G}\right) \cdot \frac{1}{1 + A_{vo} \cdot (R_S/R_m)}$$

Ideally, $R_m \gg R_S$, and the gain becomes

$$A_v|_{R_m \to \infty} = \left(\frac{R_f + R_i}{R_i}\right) \cdot \left(\frac{R_S}{R_G}\right)$$

The error current is generated as emitter current in the input diff-amp transistors. An increase in v_{i+} causes Q_1 to conduct more, resulting in an increase of v_o. Through the feedback network, Q_4 is made to conduct more and Q_3 less. The error current for $i_{e1} = v_i/R_G - i_{c3}$, where the first term is i_i. A similar expression for i_{e2} can be written. Current-feedback IAs have the same basic properties as current-feedback op-amps; bandwidth remains relatively independent of gain.

The AMP-01 has two adjustments to correct offset error. The IA requirement for two adjustments is somewhat different from that of a two-transistor diff-amp.

At the output, the total offset voltage is the sum of gain-dependent (input) and gain-independent (output) offsets:

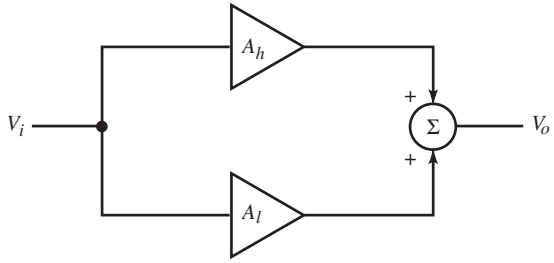
$$V_{OS} = A_v \cdot V_{IOS} + V_{OOS}$$

A similar equation holds for the offset voltage temperature coefficients (TCs). The input offset voltage adjustment varies the current ratio of the collector current sources at the output of the first stage. With R_G shorted, the offset voltage of Q_1 and Q_2 is most sensitive to their emitter current ratio. With R_G unshorted, the output offset voltage is adjusted by the emitter current ratio of Q_3 and Q_4 since they are part of the output feedback path. That is, the gain from output to R_S is $1/A_{vc}$, whereas the gain from the collector current sources to the output – the transimpedance amplifier gain – is very large.

Split-Path, Low-Frequency Feedback, and Feedbeside Amplifiers

Designing Amplifier Circuits, "CB Amplifier with r_o", expounds the effect of r_o in single-stage BJT amplifiers and the passive forward path through feedback networks. These paths are largely unintentional and generally degrade performance.

Amplifiers designed to benefit from multiple paths are called *composite* amplifiers. Various topologies at the subsystem level are possible. The *split-path amplifier* shown below purposely has parallel paths to improve performance.



Because fast amplifiers often lack good static characteristics, the idea is to combine the best of both in one amplifier. A common strategy is to shunt a fast amplifier with a low-speed, precision amplifier. The two paths are combined at the output.

A requirement for flat frequency response is that the fast and slow paths have complementary gains in the crossover frequency range so that their sum is constant. The crossover gain must also be the same as the low- and high-frequency gains. A simple design strategy is to let the slow path have a single-pole response:

$$A_l = K \cdot \left(\frac{1}{s\tau_l + 1} \right)$$

The fast path is a high-frequency (hf) amplifier (no static gain) that rolls up in gain as A_l rolls off. At a high frequency of $1/\tau_h$, it also rolls off:

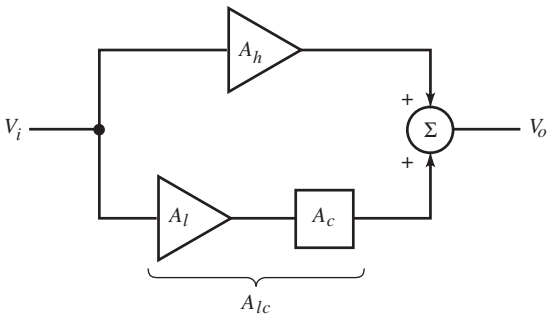
$$A_h = K \cdot \left(\frac{s\tau_l}{s\tau_l + 1} \right) \cdot \left(\frac{1}{s\tau_h + 1} \right)$$

The composite gain is

$$A = A_l + A_h = K \cdot \left(\frac{1}{s\tau_h + 1} \right) \cdot \left(\frac{s(\tau_l + \tau_h) + 1}{s\tau_l + 1} \right)$$

With $\tau_h = 0$, the two paths combine to give a flat response. To approach this ideal, the crossover frequency $1/\tau_l$ must be much less than the fast path bandwidth of $1/\tau_h$.

This constraint to flat response can be eliminated by adding a compensation stage A_c in the slow path.



It compensates for τ_h of the fast path and is

$$A_c = \frac{1}{s\tau_h + 1}$$

The modified slow-path response is thus

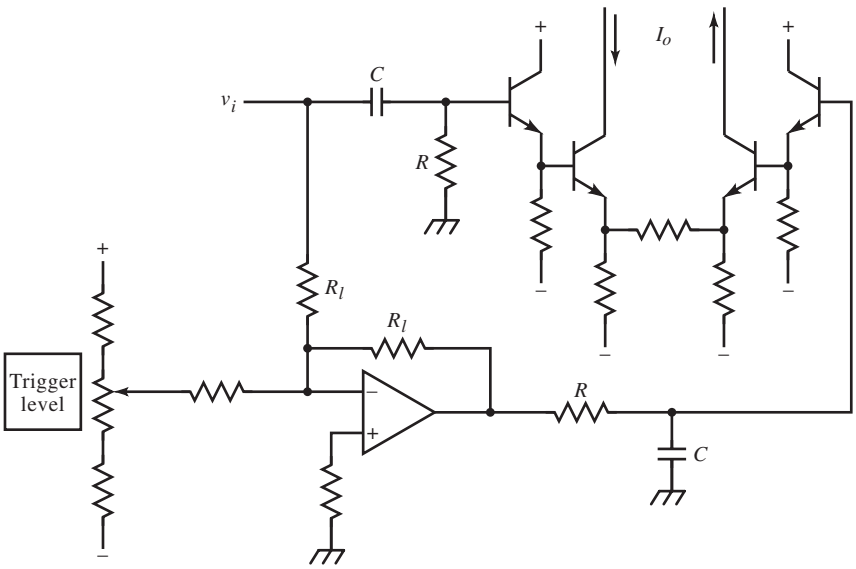
$$A_{lc} = K \cdot \left(\frac{1}{s\tau_l + 1}\right) \cdot \left(\frac{1}{s\tau_h + 1}\right)$$

Now the composite gain is

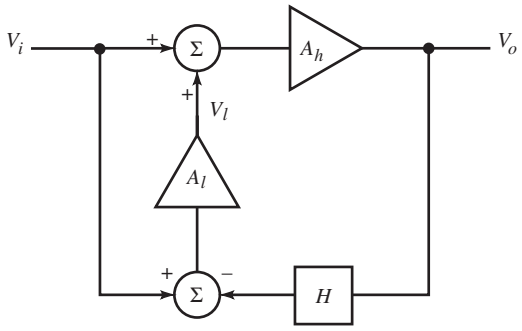
$$A = A_{lc} + A_h = K \cdot \left(\frac{1}{s\tau_h + 1}\right)$$

and the crossover frequency $1/\tau_l$ is independent of fast-path bandwidth.

Example: Split-Path Composite Trigger Amplifier



This amplifier is typical of the input stage of an oscilloscope trigger generator. Various sources are selected for v_i, and either low-frequency (lf) or hf paths can be turned off for "lf reject" or "hf reject" trigger functions. The single-ended v_i is converted to a differential signal for the diff-amp input by the inverting op-amp in the slow path. The trigger level control is summed in the slow path. For good dynamic response, the fast-path *RC* differentiator and slow-path *RC* integrator time constants must be equal. The low-pass pole of the slow path could be realized by placing a capacitor around the slow-path op-amp, but its inability to reject high frequencies makes this capacitor an undesirable hf feedthrough. A passive *RC* integrator following the op-amp is a broadband low-pass filter.



Another composite amplifier topology that is intended to accomplish the same design objective is the *low-frequency feedback* topology shown above. Low-frequency correction is made at the input to the fast amplifier by adding to V_i the error quantity,

$$V_l = A_l \cdot (V_i - H \cdot V_o)$$

The output is

$$V_o = A_h \cdot (V_i + V_l)$$

Combining these equations yields the composite gain

$$A = A_h \cdot \frac{1 + A_l}{1 + HA_hA_l} = \frac{A_h}{1 + HA_hA_l} + \frac{A_hA_l}{1 + HA_hA_l}$$

$$\uparrow \qquad \uparrow$$

fast path slow path

Both paths benefit from feedback, with a loop gain of $H \cdot A_h \cdot A_l$. For large A_l at low frequencies, the loop gain is large and static characteristics are improved over those of A_h alone.

To frequency-compensate this amplifier, assume that A_l rolls off with frequency, so that

$$\lim_{f \to \infty} A_l = 0$$

Consequently,

$$\lim_{f \to \infty} H \cdot A_h \cdot A_l = 0$$

and the fast-path response is

$$\lim_{f \to \infty} \frac{A_h}{1 + H \cdot A_h \cdot A_l} = A_h$$

The slow-path response is

$$\lim_{f \to \infty} \frac{A_h \cdot A_l}{1 + H \cdot A_h \cdot A_l} = 0$$

At high frequencies, only the fast path contributes A_h to the gain. If we also assume that A_l is an op-amp with infinite static gain, then, the fast-path response at 0 Hz is

$$\lim_{f \to 0} \frac{A_h}{1 + H \cdot A_h \cdot A_l} = 0$$

and for the slow path,

$$\lim_{f\to 0}\frac{A_h\cdot A_l}{1+H\cdot A_h\cdot A_l}=\lim_{f\to 0}\frac{A_h}{1/A_l+H\cdot A_h}=\frac{1}{H}$$

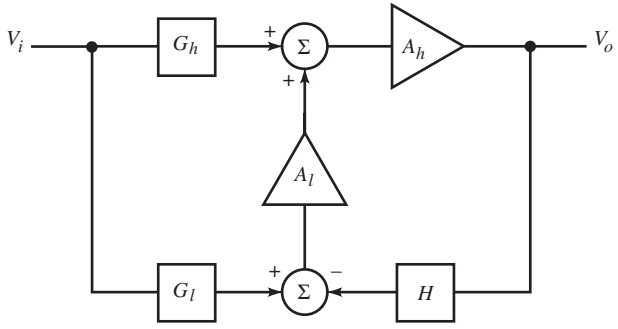
Then $A = 1/H$ at 0 Hz. This is the same gain as for the inverting op-amp configuration. For flat response, the gain at 0 Hz must be the same as at high frequencies, or

$$\frac{1}{H}=A_h\Rightarrow H=\frac{1}{A}$$

This condition is necessary but not sufficient for flat response; the mid-frequency range might not be flat. In A, the $(1 + A_l)$ factor in the numerator cancels the denominator only under this condition and ensures wideband flatness. The compensated gain is

$$A=A_h,\quad H=A_h^{-1}$$

This topology has the practical disadvantage that a slow amplifier A_l is driven by a fast signal V_i. Unless the feedback compensation is correct, A_l responds to a fast error voltage but is unable to follow it. The effect is a low-frequency response anomaly; the cause, however, is that a slow amplifier is responding to a fast input. In addition, because of loop delay (or phase lag) through A_h and H, the output of H cannot match V_i in phase. At frequencies for which the delay is significant, this error quantity becomes large.



The more general low-frequency feedback topology is shown above. Its voltage gain is

$$A = A_h \cdot \frac{G_h + G_l \cdot A_l}{1 + H \cdot A_h \cdot A_l}$$

The fast path is through G_h and A_h and is represented by the first term; the slow path is through G_l, A_l, and then A_h, which are factors of the second term. The feedback loop contains the blocks represented in the denominator. For $G_l = H \cdot G_h \cdot A_h$, $A = A_h \cdot G_h$ at low and high frequencies.

A special case has $H = G_h = 1$. It has the same problem with high-frequency input to the slow path as that of the previous topology, but coming from V_o. For it, the composite gain is

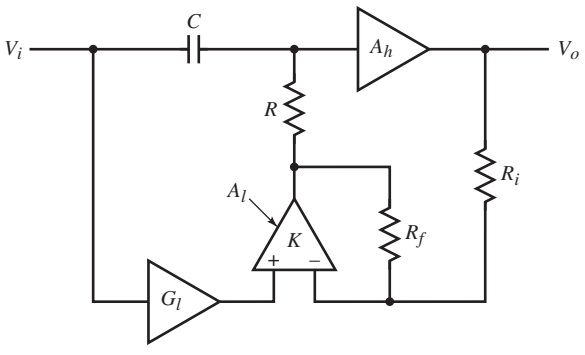
$$A = \frac{A_h \cdot (1 + G_l \cdot A_l)}{1 + A_h \cdot A_l} = \frac{A_h}{1 + A_h \cdot A_l} + \frac{G_l \cdot A_h \cdot A_l}{1 + A_h \cdot A_l}$$

With the compensation criterion of this amplifier, the gain is

$$A = A_h, \quad G_l = A_h$$

This topology has the advantage that a matching A_h can be used to provide tracking compensation of the fast path.

Example: Low-Frequency Feedback Composite Amplifier



This amplifier design is based on the following reasoning. Because the static characteristics of A_h are poor, the hf input is coupled to it through an RC differentiator which adds op-amp feedback from A_l at its input. The G_l block of the general lf feedback topology is set according to the compensation condition and A_l is made differential. Will the amplifier provide a flat frequency response?

For this amplifier, the corresponding general lf-feedback topology blocks are as follows:

$$G_l = A_h \cdot \left(\frac{R_f}{R_f + R_i} \right), \quad G_h = \frac{sRC}{sRC + 1}$$

Then

$$A = A_h \cdot \left(\frac{G_h + G_l \cdot A_l}{1 + G_l \cdot A_l} \right), \quad G_l \cdot A_l = A_h \cdot H \cdot A_l = A_h \cdot \left(\frac{R_f}{R_i} \right);$$
$$A_l = \left(\frac{R_f}{R_i} + 1 \right), \quad K \to \infty$$

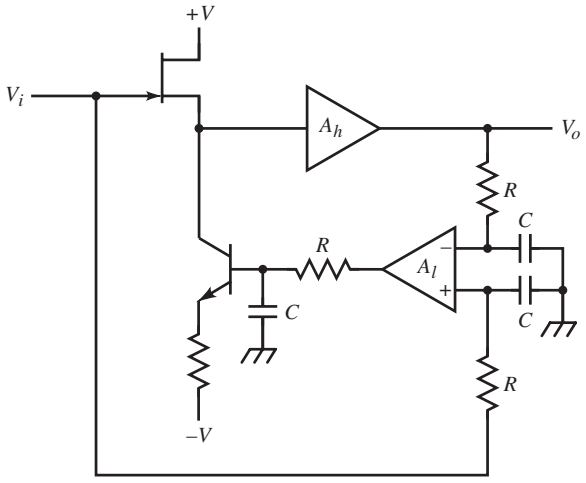
Substituting for G_h,

$$A = A_h \cdot \left(\frac{G_l \cdot A_l}{1 + G_l \cdot A_l} \right) \cdot \frac{sRC \cdot \left(\frac{1 + G_l \cdot A_l}{G_l \cdot A_l} \right) + 1}{sRC + 1}$$

and has a flat response only for $G_l \cdot A_l \gg 1$

Example: Low-Frequency Feedback Input Buffer

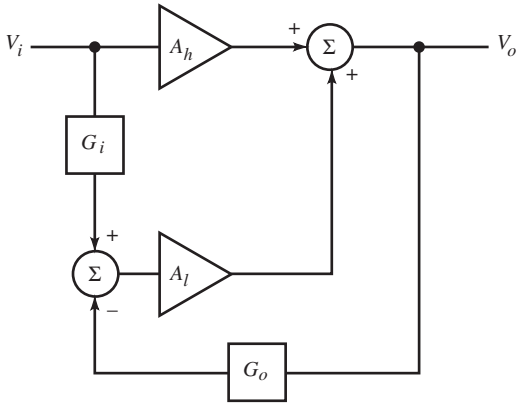
The buffer amplifier has a high-impedance input because of the field-effect transistor (FET) common-drain (CD) input stage. The FET has voltage offset and drift that is corrected by a slow path that varies the FET current to achieve zero static offset voltage.



This amplifier is represented by the general low-frequency feedback topology and A. Because it is a buffer (×1 gain), $A_h \cong 1$. From the circuit diagram, the FET CD is $G_h = 1$, and $G_l = H$. Both G_l and H are RC integrators. A_l is thus driven only by low-frequency waveforms. Substituting into A gives

$$A = A_h \cdot \frac{1 + G_l \cdot A_l}{1 + G_l \cdot A_l \cdot A_h}$$

For $A_h = 1$, the fraction is one, and the gain is A_h; otherwise, gain error in A_h contributes to compensation error. For large A_l, the fraction approaches $1/A_h$ and A approaches one. The RC time constant is set well within the bandwidth of A_l. This buffer eliminates the need for matched high-frequency FETs.



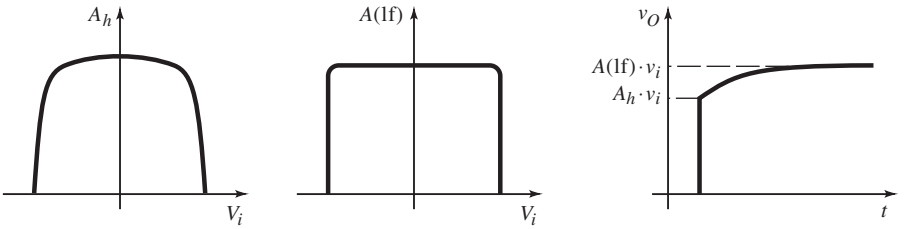
The two concepts of split-path and low-frequency feedback are combined in the amplifier shown above. Its gain is

$$A = \frac{A_h + G_i \cdot A_l}{1 + G_o \cdot A_l}$$

Factoring A_h from the numerator results in the gain A_h under the condition:

$$A = A_h, \quad G_i = G_o \cdot A_h$$

If G_o is a low-pass filter, then the input of A_l is always low in frequency. This eliminates high-frequency rectification at the inputs due to op-amp inability to respond quickly enough. It also keeps A_l from responding slowly to fast inputs.



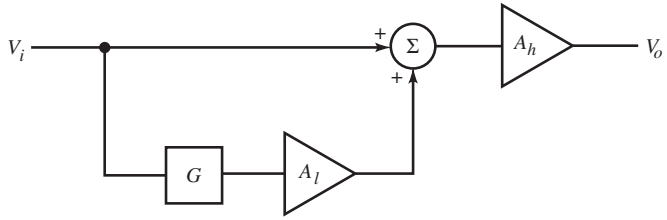
(a) (b) (c)

Flat frequency response is not the only criterion of precision. Fast amplifiers typically have poorer linearity and more thermal drift and distortion than low-frequency amplifiers. Nonlinearity is greatest at the extremes of the dynamic range, as in (a) above.

Usually, gain decreases at the extremes of V_i. The amplifier has constant gain over its dynamic range at low frequencies, as in (b), due to feedback. The step response is shown in (c); the gain of A_h is less than the low-frequency gain of A. As the slow path begins to respond, the gain increases.

To correct for this nonlinearity, A_h can be linearized. Also, temperature effects of A_h can be compensated. However, it is better in wideband amplifiers to mini-

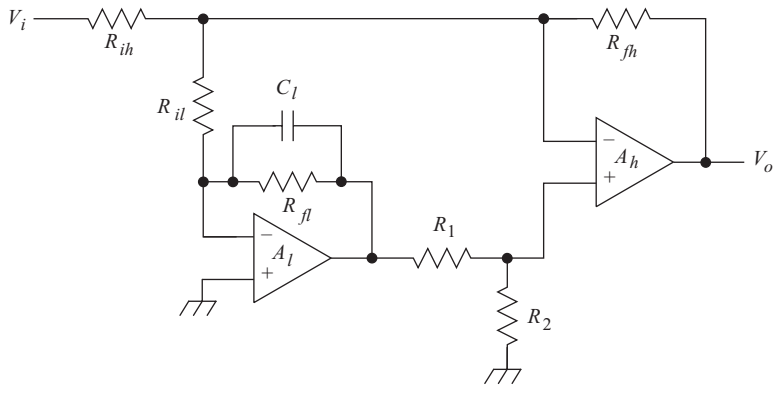
mize the complexity of the compensation networks at high-frequency nodes due to their additional stray reactance. A complicated slow-path linearity or thermal correction network avoids this problem.



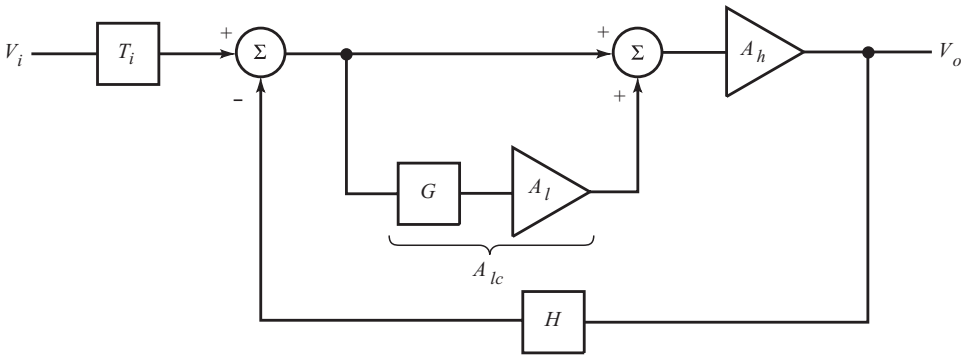
One such scheme, shown above, sums at the input to A_h. The input also drives a slow compensation path that corrects for imperfections in A_h. This scheme, called *feedbeside*, is open-loop. Its compensation network parameters are adjusted to correct for low-frequency imperfections in A_h. Its gain is

$$A = A_h \cdot (1 + G \cdot A_l)$$

where G is a passive, attenuating compensation network. A_l provides scaling and inversion, if needed, so that $G \cdot A_l$ matches the reciprocal of the error in A_h. The feedbeside path can itself be a split-path amplifier, in which each path independently compensates for an anomaly in A_h.



An amplifier involving feedbeside is shown above. A fast op-amp A_h requires input offset-voltage correction provided by the slower, more precise A_l. C_l limits the bandwidth of the feedbeside path to near 0 Hz; R_1 and R_2 reduce the gain of the slow path.



This amplifier has the block diagram shown above and is not, overall, a feedbeside topology because the slow path is within the fast-path feedback loop. The general form of the gain found from the block diagram is

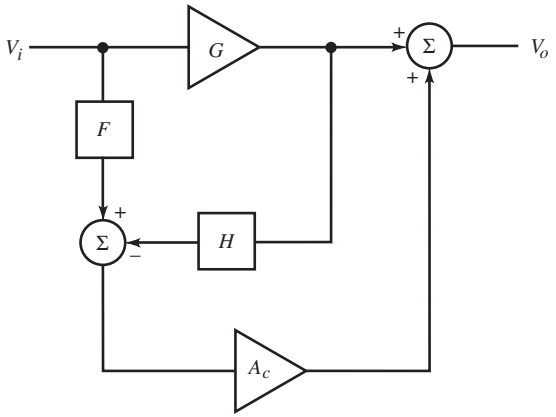
$$A = T_i \cdot \frac{A_{hc}}{1 + H \cdot A_{hc}}, \quad A_{hc} = A_h \cdot (1 + G \cdot A_l)$$

This topology has the classical feedback form with feedbeside within the loop, modifying A_h. The feedbeside loop provides added gain in the fast loop at low frequencies. The attenuator consisting of R_1 and R_2 reduces the gain of this low-frequency loop to stabilize it.

The feedbeside topology is used by John Addis (and invented with Bruce Hofer) in the Tektronix 7104 1 GHz oscilloscope vertical amplifier to correct wideband vertical stages for thermal effects. It is used in a different way in the LMC669 as an offset voltage compensator for op-amp inputs. This auto-zero integrated circuit (IC) samples the inverting virtual ground input of an op-amp and controls its noninverting input to null the voltage offset, much like the circuit shown above.

Feedforward and Linearized Differential Cascode Amplifiers

Feedback compares output with input and drives the forward-path amplifier with the error, thus correcting the output. Instead of applying the error to the input, the *feedforward* scheme adds a compensating error quantity at the output.



The feedforward topology has a forward path through G. Its output is fed back through H and subtracted from the amplifier input through F, resulting in the error of G in amplifying the input. This error is scaled by compensation amplifier A_c to the same magnitude as the output of G. When it is then added to the output of G, the error terms in GV_i are nulled.

The gain of the feedforward topology is

$$A = G + A_c \cdot (F - G \cdot H)$$

To demonstrate error reduction, let

$$G = K + \varepsilon$$

where K is the linear gain of G and ε is the nonlinear distortion terms. For feedforward error nulling, the scaling of F, H, and A_c must be correct. F and H

are passive attenuators and are assumed linear. The input to A_e should be only error terms; V_i is nulled by scaling F and H so that

$$K \cdot H = F \Rightarrow H = \frac{F}{K}$$

In addition, since F has scaled down the input, A_e must amplify it to correct for scaling and have a gain of K/F. Because A_e also has distortion, and assuming nonlinear terms scale with amplitude, then let

$$A_e = \frac{K + \varepsilon_e}{F}$$

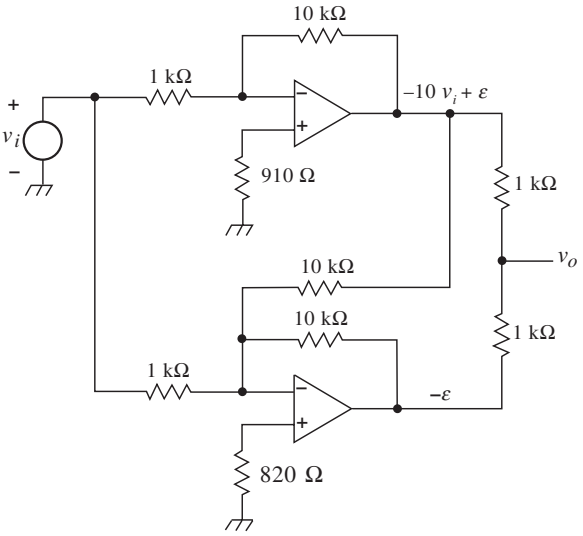
Substituting from these three equations into A,

$$A = K - \frac{\varepsilon \cdot \varepsilon_e}{K}$$

The feedforward advantage is that distortion is reduced from ε to a magnitude of about ε^2/K. For small ε, this is a large improvement in linearity. In practice, it requires accurate scaling and gain-matching of transmittances, whereas feedback does not. Feedforward amplifiers can be made faster than feedback amplifiers because loop delay is not a limitation. However, the error-path delay must match the main path for correct output summation. Although F and H are passive and can have little delay, the gain-bandwidth demand on A_e is greater than G because it has more gain and requires slightly greater bandwidth due to F and H. Because it is amplifying an already small error, its linearity need not be high.

Example: Op-Amp Feedforward Amplifier

The feedforward amplifier uses two inverting op-amps to give a gain of −10. Input summing occurs at the inverting input of the error amplifier, and output summing is done with resistors. The output might need to be buffered, and the output summer could be the input of another inverting op-amp. A dual op-amp IC provides matched op-amps. The gain-determining resistors must be matched.



The feedforward topology can be used to linearize fast BJT diff-amps. Nonlinearity is caused by variation in transistor parameters r_e, β, r_o, and f_T. The most significant is r_e. As v_i varies, the emitter current,

$$i_e = \frac{v_i}{r_{e1} + r_{e2} + 2R_E + 2R_B/(\beta + 1)}$$

also varies with r_{e1}, r_{e2}, and β. Transconductance is largest when v_i is zero (see "Transconductance Linearity of BJT Diff-Amp") and decreases with $|v_I|$. This is a *compressive* (versus *expansive*) gain characteristic. The emitter-referred base resistance term in i_e also varies with $\beta(i_E)$. Input impedance varies expansively with r_e variation, forming a nonlinear divider with base resistance. As f_T also varies with emitter current, C_π varies, causing input capacitance to vary too. And because r_e is proportional to V_T and hence absolute temperature, gain varies with temperature. To stabilize, either R_E must be increased or the emitter current source I_0 must track V_T.

All of these causes of distortion can be compensated individually, but this approach is complicated and difficult to achieve. A simpler, more elegant solution is to remove them all together by nulling the combined error with

feedforward. In particular, we concentrate on *b-e* junction nonlinearity. Techniques for correcting β or α follow.

In 1976, Pat Quinn developed the feedforward amplifier shown below. Q_1 and Q_2 comprise the diff-amp, and Q_3 and Q_4 comprise the error amplifier, also a diff-amp. The error amplifier output current is summed with the main output current at their collectors. Instead of feeding back from the main collectors, the error is sensed at the emitters and summed with resistors R_1 through R_4. Applying Kirchhoff's voltage law (KVL) around the main diff-amp input loop,

$$v_I = v_2 - v_1 = v_{BE2} - v_{BE1} + (i_{E2} - i_{E1}) \cdot R_E = \Delta v_{BE} + \Delta i_E \cdot R_E$$

By superposition, the input voltages to the error amplifier are:

$$v_3 = v_1 \cdot \left(\frac{R_3}{R_1 + R_3} \right) + v_{E2} \cdot \left(\frac{R_1}{R_1 + R_3} \right)$$

and

$$v_4 = v_2 \cdot \left(\frac{R_4}{R_2 + R_4} \right) + v_{E1} \cdot \left(\frac{R_2}{R_2 + R_4} \right)$$

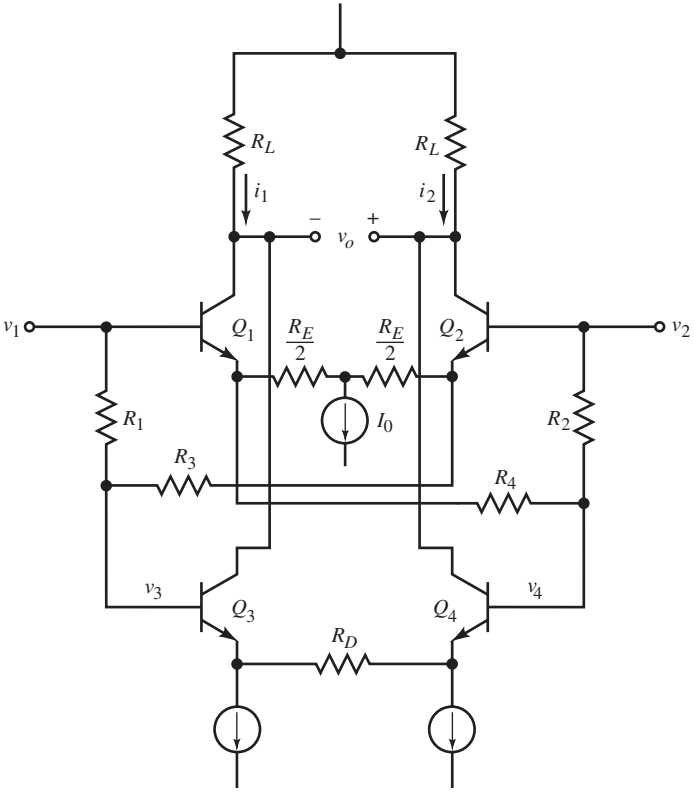
Feedforward scaling requires

$$R_1 = R_3, \quad R_2 = R_4$$

Then the dividers on each side have the same attenuation F of 1/2, and the error amplifier differential input is

$$v_4 - v_3 = \frac{\Delta v_{BE}}{2}$$

This input contains only the scaled Δv_{BE} error term of v_I, as required for feedforward. The error amplifier transconductance must now be scaled to cancel the Δv_{BE} terms of the main output current. The main amplifier output current is



$$i_O = \alpha \cdot \Delta i_E = \alpha \cdot \left(\frac{v_I}{R_E} - \frac{\Delta v_{BE}}{R_E} \right)$$

where $\alpha_1 = \alpha_2 = \alpha$. The output current of the error amplifier is

$$i_O = \frac{\Delta v_{BE}}{2} \cdot \frac{\alpha}{R_D + r_{e3} + r_{e4} + 2R_B/(\beta + 1)}$$

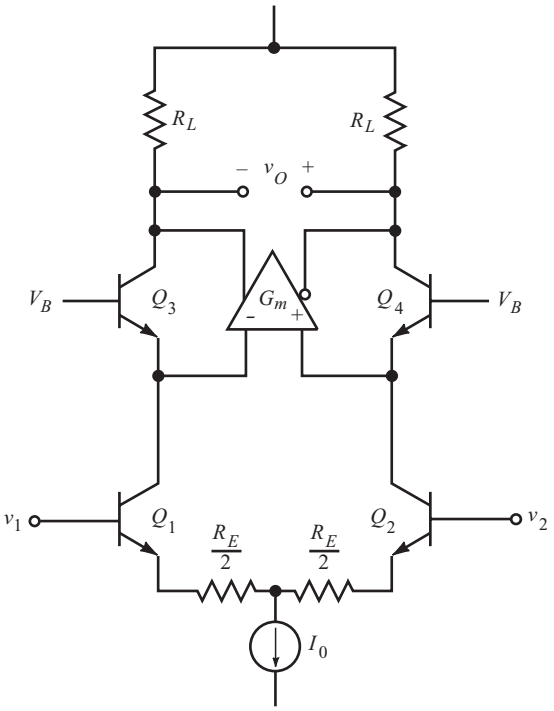
where $R_B \cong$ the Thevenin resistance of the dividers. Error amplifier gain scaling requires that

$$R_E = 2 \cdot \left(R_D + r_{e3} + r_{e4} + \frac{2R_B}{\beta + 1} \right) \cong 2R_D, \quad R_D \text{ dominant}$$

With this scaling, Δv_{BE} currents cancel at the output, resulting in an amplifier gain of

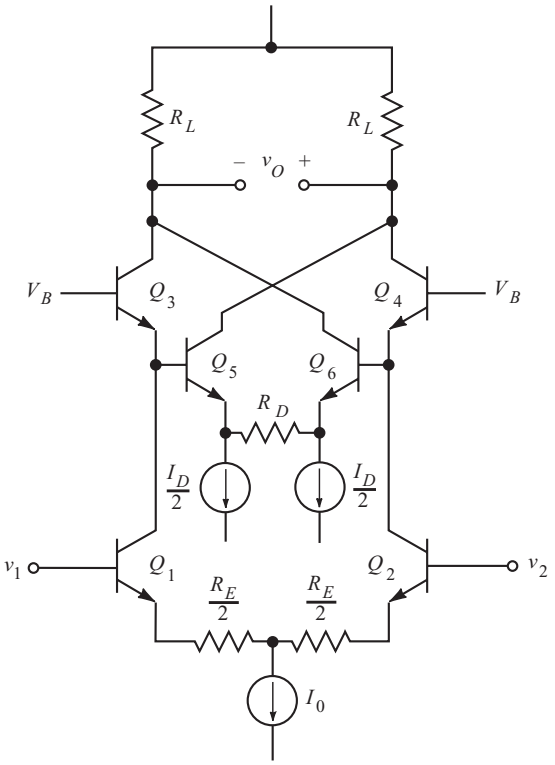
$$A = -2\alpha \cdot \frac{R_L}{R_E}$$

with *b-e* nonlinearity removed. The error amplifier has twice the gain of the main amplifier. If the transistors are of the same type, then precision amplification extends to about half the bandwidth of the main path.



The feedforward diff-amp was succeeded by the *cascomp* (shown above), an invention of Barrie Gilbert, patented by Quinn at about the same time. The main amplifier is a differential cascode. Because the same currents (apart from i_B) flow through the common base (CB) as through the common emitter (CE) stage, the nonlinearity of the CE transistors is duplicated at the emitters of the CB. Because the CB base voltages are fixed, the differential emitter voltage is

the error voltage, conveniently ground-referenced. The resistive dividers are not needed, and the error amplifier is driven directly by the CB Δv_E. The error amplifier can be another diff-amp, as shown below.



The error is nulled when

$$G_m = \frac{\alpha}{R_D + r_{e5} + r_{e6}} = \frac{\alpha}{R_E}$$

For dominant R_D, the condition for cascomp error-path gain is that $R_D = R_E$. The error amplifier corrects for main amplifier error, which is worst at the limits of its dynamic range. It is here that error-amplifier correction is most needed, and its gain is matched at these range limits.

The emitter bias current I_D is arbitrary, and criteria for its optimization can be sought. Increasing I_D reduces error amplifier Δr_e, increasing its linearity.

It also increases error-amp base current I_B, which is part of the CB emitter current. This causes v_{BE}(CB) error but tends to compensate for static α loss of the CE input transistors. By adding I_B-compensating sources at the error amplifier bases, total amplifier linearity is improved and dynamic range extended. For $I_D/I_0 > 3$, a total current $I_D + I_0$ in an uncompensated cascode reduces its Δr_e to the point of comparable linearity to the cascomp. An optimum ratio of I_D/I_0 is about 2.

The error amplifier must be linear over an input range of Δv_{BE}(CB), which is limited by I_0. Its maximum dynamic range is

$$\Delta v_{BE}(\text{CB}) = V_T \ln\left(\frac{I_0}{I_B}\right)$$

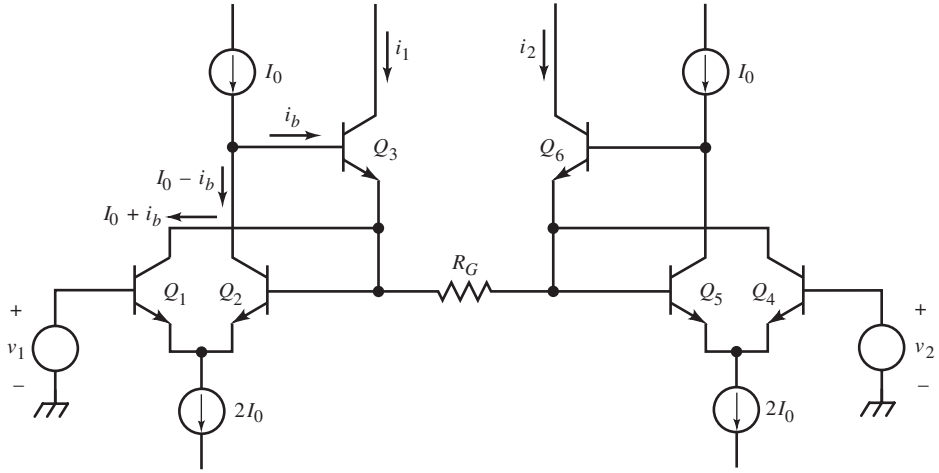
because the minimum emitter current of Q_3 and Q_4 is I_B. This base current reduces Δv_{BE}(CB), effectively reducing the gain and dynamic range of the error amplifier. Second, the CE emitter current generates v_{BE}(CE), but its collector current generates v_{BE}(CB). This current mismatch causes a corresponding mismatch between Δv_{BE}(CE) and Δv_{BE}(CB). This "α error" can be compensated by adding a base resistor to the CB. It drops additional voltage that compensates for the loss of i_B.

The cascomp is thermally compensated by stacking another differential CB onto the output of the cascomp CB transistors. The output CB base voltage is set to provide the same thermal operating point as the CE stage beneath.

α-Compensated Gain Cells

Amplifier stages with sub-1% nonlinearities must compensate for α error due to finite β and loss of base current. Two multiple-transistor stages with feedback that function as a gain unit, or *cell*, were devised by Ken Schlotzhauer and Stewart Taylor at Tektronix. These gain cells both use α-compensation techniques.

The Schlotzhauer cell, shown below, is a diff-amp with CC feedback to Q_2. The feedback loop has little delay, and the cell is fast. Two of these cells are used as ×1 buffer amplifiers. They apply the differential input voltage that appears across R_G. The differential transconductance is therefore R_G, or $R_G/2$



per side. As a noninverting configuration, the input impedance is large due to feedback.

In the left cell, the α loss in the CC (Q_3) is compensated by connecting the collector of Q_1 to the emitter of the CC instead of the supply. The current through R_G is then the output current, i_1. The base current added to i_1 is removed as Q_1 collector current. Because

$$i_{c1} + i_{c2} = 2I_0 = \text{constant}$$

the base current of Q_3 lost from i_{C2} must add to i_{C1}. The resulting current in R_G is i_1. The shunting base impedance of Q_2 is negligible if the loop gain is large.

The error quantity on the left side is

$$E = v_1 - v_{B2}$$

Applying feedback analysis, the forward gain to R_G is

$$G = \frac{v_{B2}}{v_1} = \alpha \cdot \frac{(\beta+1)(r_e + R_G/2)}{2r_e} \cdot \left(\frac{R_G/2}{R_G/2 + r_e} \right) - \alpha \cdot \frac{R_G/2}{2r_e}$$

assuming symmetrical circuitry for a virtual ground at the midpoint of R_G and equal r_e. The first term is due to i_{C2} and the second to i_{C1}. The load resistance of Q_2 is the input resistance of Q_3. G simplifies to

$$G = \alpha \cdot \frac{\beta \cdot (R_G/2)}{2r_e}$$

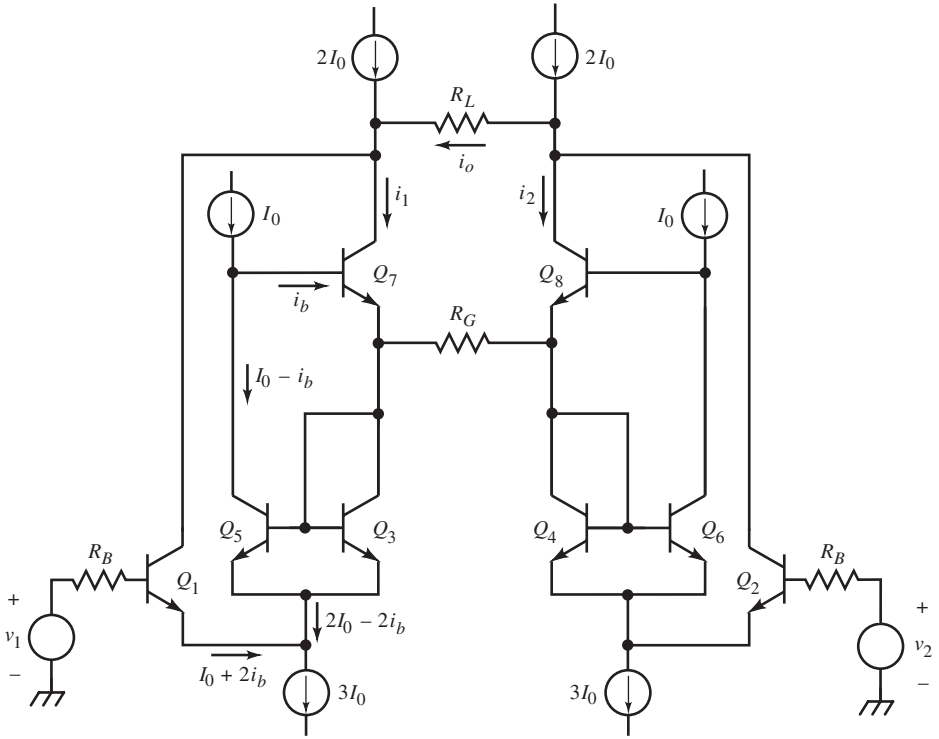
and $H = 1$, due to the direct connection. The closed-loop voltage gain is thus

$$\frac{v_{B2}}{v_1} = \frac{R_G/2}{2r_e/\alpha \cdot \beta + R_G/2}$$

When $2r_e/\alpha\beta << R_G/2$, the gain approaches 1, and the nonlinearity due to r_e is β times less than that of a diff-amp with $R_G/2$ emitter resistance. The single-ended incremental input resistance is

$$r_{in} = \frac{v_1}{i_{b1}} = \frac{v_1}{i_{b3}/\beta} \cong \beta \cdot \frac{v_{b3}}{i_{b3}} = \beta^2 \cdot \frac{v_{b3}}{i_{c3}} = \beta^2 \cdot \frac{R_G}{2}$$

an improvement of about β times over a diff-amp.



The conceptually similar Taylor cell, shown above, is a differential amplifier. The input voltage appears across R_G via Q_1 and Q_3 on the left side and Q_2 and Q_4 on the right. It causes an incremental current through R_G that flows through Q_7 and Q_8 to the output. The goal is to correct the α error of Q_7 and Q_8. Analyzing the left side, beginning with Q_5, the collector current is

$$i_{C5} = I_0 - i_{B7}$$

Q_5 and Q_3 comprise a current mirror, and i_{C5} is replicated as

$$i_{E7} = I_0 - i_{B7}$$

on the assumption that the areas of Q_5 and Q_3 are ratioed for a current-mirror gain of 1, not $(\beta + 2)/\beta$ as for equal areas. Q_7 loses i_{B7} by α loss and

$$i_{C7} = I_0 - 2 \cdot i_{B7}$$

At the emitter node of Q_1, Kirchhoff's current law (KCL) yields

$$i_{E1} = 3 \cdot I_0 - (i_{E5} + i_{E3}) = 3 \cdot I_0 - (2 \cdot I_0 - 2 \cdot i_{B7}) = I_0 + 2 \cdot i_{B7}$$

At the output node, i_{C1} and i_{C7} add. Applying KCL,

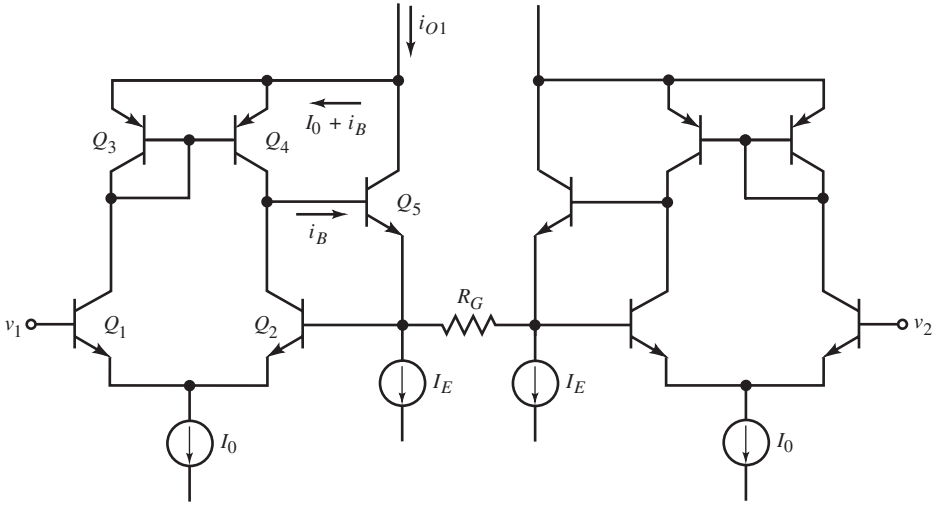
$$i_o = [3 \cdot I_0 - (1 + A) \cdot (I_0 - i_B)] + \alpha \cdot A \cdot (I_0 - i_B) - 2 \cdot I_0 = 0$$

If Q_1 has α = 1, then this equation equals zero, and Q_7 α error is compensated.

The last cell, shown below, is similar to the Schlotzhauer cell except that a current mirror is used. The total mirror current is I_0, conducted by the diff-amp Q_1–Q_2, plus the base current of the output CC, Q_5. The mirror current is part of the output, and includes i_{B5};

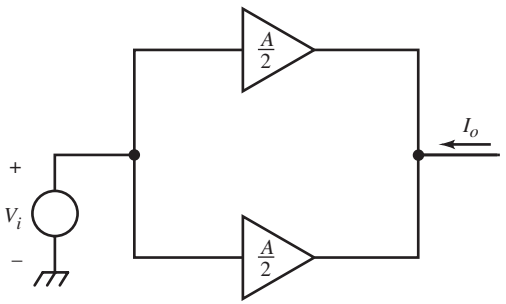
$$i_{O1} = I_0 + i_{E5} = I_0 + I_E + \frac{v_1}{R_G/2}$$

The cell is also differential, and R_G determines the transconductance.



f_T MULTIPLIERS

The f_T (gain-bandwidth product) of current-feedback amplifiers is not gain-limited, but this property is not unique to them. Amplifiers with a fixed f_T can be combined to achieve a greater f_T. A f_T-multiplying stage output can avoid bandwidth reduction by summing current outputs from individual amplifiers and converting them to a voltage in a succeeding stage. For very fast amplifiers, a current input creates an input voltage to paralleled transconductance amplifiers across source resistance R_S. The composite amplifier then has a current gain instead of a voltage gain.



As the gain of an individual amplifier is reduced, its bandwidth increases. The figure shows two amplifiers connected in parallel. The transfer function for each amplifier individually is

$$A = \frac{A_o}{s/(\omega_T/A_o)+1}$$

where the bandwidth is

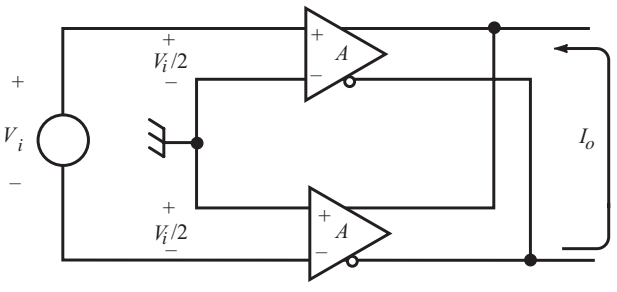
$$\omega_{bw} = \frac{\omega_T}{A_o}$$

When the lf gain A_o is halved, the bandwidth doubles. Two of these amplifiers in parallel have an additive output and the total gain is equal to that of A, or

$$\text{composite } A = 2 \cdot \left(\frac{A_o/2}{s/(\omega_T/(A_o/2))+1} \right) = \frac{A_o}{s/(2 \cdot \omega_T/A_o)+1}$$

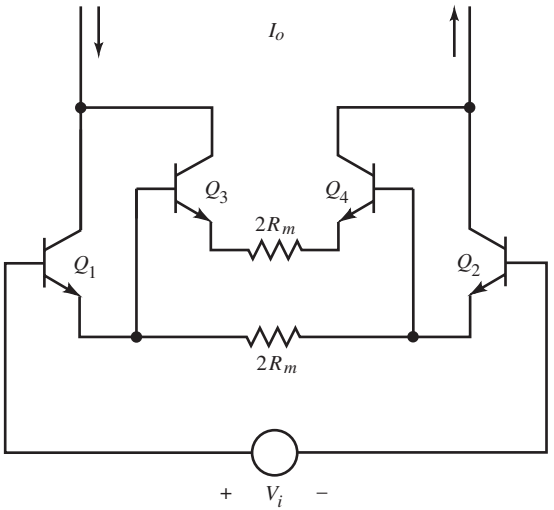
Compared with A, the composite amplifier has the same lf gain but twice the bandwidth. In effect, f_T has doubled.

Such amplifiers, called f_T *doublers*, were invented by Carl Battjes for use in oscilloscope vertical amplifiers. He extended the idea to an arbitrary number of amplifiers; three amplifiers triples f_T. In practice, parasitic elements of most f_T multiplier topologies cause diminishing returns above two amplifiers. The input loading increases, and the input pole decreases bandwidth faster than extra amplifiers increase it.



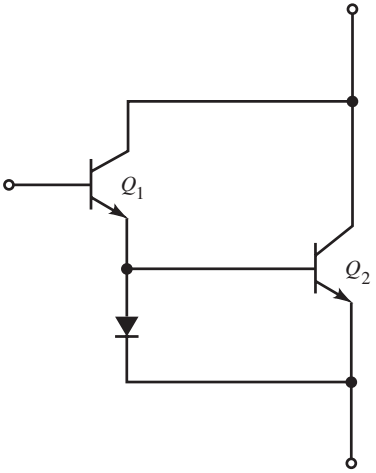
The differential circuit realization of the f_T doubler is shown above. Two diff-amps are connected so that their inputs are in series across V_i, and their outputs are in parallel, cross-coupled so that they add. With a gain of A and input of $V_i/2$ for each diff-amp, the f_T doubler output is AV_i, the same as a single diff-amp with input of V_i and gain of A. But the f_T doubler has twice the gain at the same bandwidth. The difference is illustrated in one equation for f_T multiplication, m:

$$\begin{array}{ccccc} m & \cdot & \left[A \cdot \left(\dfrac{V_i}{m} \right) \right] & = & A \cdot V_i \\ \uparrow & & \uparrow & & \uparrow \\ m & & f_T & & f_T \end{array}$$



For the differential f_T doubler shown above, the input to the second diff-amp is taken from across $2R_m$ of the input diff-amp. For $R_m >> r_e$, the input to the second diff-amp is approximately V_i also.

Single-ended f_T doublers are also possible. The topology, shown below, is basically that of a Darlington configuration except for the diode. This gain cell has three terminals and can be substituted for a single BJT. The base current



of Q_1 generates emitter current i_{E1}, which flows through the diode, creating the same voltage drop, v_{BE2}, as v_{BE1}. This results in i_{C2}, which sums with i_{C1}, producing $2i_{C1}$. This current output is caused by an input of i_{B1}; therefore, the current gain is $2i_{C1}/i_{B1} = 2\beta$. The current gain is double that of a single BJT with the same f_β. Thus, f_T has doubled.

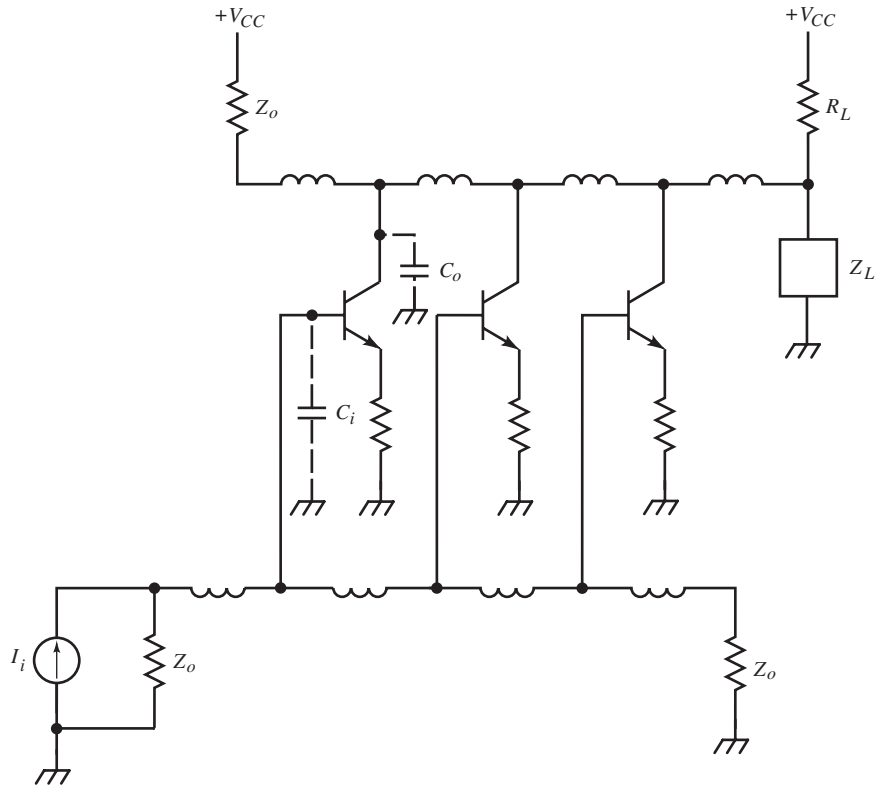
More precisely, some of i_{E1} goes into the base of Q_2. With matched junctions, the diode current is i_{C1}, and the current mirror replicates this current in the emitter of Q_2, or $i_{E2} = i_{C1}$. The output current is slightly less than before, or

$$i_O = (1+\alpha)\cdot i_{C1} \cong 2\cdot i_{C1}, \quad \beta >> 1$$

The f_T-doubler current gain is consequently that of a single-ended f_T doubler:

$$A_i = (1+\alpha)\cdot\beta(s)$$

The idea of adding the outputs of multiple amplifiers instead of cascading them leads to the fastest amplifier topology, the *distributed amplifier*, shown in simplified form below. Start with two discrete delay lines. The input source drives the input line. As the waveform propagates along the line, it drives the inputs of gain stages. They are typically a single transistor amplifier with large input resistance and a controlled input capacitance that is part of the delay line.

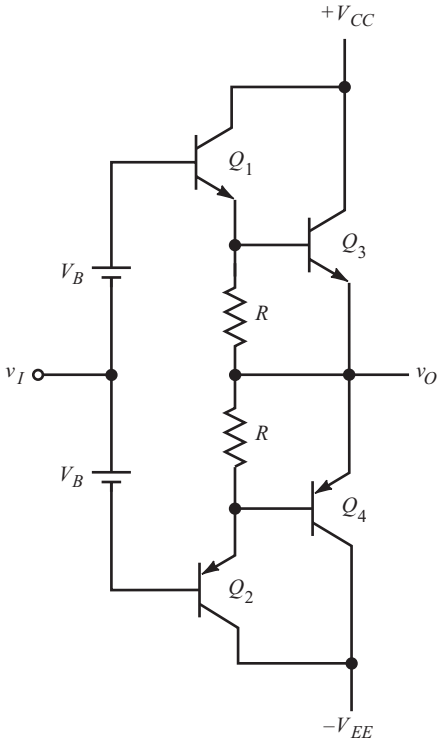


Its output feeds a tap on the output line, also with controlled output capacitance. As the individual stages respond, their outputs accumulate as they propagate down the output line toward the load. Input and output waveforms propagate in synchronism for minimum phase error in the summation of outputs. Each line is terminated for no reflections at each end.

Although distributed amplifiers have been implemented in various oscilloscope vertical amplifiers, they are a last resort at speed improvement because of the number of stages required for a given gain. For n stages with stage gain A, the composite gain is $n \cdot A$ rather than A^n as for cascaded stages. For large gain requirements, the number of stages can be excessive. With increasing IC density, however, a novel form of IC distributed amplifier with silicon-based delay-line structures might be feasible.

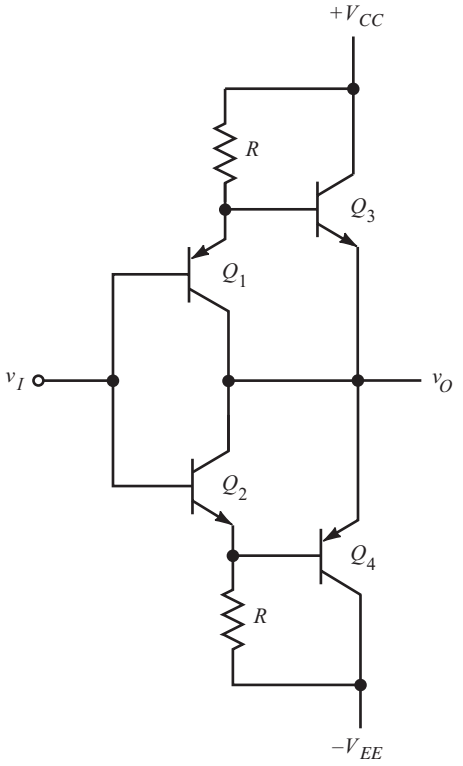
High-Performance Buffer Amplifiers

Amplifiers with a high-impedance input, accurate (usually ×1) voltage gain, and a low-impedance, large-signal output are called buffer amplifiers, or *buffers*. They are typically used to drive low-impedance loads and have a large-signal dynamic range. Examples of loads are video distribution cables, telephone lines, audio speakers, cathode-ray tube (CRT) deflection plates, magnetic deflection yokes, electromechanical devices, and pulse- and function-generator outputs.



The BJT two-stage complementary emitter-follower shown above – a variation on that in the section "Complementary Emitter-Follower Output Amplifier" – is preceded by a complementary CC stage. This results in a higher input imped-

ance. In addition, instead of returning the input-stage emitters to the supplies, they are connected to the output. This has the effect of reducing the output-stage deadzone by supplying output drive from the input stage. The deadzone is around 0 V at the output, making the output current requirement within it small. Consequently, the input stage can supply much of the needed current, putting otherwise wasted current to use.

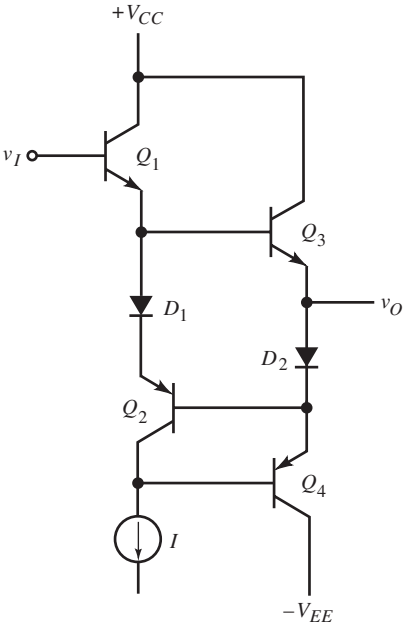


The buffer shown above also supplies input-stage current to the output and reduces the deadzone. Its complementary CC input stage is inverted from that of the previous buffer, and the collectors connect to the output. In this circuit, as v_I increases, Q_2 increases conduction, supplying emitter current to Q_3. In the deadzone, this current biases Q_3 on, reducing r_{e3} and increasing gain. By symmetry, Q_1 similarly biases Q_4. This buffer does not necessarily require input-bias

voltage sources since the *b-e* junction voltages of Q_1–Q_3 and Q_2–Q_4 tend to cancel. The choice of R in both buffer circuits is based on a tradeoff between input characteristics and deadzone reduction.

The output buffer shown below has equal diode currents at 0 V output. The current source sinks I through Q_2 and D_1. Q_2 and Q_4 are coupled to exchange base currents and are thereby α-compensated. Because

$$v_{D1} + v_{BE2} = v_{BE3} + v_{D2}$$

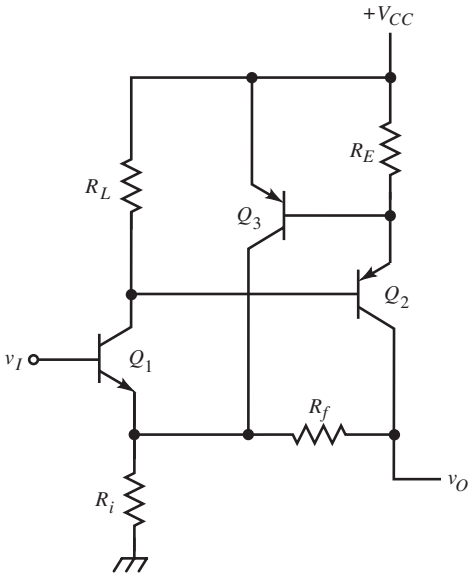


with matched junctions and no load current, $i_{D2} = i_{D1}$. Consequently, the base currents of Q_2 and Q_4 are equal (for matched devices) and cancel. As v_I varies, load current upsets the compensation somewhat as v_{BE3} no longer matches v_{BE2}. However, the deadzone is reduced because Q_3 is biased by an emitter current source of I.

The buffer can sink up to $(\beta + 1) \cdot I$ current. Applying KCL at the bases of Q_2 and Q_4 and solving

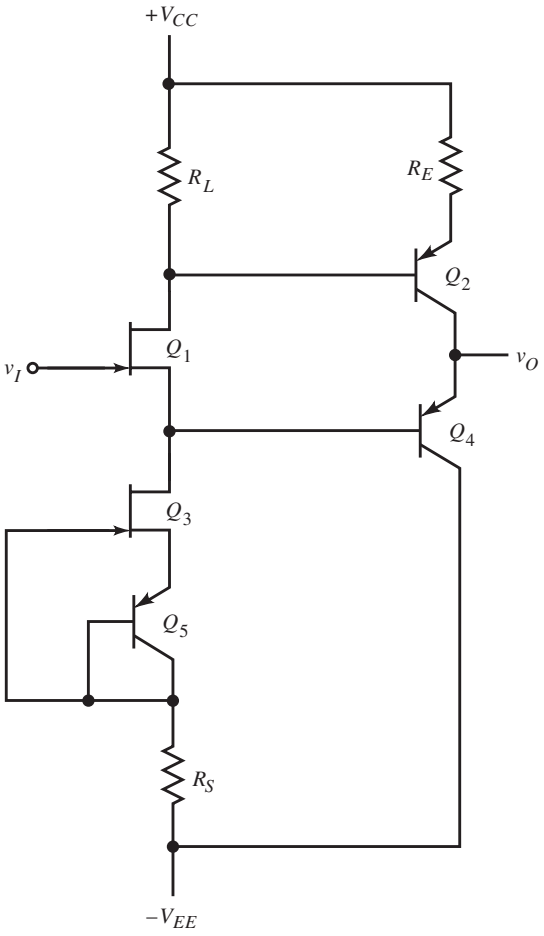
$$i_{D2} = (\beta+1)\cdot I - \frac{\beta^2+\beta+1}{\beta+1}\cdot i_{D1} \cong (\beta+1)\cdot I - \beta \cdot i_{D1}$$

For large negative inputs, $i_{D1} = 0$, and I is the base current of Q_4. Thus, I can be kept small to reduce the base current of Q_1. This buffer has an inherent voltage offset of $v_{BE1} + v_{BE3}$.



Output current-limiting does not appreciably reduce the dynamic range in the noninverting feedback buffer shown above. Without Q_3 and R_E, the maximum output occurs when Q_2 saturates and is V_{CC}. When the current-limit circuit is added, some drop occurs across R_E, but it can have a small value because of the gain through Q_3 to the amplifier input loop.

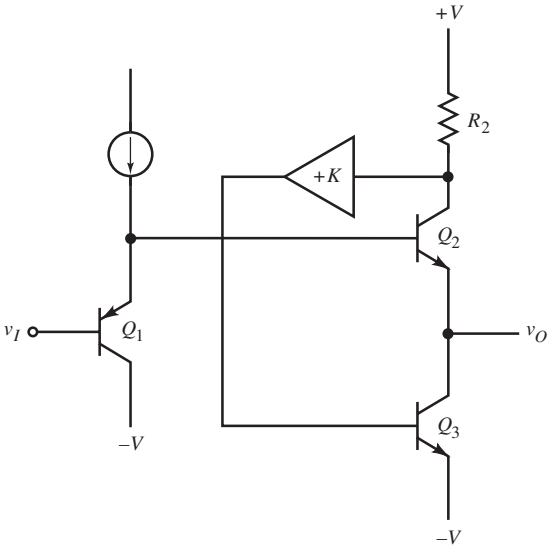
More precise buffer amplifiers have better input characteristics and less offset voltage. In the figure below, Q_4 is matched with Q_5. Then with matched FETs,



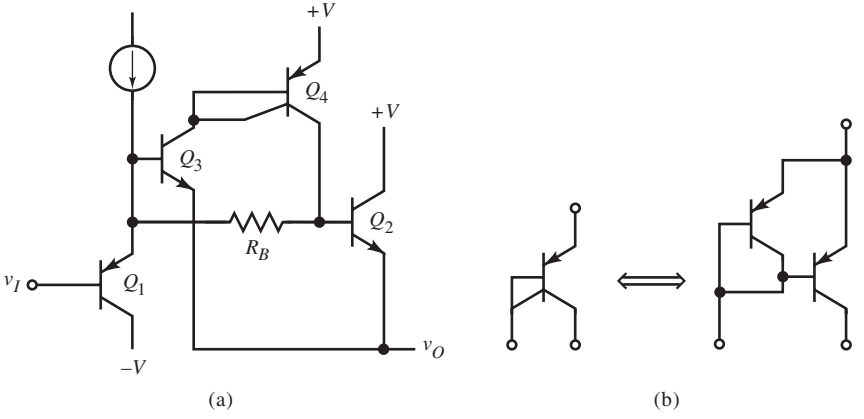
$$V_{GS1} = -V_{BE5} = -V_{BE4}$$

and offset voltage is canceled. Because V_{GS3} is forced to be V_{BE5}, any FET drift in Q_3 causes a corresponding drift in Q_1, thus nulling FET TC effects on offset voltage.

The current-sourcing capability of the buffer is improved by the Q_2 CE BJT, though this path is slower than for current-sinking through Q_4. R_S might be needed for thermal compensation.

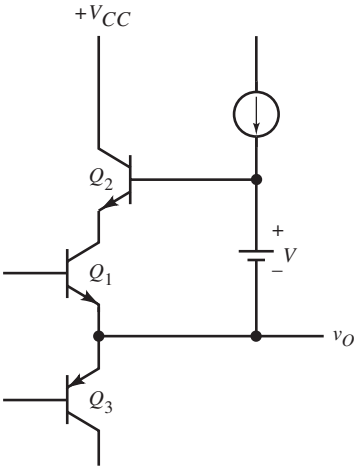


A similar two-path buffer topology is found in the Linear Technology LT1010 and is simplified in the above figure. Q_1 and Q_2 are complementary CCs of the fast path. Q_2 also functions as a CE to drive the slow path through a gain of K to Q_3, which provides active current-sinking, similar in function to Q_2 of the previous buffer.



The general technique of bootstrapping is applied to the above buffer in (a) above. (This circuit is also used in the LT1010.) Q_3, Q_4, and R_B form the boot-

strap circuit. Q_4 is an integrated split-collector BJT with one collector connected to the base. This is equivalent to the circuit in (b) and functions as a simple current mirror. The bootstrap circuit is driven from the input to Q_2 and forms an active parallel path to its base, supplementing the drive of the passive path through R_B. To increase efficiency, the emitter current of Q_3 is supplied to the output.

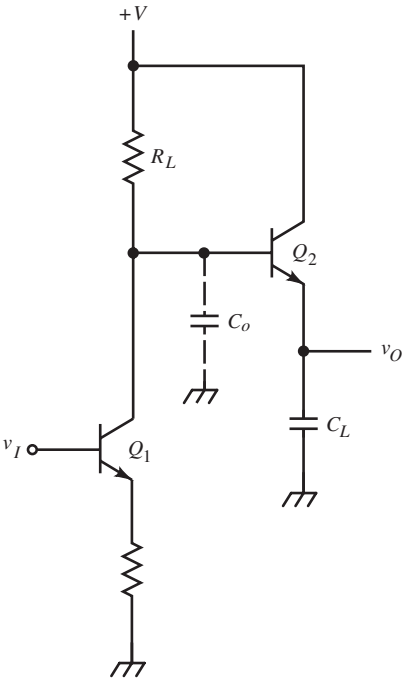


Finally, a general bootstrapping technique for increasing the dynamic range of an output stage is shown above. Q_2 is bootstrapped via V atop Q_1. This minimizes power dissipation in Q_1 but gives it more collector voltage as the output requires it. A complementary circuit is applied to Q_3.

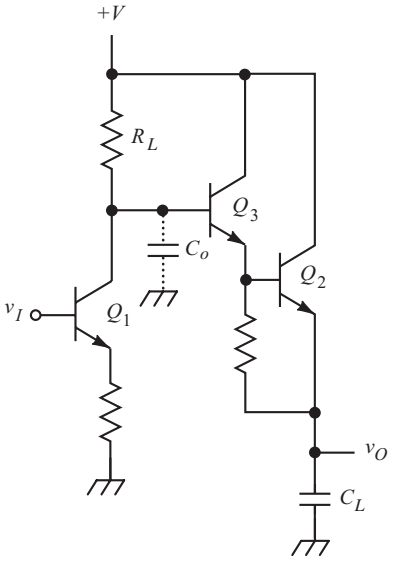
Unipolar Voltage-Translating Amplifiers

Some amplifier stages require control from near circuit ground to supply a unipolar drive at an elevated voltage. Examples are the current-sourcing drivers of H-bridge power switches, oscilloscope horizontal deflection amplifiers, and voltage translators. The basic problem to be solved is that of supplying adequate drive to the elevated (high-side) output device from ground.

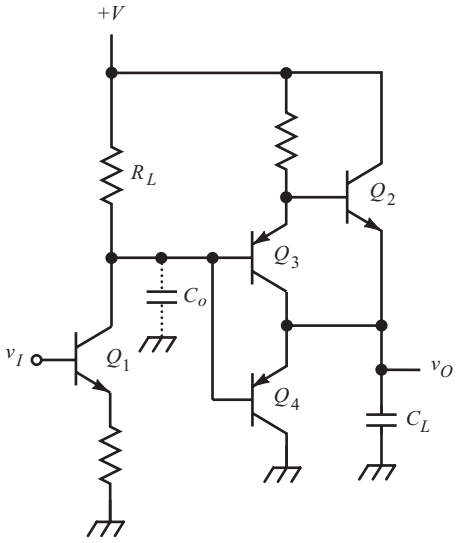
The situation is shown below. Q_1 operates near ground and drives Q_2. Since Q_1 can only sink current, it cannot actively drive Q_2 on. R_L supplies Q_2 drive



but must be kept large to limit Q_1 current. The capacitance C_o at the base of Q_2 and load capacitance referred to the base node limit amplifier speed. Various schemes have been devised to replace R_L with a high-side driver to reduce power dissipation in Q_1 and increase slew rate.

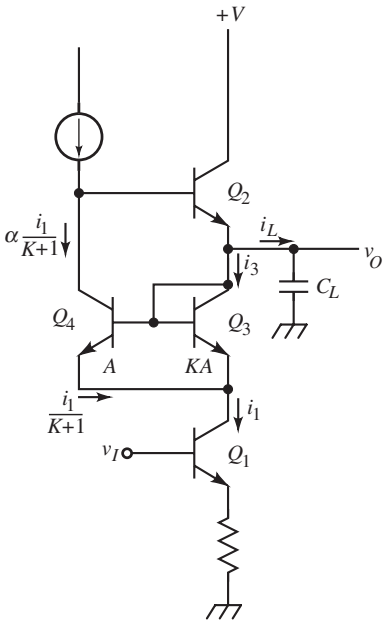


If Q_2 is made a Darlington stage instead, less drive current is needed and the slew-rate-limiting base node is further isolated from the load. Yet the time constant, $R_L \cdot C_o$, is not reduced, and with the higher input resistance of the Darlington, the node response time is increased for small signals.



A complementary Darlington, as shown above, is used in the LM3900 with some advantages over the previous circuit. The output voltage range is increased by a junction drop. Also, the *b-c* junction of Q_3 conducts in the forward direction for large, quick negative-going outputs and provides a low-impedance path from the output to Q_1. In effect, Q_1 is the output current-sinking driver. (This feature can be easily added to the basic high-side driver by placing a diode in reverse across the *b-e* junction of Q_2 but with no current gain advantage.) When the output voltage difference from the input is small (as in the LM3900), Q_4 can be added to buffer Q_1 for sinking output current. In amplifiers with large V_{BC4}, Q_4 (like Q_1 in the basic high-side driver) can dissipate excessive power. Therefore, high-side drivers are needed that do not sink large currents to ground across large voltage drops.

One approach is to minimize C_o. The input stage can be made a cascode; the Miller effect is eliminated and C_o substantially reduced. A shunt-feedback cascode with high side driver has feedback benefits as well, but feedback itself cannot overcome slew-rate limitations.



In the early 1970s, Battjes devised the floating current mirror scheme shown above. The current mirror consists of Q_3 and Q_4 with area ratio $A_3/A_4 = K$. The emitter current splits in the proportions of

$$i_{E3} = \frac{K}{K+1} \cdot i_1, \quad i_{E4} = \frac{1}{K+1} \cdot i_1$$

The output sink current is

$$i_3 = i_{E3} + i_{B4} = \frac{K}{K+1} \cdot i_1 + \frac{i_1/(K+1)}{\beta+1}$$

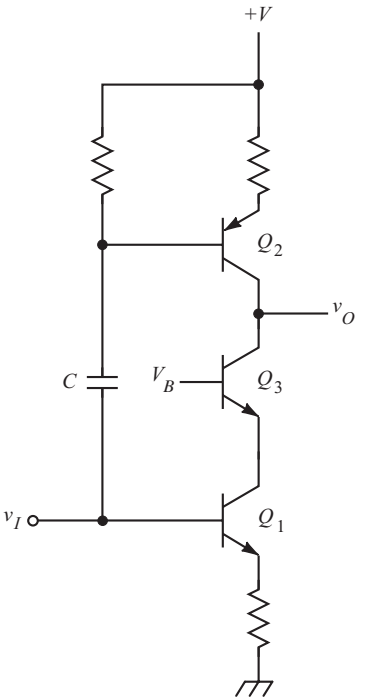
The source current from the emitter of Q_2 is $-(\beta + 1) \cdot i_{C4}$. Voltage inversion occurs at the collector of Q_4, reducing i_{E2} when i_1 increases. Then

$$i_{E2} = -(\beta+1) \cdot \left[\alpha \cdot \frac{i_1}{K+1} \right] = -\frac{\beta}{K+1} \cdot i_1$$

This leads to the transfer ratio

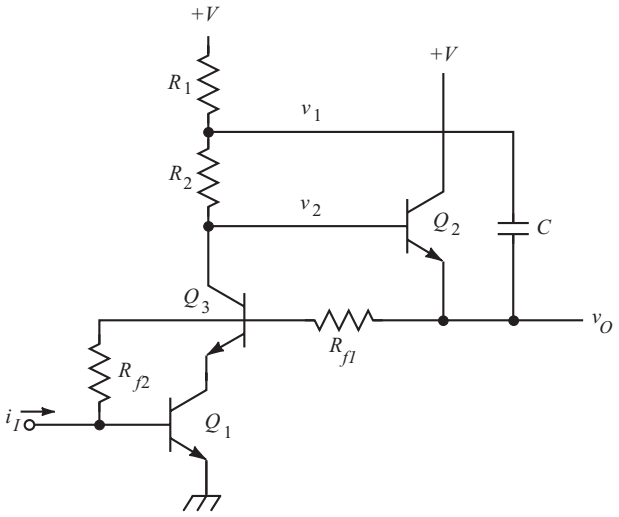
$$\frac{i_L}{i_1} = \frac{i_{E2} - i_3}{i_1} = -\frac{\beta + 1/(\beta + 1) + K}{K + 1}$$

The floating current mirror sinks output current and drives Q_2.

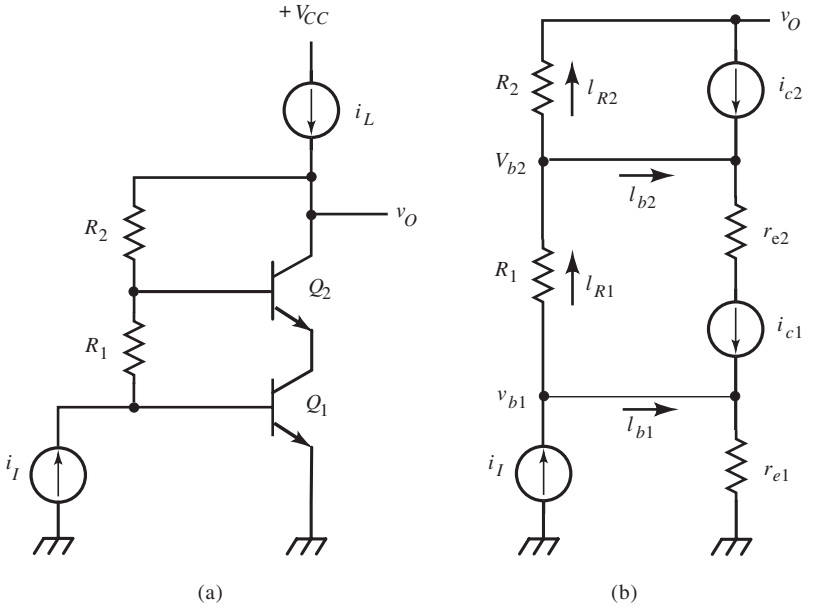


Bootstrapping techniques can be applied to the high-side driver from either the input, as shown above, or output, in the circuit below. Both are dynamic bootstraps that can improve speed. In the above circuit, the cascode input drives Q_2 through C, making it an active current source for fast inputs.

The circuit shown below was developed by Art Metz for oscilloscope horizontal amplifiers. The output, through C, bootstraps R_2; as v_O increases, v_1 does also. This keeps the voltage across R_2 and its current from decreasing as v_2 rises, thereby maintaining current drive to Q_2. Metz further improved the circuit by making Q_1 and Q_2 f_T doublers. (See "f_T Multipliers.") For Q_2, the gain and bandwidth are higher, and base current drive is reduced, leaving more current to charge node capacitance. (This circuit is an implementation of the amplifier in the section "Bootstrap Speed-Up Circuit.")



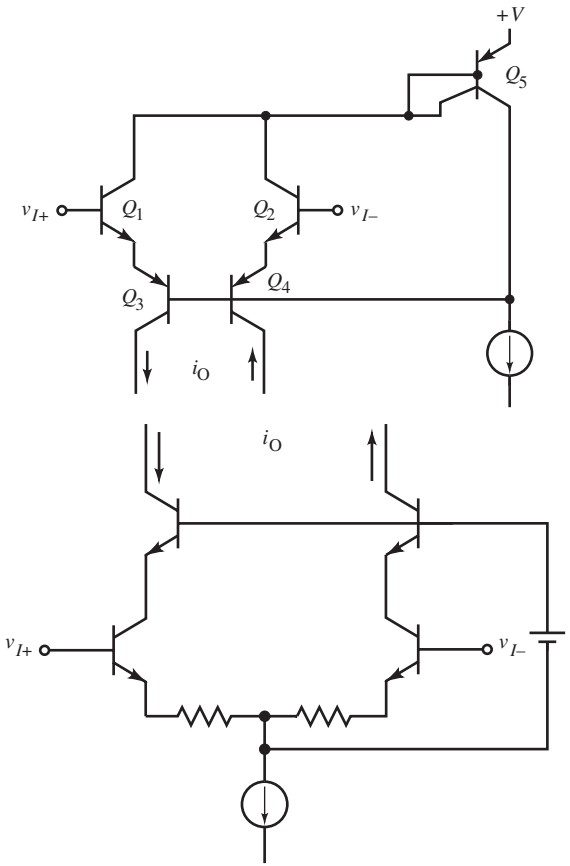
The final embellishment is to tap the shunt-feedback resistor and drive the base of the cascode CB, Q_3, from the tap. The *c-b* junction of Q_3 is bootstrapped as output and base voltage vary together. This increases the dynamic range and reduces C_o and the breakdown voltage requirements of Q_3. The topology is similar to that of the split cascode, shown below. (See "Cascode and Differential Shunt-Feedback Amplifiers" in *Designing Amplifier Circuits.*)



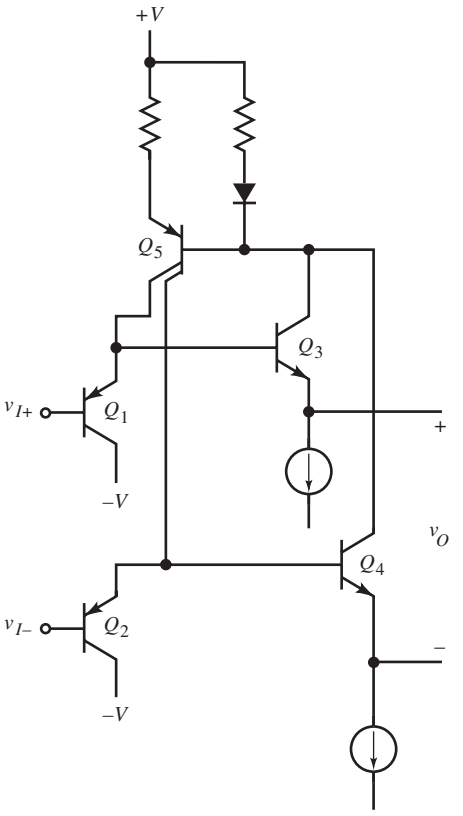
Bootstrapped Input Stages

Bootstrapping is commonly applied to the inputs of op-amps to increase their dynamic range and improve performance. A simplified topology of the classic 741 op-amp input circuit is shown below.

The input differential BJT pair, Q_1, Q_2, is a CC stage driving a complementary differential CB stage, Q_3, Q_4. The CB bases are bootstrapped to follow the inputs and are controlled through the feedback loop through Q_5. As the input pair conduct more current due to a common-mode voltage increase, Q_5 operates as a current mirror and sources more current, causing the CB bases to rise. This bootstrapping effect takes the form of noninverting feedback and also bootstraps the input impedance.

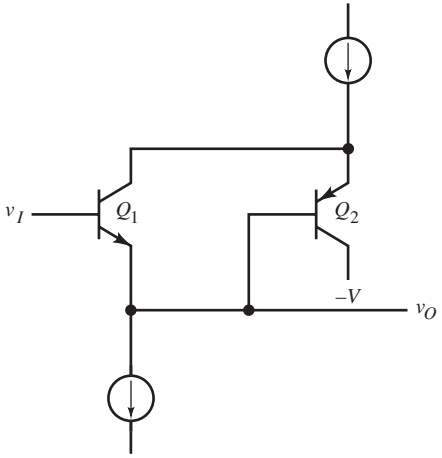


This more conventional diff-amp input stage drives bootstrapped CB transistors. In the circuit, the common-mode input voltage is taken from the emitter virtual ground node. The voltage translator, shown as a battery, is typically a zener diode or resistor driven by a current source.



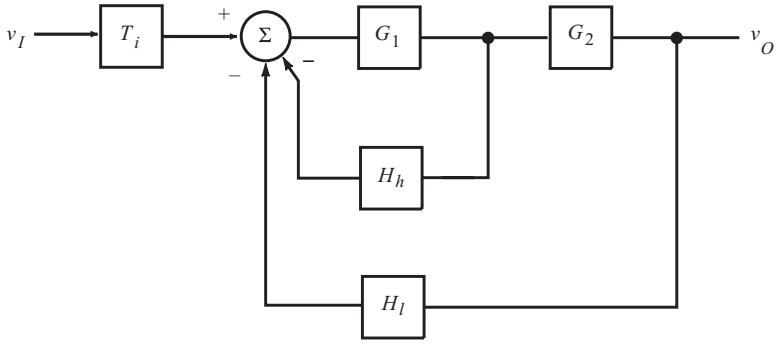
The Linear Technology LT1011 comparator has an input like that shown above. A differential two-stage follower is bootstrapped to control the emitter bias currents in the first stage CCs. The variation in bias currents of the output CCs is tracked in the input CCs.

Finally, the simple input bootstrapping scheme for the collector of Q_1 below is provided by Q_2. Although V_{CE} of Q_1 is limited to V_{BE2}, its collector follows its emitter as the complementary CC output of Q_2.



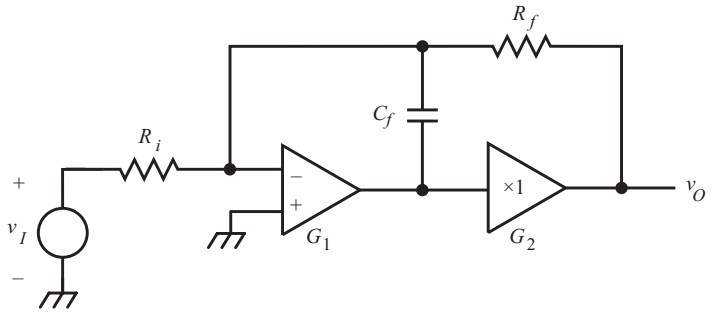
Composite-Feedback and Large-Signal Dynamic Compensation

The composite amplifiers of "Split-Path Amplifiers" had a single feedback path. Now consider composite amplifiers with multiple feedback paths.



In the figure above, the paths are nested. The inner loop of G_1H_h is part of the forward path of the outer loop.

In the corresponding op-amp circuit shown below, the input op-amp drives a more powerful, reactively loaded buffer amplifier. The high-frequency path through C_f is isolated by the buffer and R_f from the load, whereas static feedback



is taken at the load through R_f. (This topology is an alternative to that in "Output Load Isolation" in *Designing Dynamic Circuit Response*.) The block diagram transfer function reduces to

$$\frac{V_o}{V_i} = T_i \cdot \frac{G_1 \cdot G_2}{1 + G_1 \cdot H_h + G_1 \cdot G_2 \cdot H_l}$$

The noninverting version omits T_i. H_h, H_l, and T_i have poles at $1/\tau = 1/(R_f \| R_i) \cdot C_f$, and H_h has a zero at the origin, being an RC differentiator. These blocks are of the form

$$T_i = \frac{T_{io}}{s\tau + 1}, \quad -H_h = \frac{s\tau}{s\tau + 1}, \quad -H_l = \frac{H_o}{s\tau + 1}$$

where

$$T_{io} = \frac{R_f}{R_f + R_i}, \quad H_o = \frac{R_i}{R_f + R_i}$$

Substituting into V_o/V_i, setting $G_1 = -K$ and $G_2 = 1$, the voltage gain is

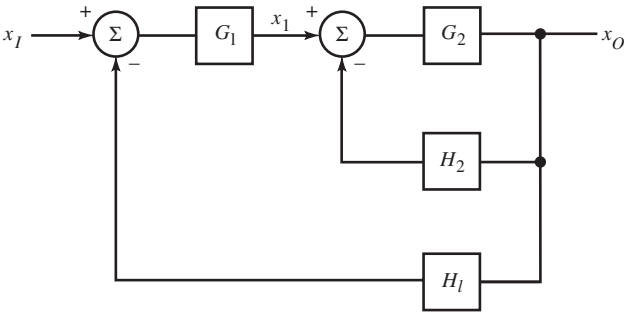
$$\frac{V_o}{V_i} = -T_{io} \cdot \left(\frac{K}{1 + KH_o} \right) \cdot \frac{1}{s\tau \left(\frac{1+K}{1+KH_o} \right) + 1}$$

For infinite K,

$$\frac{V_o}{V_i} = -\frac{R_f}{R_i} \cdot \frac{1}{sR_fC_f + 1}$$

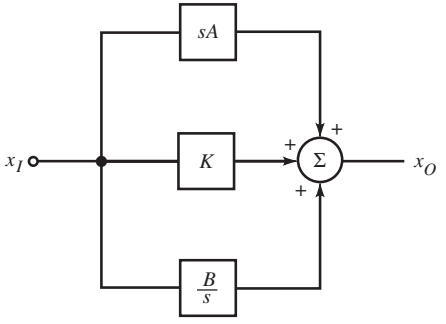
The low-frequency gain is that of an inverting op-amp with a pole at $1/R_f \cdot C_f$. This pole can be placed at a high frequency, away from the poles of the forward path and load.

Nested feedback loops are common in power electronics. In both power converters and motor controllers, a current-controlling inner loop is controlled by a voltage or speed-controlled outer loop. In position controllers for motors, a third outer position loop is added to allow speed to be controlled by position error. The general topology is shown below.



For a position controller, x_1 is speed and G_2, H_2 is a speed-control loop. For a motor speed controller, x_1 is current (or scaled torque), and for current-mode switching power supplies, x_I is voltage and x_1 is current.

Voltage-mode power converters have the same topology, with x_1 as duty ratio and the inner loop as a pulse-width modulator. This topology can be equally well applied to other analog applications and is sometimes called *pseudoderivative feedback* (PDF) control.

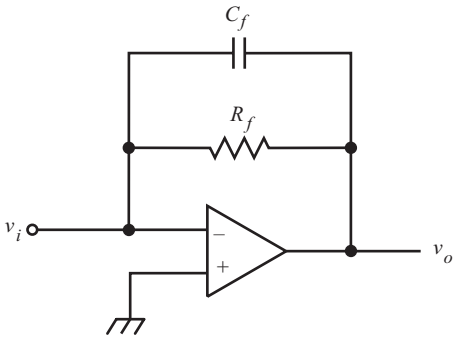


The most common form of compensator for this topology is a variant of lead-lag compensation called *proportional integral differential* (PID) control. This multipath topology has gain, integral, and derivative paths that add. Each path can be adjusted independently, with transfer function

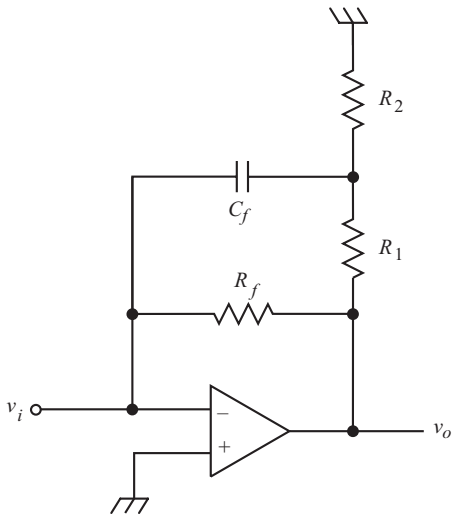
$$\text{PID compensator} = K + sA + \frac{B}{s} = B \cdot \frac{s^2(A/B) + s(K/B) + 1}{s}$$

A, K, and B are independently adjustable, allowing the coefficients to be set with only scale interaction. The static gain, B, scales the coefficients and is set first. Note that this transfer function has more zeros than poles and can deter phase lag more effectively than lead-lag compensation.

The complexity of topologies beyond those given here requires a more abstract and simplifying formulation. State-variable control theory is a good foundation for more complex designs, and as with classical control, some techniques do appeal to design insight. Formulations must be sought that provide physically meaningful insight into circuit operation; see Lorenz (1986).



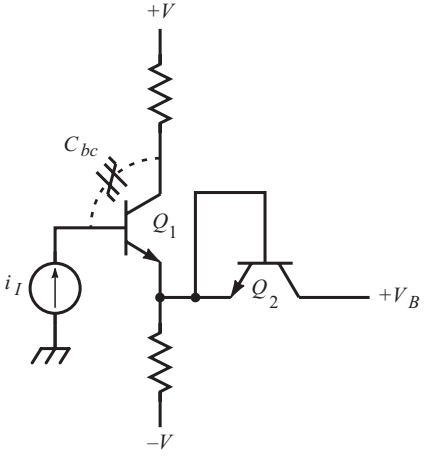
A design technique is needed for compensating feedback amplifiers with large feedback resistors, as shown above. The problem caused by the large value of R_f is that C_f must be made too small to be practical for proper compensation. Not uncommonly, fractions of a picofarad are required. Another problem caused by C_f, due to the Miller effect, is large input capacitance. Some sources must have minimum capacitive loading, such as D/A converter outputs.



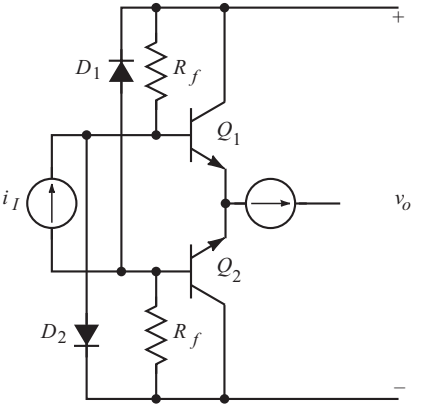
A larger C_f can be used in the circuit above, in which a voltage divider is formed to drive C_f. If the divider Thevenin resistance is negligible, then the loop gain is reduced by the divider. The voltage gain across C_f is also reduced from $(A_v + 1)$, so that the Miller capacitance is

$$C_{in} = C_f \cdot (A_v + 1) \cdot \left(\frac{R_2}{R_1 + R_2} \right)$$

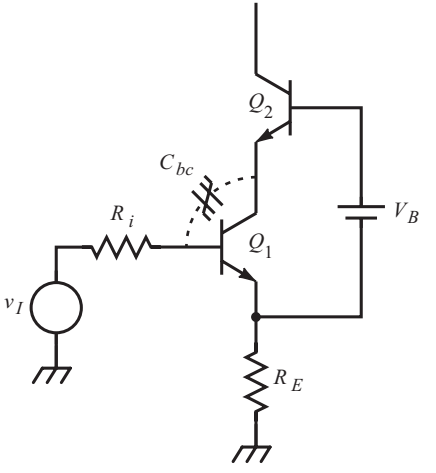
If $R_1 \| R_2$ is not negligible, various poles and zeros appear. A divider in the feedback path can similarly be used to reduce R_f.



Nonlinear dynamic compensation is often necessary in circuits with large dynamic ranges. Transistor parameters change with waveforms, causing otherwise well-compensated circuits to show errors in their response. The *b-c* capacitance of BJTs varies nonlinearly with v_{BC}. Some circuits compensate for ΔC_{bc} with another *b-c* junction of a similar transistor. As shown above, emitter compensation is applied by reverse-biasing the *b-c* junction of Q_2 with V_B to track the C_{bc} variation of Q_1.



The tracking is more easily accomplished in a differential amplifier, as shown above, where the opposing sides of the shunt-feedback amplifier vary by $\pm\Delta v_{BC}$ around the same bias point.

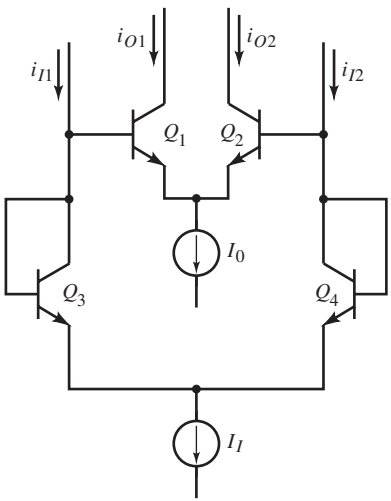


In this circuit, C_{bc} is bootstrapped from the emitter of Q_1. As v_I varies, v_{E2} follows it, keeping the voltage across C_{bc} constant. However, now that v_{B2} varies, the effect is transferred to C_{bc} of Q_2. But if V_{CB} of Q_2 is large relative to that of Q_1, its C_{bc} is smaller and varies less.

The basic technique of adjusting a semiconductor parameter with a fixed quantity for dynamic compensation is used in high-speed IC circuits, in which compensation of internal nodes is feasible only by means of an external adjustment outside the high-frequency signal path.

The Gilbert Gain Cell and Multiplier

In the mid-1960s, Gilbert discovered another basic amplifier technique with wide bandwidth and high linearity. Instead of increasing circuit complexity to compensate for circuit error, the technique is based on the accurate logarithmic function of *b-e* junctions and good junction matching.



A differential current mirror, or *translinear cell*, based on the one in "Current-Input and Feedback Amplifiers," is shown above; it consists of two pairs of emitter-coupled transistors. The input pair is connected as diodes, with input current $i_{I2} - i_{I1}$. If the areas are matched, then I_S is the same for all transistors.

The differential voltage across the bases of Q_1 and Q_2 is the same as across Q_3 and Q_4, or

$$\Delta v_{BE} = V_T \ln\left(\frac{i_{O2}}{i_{O1}}\right) = V_T \ln\left(\frac{i_{I2}}{i_{I1}}\right) \Rightarrow \left(\frac{i_{O2}}{i_{O1}}\right) = \left(\frac{i_{I2}}{i_{I1}}\right)$$

In other words, the ratio of output currents equals the ratio of input currents, assuming that BJT $\alpha = 1$. As this result is a current gain, it cannot exceed β but is very linear up to gains near β. Linearity is reduced by junction area mismatch and ohmic base and emitter resistance but is much better than for comparable diff-amps with emitter resistance. Also, because Q_3 and Q_4 have $V_{CE} = V_{BE} \cong 0.8$ V, additional voltage drop across the collector ohmic resistance can cause V_{CE} to approach zero, causing diode error. The bandwidth is limited by BJT f_T.

The differential current gain is found by expressing the current ratios of ΔV_{BE} in the form

$$\frac{i_{O2}}{i_{O1}} = \frac{(i_{O2} + i_{O1}) + (i_{O2} - i_{O1})}{(i_{O2} + i_{O1}) - (i_{O2} - i_{O1})} = \frac{i_{I2}}{i_{I1}} = \frac{(i_{I2} + i_{I1}) + (i_{I2} - i_{I1})}{(i_{I2} + i_{I1}) - (i_{I2} - i_{I1})}$$

Solving for the current gain,

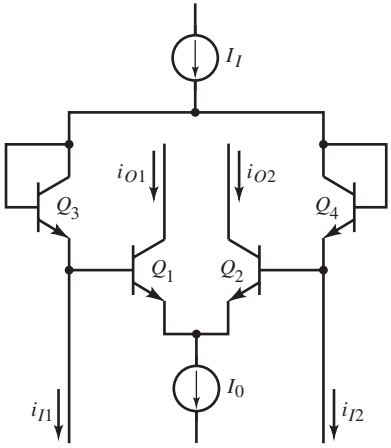
$$A_i = \frac{i_O}{i_I} = \frac{i_{O2} - i_{O1}}{i_{I2} - i_{I1}} = \frac{i_{O1} + i_{O2}}{i_{I1} + i_{I2}} = \frac{I_0}{I_I}$$

The significance of this result is that ratios of input and output variables can be expressed as ratios of their sums and differences. In general,

$$\frac{a}{b} = \frac{c}{d} \Rightarrow \frac{a-b}{a+b} = \frac{c-d}{c+d}$$

A gain-inverting translinear cell is shown below. The outer pair Δv_{BE} is negative that of the basic translinear cell, and has current ratios

$$\frac{i_{O2}}{i_{O1}} = \frac{i_{I1}}{i_{I2}}$$

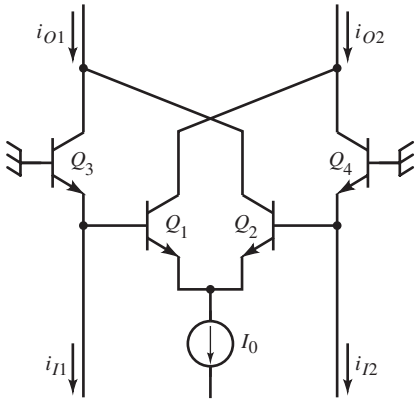


and

$$A_i = -\frac{I_0}{I_I}$$

Complementary pairs of inner and outer junctions are also possible, though I_S matching is more difficult.

The *Gilbert gain cell* follows from the inverting translinear cell by connecting Q_3 and Q_4 as CB transistors and cross-coupling their collectors with Q_1 and Q_2 for additive outputs.



The current gain of the outer pair is $\alpha \cong 1$, apart from the additional current gain of the inner pair. The inner-pair gain is that of A_i, or

$$\frac{i_{C2} - i_{C1}}{i_{I2} - i_{I1}} = -\frac{I_0}{I_I}$$

Approximating α as one, the Gilbert cell gain is

$$A_i = \frac{i_{O2} - i_{O1}}{i_{I2} - i_{I1}} = \frac{(i_{C1} + i_{I2}) - (i_{C2} + i_{I1})}{i_{I2} - i_{I1}} = -\frac{i_{C2} - i_{C1}}{i_{I2} - i_{I1}} + 1 = \frac{I_0 + I_I}{I_I}$$

Two currents such as i_{I1} and i_{I2} can be expressed as fractions of I_I. Let the "modulation index" be the fraction x:

$$i_{I1} = x \cdot I_I \quad \text{and} \quad I_{I2} = (1 - x) \cdot I_I$$

Their sum is still I_I. Applying the current ratios to the Gilbert cell, their ratio is

$$\frac{i_{I1}}{i_{I2}} = \frac{x}{1 - x} = \frac{i_{C2}}{i_{C1}}$$

Thus,

$$i_{C2} = x \cdot I_0 \quad \text{and} \quad i_{C1} = (1 - x) \cdot I_0$$

The Gilbert cell output currents are then

$$i_{O2} = i_{C1} + i_{I2} = (1 - x) \cdot (I_0 + I_I)$$
$$i_{O1} = i_{C2} + i_{i1} = x \cdot (I_0 + I_I)$$

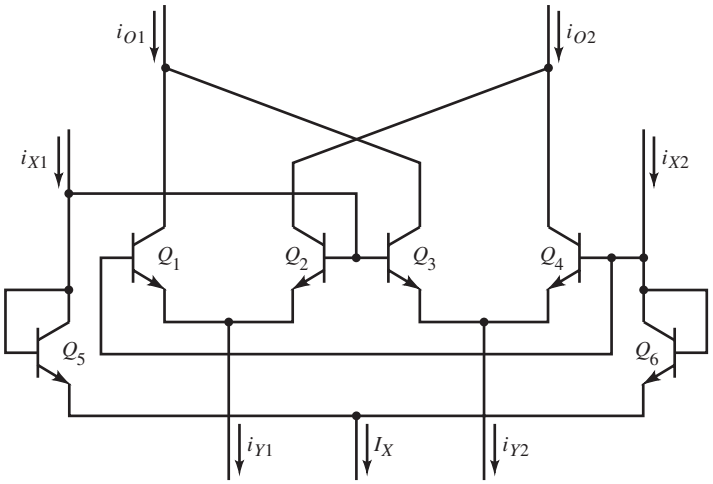
The dynamic range is $0 \le x \le 1$; unlike the cascomp, I_0 and I_I are fully used as signal currents over this range. Gilbert cells are easily cascaded with little bandwidth loss due to interstage coupling.

The differential current mirror and Gilbert cell have important uses beyond current amplification. Equations for noninverting A_i and inverting A_i suggest

that they also function as multipliers. Because I_0 and I_I can both be varied, the current gain can be changed; I_0 multiplies the gain and I_I divides it. If I_0 is not held constant but allowed to be the function i_Y, then the output from the noninverting A_i is

$$i_O = \frac{i_I \cdot i_Y}{I_I}$$

A divider is similarly realized if I_I is a waveform instead. Given that i_I is differential, it is a bipolar waveform; i_Y is unipolar, resulting in a two-quadrant multiplier. Two differential pairs result in full four-quadrant multiplication.



The collectors of the pairs are cross-coupled. When $i_{Y1} = i_{Y2}$, the BJT pairs have equal but opposite gains, and their output currents cancel. At either range extremum, i_{Y1} or $i_{Y2} = 0$, and only one pair amplifies. Within the range, the proportions of collector currents are set by the ratio of i_{Y1} and i_{Y2}.

The four-quadrant multiplier output is derived in terms of the inputs

$$i_X = i_{X2} - i_{X1} \quad \text{and} \quad i_Y = i_{Y2} - i_{Y1}$$

where

$$i_{X1} + i_{X2} = I_X \quad \text{and} \quad i_{Y1} + i_{Y2} = I_Y$$

Define the ratio of input currents to be

$$i_{X2} = x \cdot I_X, \quad i_{X1} = (1-x) \cdot I_X$$

Then from the input equations,

$$i_X = x \cdot I_X - (1-x) \cdot I_X = (2 \cdot x - 1) \cdot I_X$$

The output currents are

$$i_{O1} = i_1 + i_3 = x \cdot i_{Y1} + (1-x) \cdot i_{Y2} = -x \cdot i_Y + i_{Y2}$$

and

$$i_{O2} = i_2 + i_4 = (1-x) \cdot i_{Y1} + x \cdot i_{Y2} = x \cdot i_Y + i_{Y1}$$

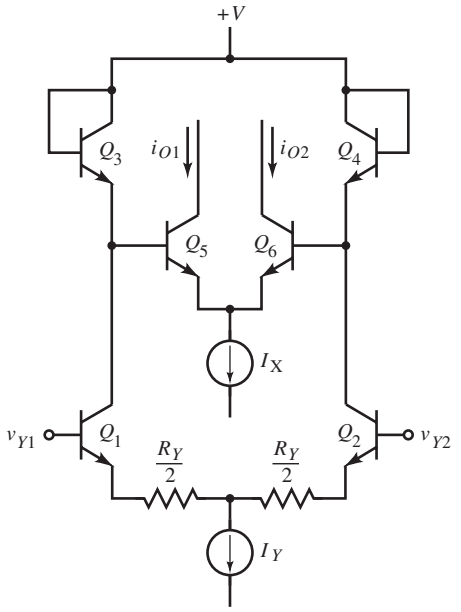
The differential output current is

$$i_O = i_{O2} - i_{O1} = 2 \cdot x \cdot i_Y - i_Y = (2 \cdot x - 1) \cdot i_Y = \frac{i_X}{I_X} \cdot i_Y$$

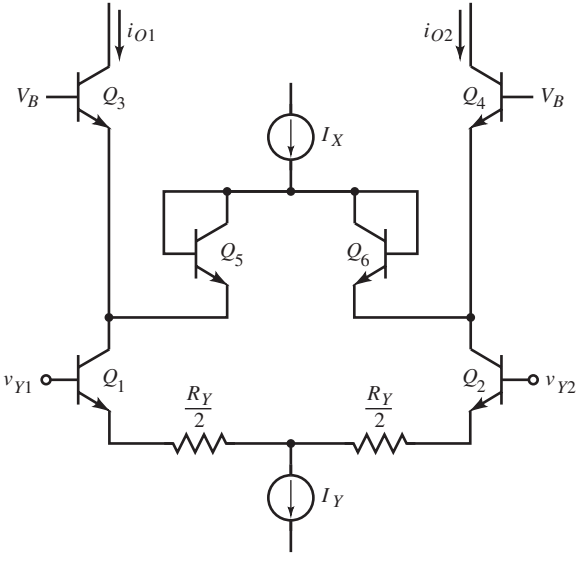
The output current is a fraction of the total output current I_Y, or

$$\frac{i_O}{I_Y} = \frac{i_X}{I_X} \cdot \frac{i_Y}{I_Y}$$

The result is ideal; i_X and i_Y are multiplied and scaled by I_X and I_Y. The fractional output depends on I_X, and it too can be varied to provide division without affecting the output scaling. As for imperfections, v_{BE} mismatch causes even harmonic distortion, junction resistance causes odd harmonics and thermal noise, and C_{bc} of $Q_1 - Q_4$ are a feedthrough path for high-frequency waveform leakage, causing appreciable error around zero output. The above circuit is the multiplier core. To multiply voltages, diff-amp transconductance amplifiers drive the multiplier and ratio I_X and I_Y by their inputs.



Two-quadrant multipliers are used as *voltage-controlled amplifiers* (VCAs), in which only magnitude (not phase) of the signal is controlled. A typical two-quadrant multiplier is shown above, using an inverting translinear cell. The fractional output $i_O/I_X = i_Y/I_Y$.



An improved multiplier topology is the two-quadrant *controlled-cascode multiplier* cell shown above. The input v_Y generates i_Y as collector currents of a differential cascode amplifier. The translinear input junctions are now the *b-e* junctions of the CB stage instead of diodes. Consequently, i_Y becomes the output current. I_X drives the other translinear pair of junctions, Q_5 and Q_6, and is proportioned by the CB currents so that xI_X subtracts from xI_Y. The outputs are the differences

$$i_{O1} = i_{Y1} - i_{X1} = (1-x)\cdot I_Y - (1-x)\cdot I_X = (1-x)\cdot(I_Y - I_X)$$
$$i_{O2} = i_{Y2} - i_{X2} = x\cdot I_Y - x\cdot I_X = x\cdot(I_Y - I_X)$$

Thus the differential output is

$$i_O = (2x-1)\cdot(I_Y - I_X)$$

and the fractional output is

$$\frac{i_O}{I_Y} = (2x-1)\cdot\left(1 - \frac{I_X}{I_Y}\right)$$

As a two-quadrant multiplier, I_X varies from 0 to I_Y, whereas the input v_Y varies x over the range from 0 to 1. The scaling effect for I_X is reversed since an increase in I_X decreases the output.

This topology has less high-frequency feedthrough because of the CB isolation of C_{cb}. In this circuit, the fractional output varies inversely with input-current scaling I_Y. In this equation, the sensitivity of i_O to I_Y is $I_Y/(I_Y - I_X) > 1$, and scaling is more sensitive to I_Y. This topology also has less nonlinearity from area mismatch and junction resistance. The Analog Devices AD539 is a dual-controlled cascode multiplier of Gilbert's design, with an i_Y gain-independent bandwidth of 60 MHz, I_X bandwidth of 5 MHz, and less than 1% nonlinearity.

The four-quadrant cascode multiplier (shown below) using Gilbert gain cells was developed by Metz. The BJT pair, Q_1, Q_2, forms a Gilbert cell with Q_5, Q_6. The second Gilbert cell, with Q_3, Q_4, is reversed from the first cell. When $i_X = 0$, both cells have the same gain and their outputs cancel, leaving i_Y unattenuated. At the extremes, only one cell is on and either adds or subtracts from i_Y.

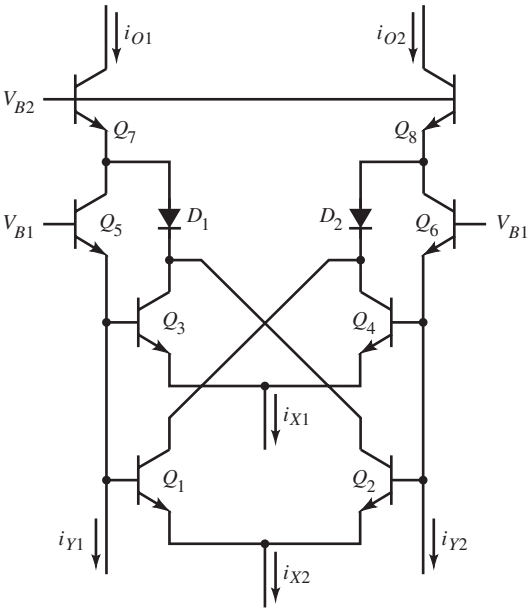
The output range is from zero to twice i_Y when $i_{X2} = I_X = I_Y$. The outputs in general are

$$i_{O1} = i_{Y1} + x \cdot i_{X1} + (1-x) \cdot i_{X2}$$

$$i_{O2} = i_{Y2} + x \cdot i_{X2} + (1-x) \cdot i_{X1}$$

For the translinear cell,

$$\frac{i_Y}{I_Y} = \frac{i_X}{I_X} = 2 \cdot x - 1$$



The output is then

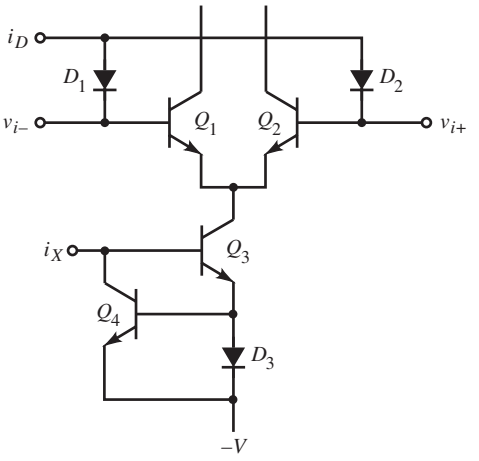
$$i_O = i_Y + (2x-1) \cdot i_X = i_Y \cdot \left(1 + \frac{i_X}{I_Y}\right)$$

The multiplier functions as a controlled-gain amplifier of i_Y, with current gain

$$A_i = \frac{i_O}{i_Y} = 1 + \frac{i_X}{I_Y}$$

As i_X is bipolar, the gain is one when i_X is zero.

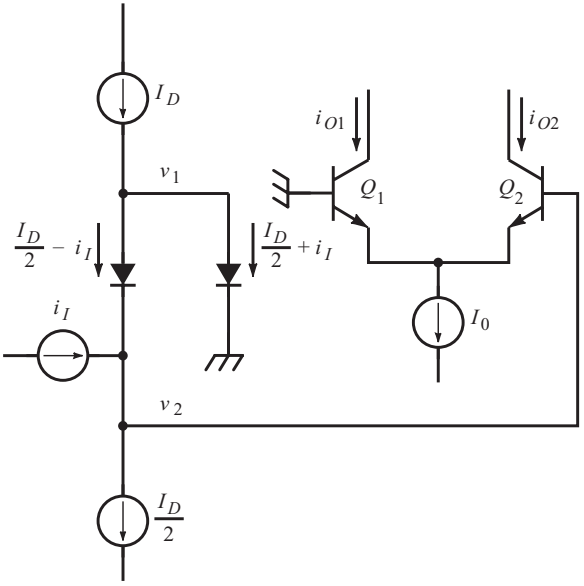
This multiplier has the same advantages as the previous one: high linearity with junction mismatch and resistance. Both also have low thermal distortion. The diodes D_1 and D_2 make the operating-point *c-b* voltage of the Gilbert-cell transistors the same as for the CB. D_1 reduces v_{CE2} and v_{CE3} to compensate for v_{BE5}. If $I_X = I_Y$, the emitter currents are also equal, and thermal balance results. The additional CB of Q_7 and Q_8 provides feedthrough isolation for the Gilbert cells.



A translinear cell is also found in the input of the LM13600, LM13700 dual transconductance op-amps (shown above) employing "linearizing diodes." A method of biasing the translinear cell is shown below.

Balanced currents of $I_D/2$ flow through diodes D_1, D_2, resulting in zero input voltages to the diff-amp. Input current i_I upsets the balance so that the junction voltages around the translinear-cell loop are

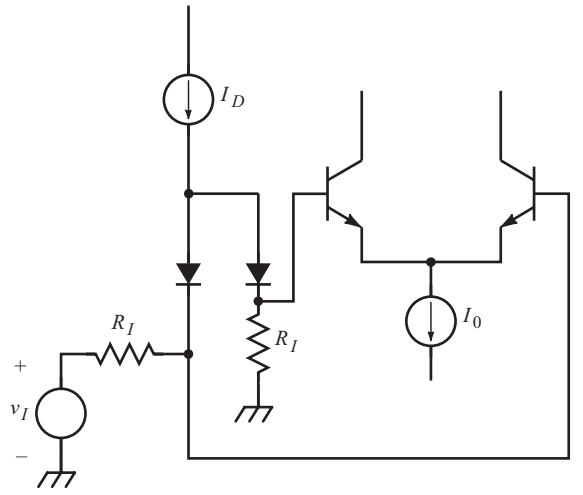
$$v_2 - v_1 = V_T \ln\left(\frac{I_D/2 + i_I}{I_D/2 - i_I}\right) = V_T \ln\left(\frac{i_{O2}}{i_{O1}}\right) = V_T \ln\left(\frac{I_0 + i_O}{I_0 - i_O}\right)$$



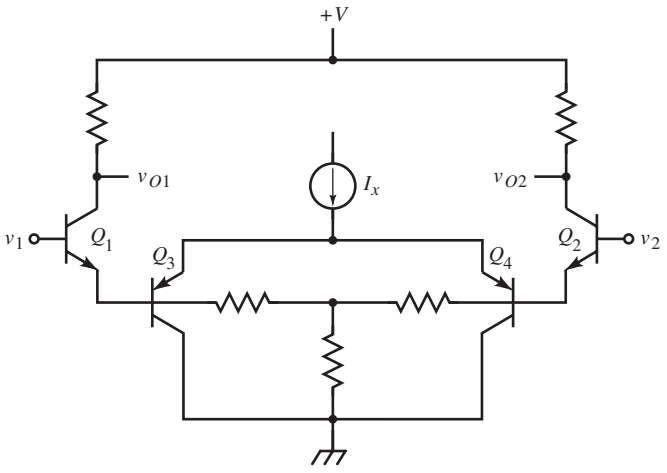
which simplifies to

$$i_O = i_{O2} - i_{O1} = 2 \cdot \left(\frac{I_0}{I_D} \right) \cdot i_I, \quad |i_I| < I_D/2$$

This is a single-ended-to-differential translinear cell. Because the matching of bias-current sources is harder to achieve than matching resistors, the voltage-input version shown below is easier for discrete design.



When amplification must be linear but amplitude control need not be, the linearizing diodes of the translinear cell can be discarded. The linear input is I_0, and VCA control input is v_I.



A class of variable-gain circuits with minimal control bandwidth and linearity requirements are *automatic gain control* (AGC) and *compandor* (compressor-expander) circuits, used in radios and broadcast audio equipment to maintain constant loudness. A variety of variable-gain circuits exist such as the one shown above. By increasing I_X, r_e in Q_3 and Q_4 are reduced and gain increased.

FET forms of the Gilbert multiplier have been studied. Because the FET is a square-law rather than an exponential device, the simplicity of high-performance multiplication has not been achieved. The *quarter-square multiplier* is based on square-law devices and the quarter-square formula

$$\frac{1}{4}\cdot\left[(x+y)^2-(x-y)^2\right]=x\cdot y$$

This technique requires three summations, two squarings, and scaling and has been implemented with MOSFETs.

PROGRAMMABLE-GAIN AMPLIFIERS

Amplifier applications having a wide range of input voltages cause amplifier range and gain constraints to conflict. For small waveforms, a high gain is needed to produce a full-scale (fs) output, but for large waveforms, the same gain drives the amplifier out of range. If the scaling accuracy is not critical, a logarithmic amplifier or compander can solve this problem though both have a nonlinear transfer curve; for small waveforms the gain is high but decreases with amplitude. The output, of course, is distorted by the log-amp or compander gain variation.

For linear systems, a simple solution is to change the amplifier gain to match the waveform amplitude. This minimizes amplifier error and keeps the waveform within the amplifier linear range. An input x_i is subject to scaling (gain) and offset errors in the amplifier, so that the output referred to the input is

$$x_o = (1 + \varepsilon_s) \cdot x_i + \varepsilon_o$$

where ε_s is scaling error, and ε_o is offset error. The amplifier error is defined as

$$\varepsilon \equiv \frac{x_o - x_i}{x_i} = \frac{\varepsilon_s \cdot x_i + \varepsilon_o}{x_i} = \varepsilon_s + \frac{\varepsilon_o}{x_i}$$

The error has two terms: ε_s is the *range error* and ε_o / x_i is the *reading error*. Accuracy specifications are often given as ε_s because it is a fraction of the fs amount for all x_i. As x_i approaches zero, the fixed offset error becomes an increasingly large fraction of the total error and varies inversely with x_i. The zero-scale (zs) end of the range partly depends on how much error can be tolerated.

Amplifier gain can be changed over a continuous range using a multiplier. Usually it is simpler to switch gain-determining elements for a small number of gain settings. These *gain-switched* or *programmable-gain* amplifiers are commonly found on measurement instruments, such as data-acquisition systems and oscilloscopes, for which a very wide range of input values is allowed. The vertical sensitivity control on oscilloscopes typically spans a range from 2 mV/div to 50 V/div. (The *div* unit is a division of vertical deflection on the CRT screen graticule; eight divisions is standard for full scale.) Instrument designers have

standardized on a 1-2-5 sequence of gain settings within a decade. These discrete settings fall short of matching the amplifier fs to the input, and some additional error due to ε_o is accepted.

For a given gain setting, the range is from zs to fs for input quantity x_i, or

$$x_{zs} = \frac{x_{fs}}{a} \le x_i \le x_{fs}, \quad a > 1$$

where x_{fs} is the maximum x_i at fs and the minimum x_i at zs is x_{zs}, the fraction $1/a$ of fs. Thus a defines the extent of the range. Now the average error over the range is

$$\bar{\varepsilon} = \frac{1}{x_{fs} - x_{fs}/a} \int_{x_{fs}/a}^{x_{fs}} \left(\varepsilon_s + \frac{\varepsilon_o}{x_i} \right) \cdot dx_i = \left(\frac{a \cdot \ln a}{(a-1) \cdot x_{fs}} \right) \cdot \varepsilon_o + \varepsilon_s$$

for $a > 1$. For amplifiers with continuously adjustable gain, for all x_i, gain is adjusted so that x_i is x_{fs}. The ratio of offset errors for the discrete to continuous cases is

$$\frac{\varepsilon_{od}}{\varepsilon_{oc}} = \left(\frac{a}{a-1} \right) \cdot \ln a$$

For a range of a-to-1, the larger a is, the larger the error ratio, to the disadvantage of discrete-gain settings.

In computer-based data-acquisition systems, a 1-2-5 gain sequence is not necessary; the computer can rescale the measurements. The question then arises as to the advantage of regular gain settings over the 1-2-5 settings. For minimum offset error, many ranges with small a are desired. If we assume that at fs, ε_s can be nulled, then for three gain settings per decade, the 1-2-5 scheme has the following a and normalized offset error:

x_{fs}	x_{zs}	a	$(\bar{\varepsilon} - \varepsilon_s)/\varepsilon_o$
2	1	2.0	1.39
5	2	2.5	1.53
10	5	2.0	1.39

The maximum normalized error is 1.53. For three regularly spaced settings covering a decade,

$$a = 10^{1/3} \cong 2.15$$

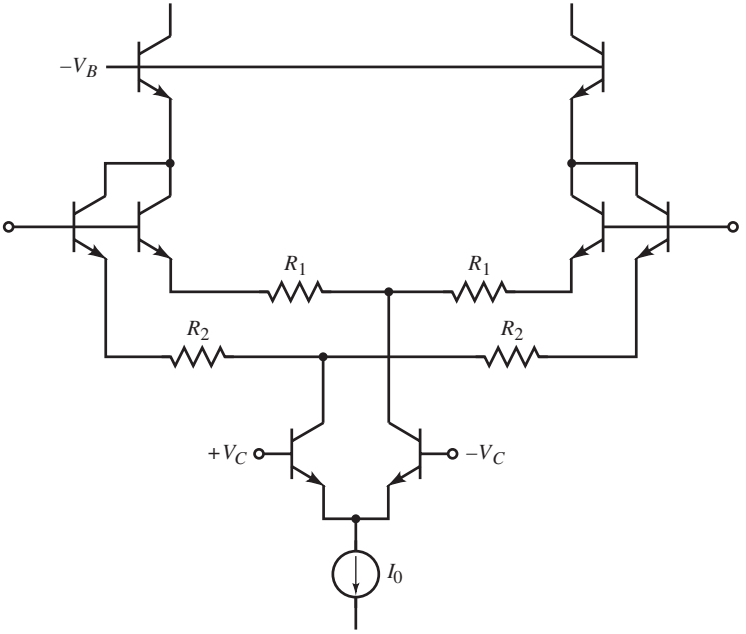
and maximum normalized error is 1.43. Therefore, the improvement of regular settings over 1-2-5 settings is 1.43/1.53 = 0.93, for a 7% improvement. This is usually not significant, and the 1-2-5 sequence appears to be well chosen.

A major consideration in gain-switched amplifiers is the imperfection of the switches. Reed relays are closest to ideal electrically but are slow, bulky, and power-intensive. Solid-state analog switches are used in all but the most demanding applications, with their parasitic series, shunt (leakage) resistances, and shunt capacitance.

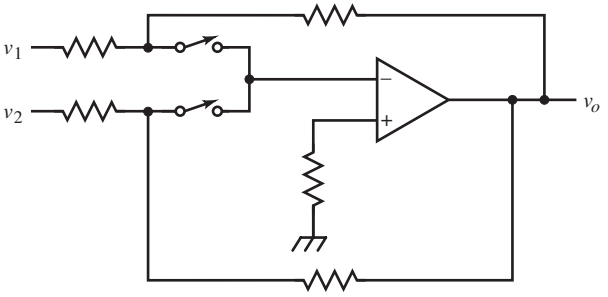
Diodes, BJTs, JFETs, and MOSFETs function as switches of voltage or current. JFETs (including those of biFET implementation) have constant on-resistance, r_{on}, of typically 100 Ω over the input voltage range. MOSFET switches have lower r_{on}, typically 10 to 20 Ω, with large Δr_{ON}. Discrete power MOSFETs have r_{on} values of less than 100 mΩ but with correspondingly large shunt capacitances. MOSFETs have a parasitic diode between the body and drain. Three-terminal MOSFETs have the body connected to the source, and current flows through this diode in the opposite direction to the gate-controlled flow. Two MOSFETs in series, back to back, form a bipolar switch. Because of MOSFET r_{ON} variation, n- and p-channel devices are used together as CMOS switches because their variations tend to cancel. The combined r_{on} still varies more than for a JFET switch and peaks at the midrange input voltage.

Discrete diodes and BJTs are sometimes used as gain switches, as in the BJT amplifier below. The diff-amp emitter resistors are different ($R_1 \neq R_2$), and gain is selected by enabling one diff-amp. The diff-amps are switched by switching the emitter-current source to the selected diff-amp. This scheme is extendable by adding more diff-amp pairs and current switches.

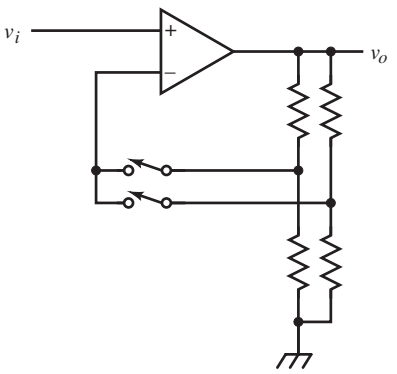
Of the two switching modes, voltage and current, parasitic elements in switches mainly determine which mode to use. A switch with high r_{on} is a poor choice for switching in a voltage divider; r_{on} adds to the divider resistance, changing the



attenuation. Except for voltage range (compliance) limitations on current sources, r_{on} does not affect current-mode switching.



An input source is selected by current-switching into the virtual ground of an inverting op-amp, as shown above. Current-mode switching also reduces the effects of shunt capacitance by minimizing voltage variation.



Sometimes voltage-mode switching is necessary and can be error-free if the switch load is an open circuit-that is, switch current (and voltage drop) is zero. One example is the gain-switching, noninverting op-amp, shown above. To keep the op-amp loop from momentarily opening during switching, the switches should be make-before-break types.

CLOSURE

Several amplifier concepts based on current amplification get around gain-bandwidth limitations, and the use of composite topologies can improve performance. The translinear cell opens a wide range of possibilities for improved analog waveform processing. Here, we investigated amplification and multiplication, but function generation is another use for it. The study of amplification has been our focus to this point, but with the introduction of gain switching, nonlinear analog circuit functions become a consideration, taken up in *Designing Waveform-Processing Circuits.*

References

Anonymous (1981, January 7). Characteristics of Flat Cable. inset, *EDN*, pp. 110–111.

Anonymous (1985, March). A New Approach to Op Amp Design. Comlinear Corp. Application Note 300-1.

Balph, T. (1972). Interconnection Techniques for Motorola's MECL 10,000 Series Emitter Coupled Logic. Motorola Application Note AN-556.

Barna, A. (1970). *High-Speed Pulse Circuits*. New York: Wiley-Interscience.

Battjes, C. R. (1973, December). A Wide-Band High-Voltage Monolithic Amplifier. *IEEE JSSC*, Vol. SC-8, No. 6, pp. 408–413.

Battjes, C., Hofer, B., and Addis, J. (1982). *Amplifier Frequency and Transient Response* (*AFTR*) course notes, Tektronix, Inc., Beaverton, OR.

Cherry, E. M. and Hooper, D. E. (1968). Amplifying Devices and Low-Pass Amplifier Design. San Francisco: Wiley.

Cochrun, B. L. and Grabel, A. (1973, January). A Method for the Determination of the Transfer Function of Electronic Circuits. *IEEE Transactions on Circuit Theory*, vol. CT-20, no. 1, pp. 16–20.

Dash, G., editor (1988). *Compliance Engineering 1988*, vol. 4. Acton, MA: Compliance Engineering.

Davis, A. M. (1979, February 20). Analyze Active-Network Responses without Complex Manipulations. *EDN*, pp. 109–112.

Doubrava, L. (1978, May). High-Frequency Bypass and Decoupling Design. *Proceedings of Powercon,* vol. 5, pp. H1-1–H1-11.

Doubrava, L. (1979, March 5). Proper Handling of Voltage Spikes Safeguards Circuit Designs. *EDN,* pp. 83–87.

Franco, S. (1989, January 5). Current-Feedback Amplifiers Benefit High-speed Designs. *EDN*, pp. 161–172.

Ghausi, M. S. (1965). *Principles and Design of Linear Active Circuits*. New York: McGraw-Hill.

Nordholt, E. H. (1983). *Design of High-Performance Negative-Feedback Amplifiers*. Amsterdam: Elsevier.

Gilbert, B. (1968a, December). A New Wide-Band Amplifier Technique. *IEEE JSSC*, Vol. SC-3, No. 4, pp. 353–365.

Gilbert, B. (1968b, December). A Precise Four-Quadrant Multiplier with Subnanosecond Response. *IEEE JSSC*, Vol. SC-3, No. 4, pp. 365–373.

Gilbert , B., Kitchin, C. and Weigel, K. (1984, October 18). Build Fast VCAs and VCFs with Analog Multipliers. *EDN*, pp. 289–299.

Gross, W. (1988, October 27). Use NPN and PNP Devices Effectively in Semicustom Arrays. *EDN*, pp. 297–308.

Hansford, A. (1987, November 26). Use of Transimpedance Amplifiers Minimizes Design Tradeoffs. *EDN*, pp. 205–214.

Hofer, B. (*ca.* 1975). Harmonic Distortion Estimation Notes. Unpublished derivations, Beaverton, Oregon.

Jarva, W. (1977, March 15). Design EMI Shielding More Accurately. *Electronic Design*, vol. 6, pp. 88–91.

Lorenz, R. D. (1986). Synthesis of State Variable Controllers for Industrial Servo Drives. *Conference on Applied Motion Control 1986*, p. 247ff.

McCully, D. A. (1980, January). Preservation of Sensor Data Integrity in Noisy Environments. Computer Products, Inc. MC1081, Fort Lauderdale, FL.

McElvein, D. (1980, October 7–9). Common Misconceptions in the Use of EMI Gaskets. *IEEE Symposium on EMC*, Baltimore, MD; reprint from Chomerics, Inc., Woburn, MA.

Morrison, R. (1983, January–March). Grounding in Instrumentation Systems. *EMC Technology*, pp. 50–52.

Metz, A. (1983, August 4). Circuit-Design/Process Combo Speeds Horizontal-Amp Slewing. *EDN*, pp. 173–180.

National Semiconductor Corp. (1986). Super Matched Bipolar Transistor Pair Sets New Standards for Drift and Noise. *Linear Applications Databook*, Application Note AN-222, pp. 517. Santa Clara, California.

Neuman, W. *Switching Power Supplies, Radio Frequency Interference, & Power Line Interference Filters*. Chicago, IL: Corcom, Inc.

Oates, E. R. (1977, January 4). Good Grounding and Shielding Practices. *Electronic Design*, vol. 1, pp. 110–112.

Oliver, B. M. (1971, February). Distortion in Complementary-Pair Class-B Amplifiers. *Hewlett-Packard Journal*, vol. 22, no. 6, pp. 11–16.

Ott, H. (1983, January–March). Ground: A Path for Current Flow. *EMC Technology*, pp. 44–48.

Pena-Finol, J. S. and Connelly, J. A. (1987, December). A MOS Four-Quadrant Analog Multiplier Using the Quarter-Square Technique. *IEEE JSSC*, Vol. SC-22, No. 6, pp. 1064–1073.

Quinn, P. (1989, January). Feedforward Amplifier. Tektronix patent #4,146,844.

Rosenstark, S. (1986). *Feedback Amplifier Principles*. London: Macmillan, pp. 67–77.

Saucedo, R. and Schiring, E. (1968). *Introduction to Continuous and Digital Control Systems*. London: Macmillan, pp. 273, 275.

Ryan, A. and Scranton, T. (1984). D-C Amplifier Noise Revisited. *Analog Dialogue*, vol. 18, no. 1, pp. 3–10.

Sansen, W. M. C. and Meyer, R. G. (1974, August). An Integrated Wide-Band Variable-Gain Amplifier with Maximum Dynamic Range. *IEEE JSSC*, Vol. SC-9, No. 4, pp. 159–166.

Severinsen, J. (1975, February 5). Designer's Guide to EMI Shielding, Part I. *EDN*, pp. 47–51.

Simpkins, S. and Gross, W. (1983, December). Cascomp Feedforward Error Correction in High Speed Amplifier Design. *IEEE JSSC*, Vol. SC-18, No. 6, pp. 762–764.

Soliman, A. M. (1981). Instrumentation Amplifiers with Improved Bandwidth. *IEEE Circuits and Systems Magazine*, vol. 3, no.1, pp. 7–9.

Starič, P. (1989). Application of T-Coil Transistor Interstage Coupling in Wideband/ Pulse Amplifiers. *Electrotechnical Review*. ELVEA 2 57 (1990) 3. 143–152 (written in Slovene with English summary).

Starič, P. (1991). Title of thesis. Ph.D. diss., University of Ljubljana, Slovenia (written in Slovene).

Starič, P. (1994). Wideband Cascode Amplifier with Emitter Peaking. *Electrotechnical Review*, vol. 61, no. 1–2, pp. 18–28.

Starič, P. and Margan, E. (2006). *Wideband Amplifiers*. New York: Springer.

Vance, E. F. (1983, January–March). Cable Grounding for the Control of EMI. *EMC Technology*, pp. 54–58.

White, D. R. J. and Mardiguian, M. (1985). *EMI Control Methodology and Procedures*, 4th ed. Gainesville, VA: Interference Control Technologies, Don White Consultants, Inc.

Williams, J. (1984, August 9). Monolithic Power-Buffer IC Drives Difficult Loads. *EDN*, pp. 153–159.

Wilson, G. R. (1968, December). A Monolithic Junction FET-N-P-N Operational Amplifier. *IEEE JSSC*, Vol. SC-3, No. 4, pp. 341–348.

Yen, C.-S. (1979, August). Distortion in Emitter-Driven Variable-Gain Pairs. *IEEE JSSC*, vol. SC-14, no. 4, pp. 771–773.

INDEX

C