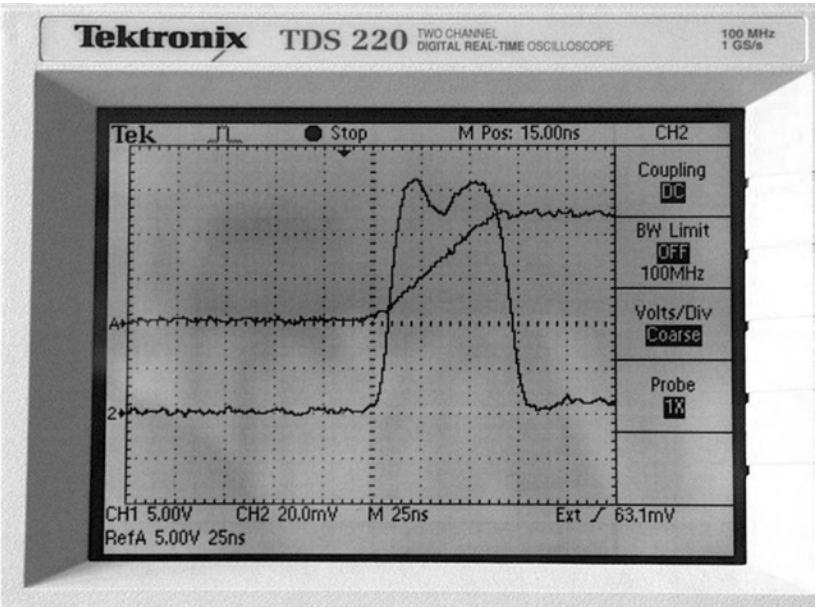
Analog Circuit Design Series Volume 4

Designing Waveform-Processing Circuits



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: 9781891121852
Series ISBN: 9781891121876

Library of Congress Cataloging-in-Publication Data
Feucht, Dennis.
Designing waveform-processing circuits / D. Feucht.
p. cm. – (Analog circuit design series ; v. 4)
Includes bibliographical references and index.
ISBN 978-1-891121-85-2 (pbk. : alk. paper) – ISBN 978-1-891121-87-6 (series)
1. Signal generators. 2. Oscillators, Electric. 3. Signal processing. 4. Electronic circuit design. I. Title.
TK7872.S5.F83 2010
621.3815′33–dc22

2009028291

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 three volumes of this Analog Circuit Design series, *Designing Amplifier Circuits, Designing Dynamic Circuit Response*, and *Designing High-Performance Amplifiers*, are primarily concerned with amplification. This fourth volume widens coverage to other kinds of analog and analog-digital (or "mixed-signal") circuits in two chapters. The first starts with a solid-state coverage of voltage references, mainly bandgap references. Much is also made of various current source circuits, then light coverage of filters, hysteretic switches (Schmitt triggers), clamps and limiters, monostable multivibrators (MMVs) and timing circuits – a mix of circuitry that appears repeatedly in electronics. Analysis of some of this familiar circuitry is taken in directions largely unfamiliar, such as what affects the frequency of astable MMV oscillators. Capacitance and resistance multiplier circuits are less common though exhilarating in how they function. Some oscilloscope-specific circuitry appears; trigger generators also find application in digital synchronizing circuits. Ramp and sweep generators can be used in other kinds of instruments such as mass spectrometers, as can the class of function-generating circuits that include logarithmic and exponential amplifiers and power-series generators. Triangle-wave generators continue a test instrument theme as the core of function generators. Absolute-value or precision rectifier circuits and peak detectors close out the wide range of different circuits with which the competent analog circuit designer must be familiar.

The second and last chapter is mainly about analog-digital conversion, both A/D and D/A, beginning with some characterizing concepts that are then applied to a catalog presentation of DACs followed by ADCs. Voltage-to-frequency converters are also included as a kind of ADC. All of these circuits receive analysis resulting in equations useful for design. The chapter continues with the mathematically oriented theory of time- and frequency-domain sampling theory. I try to present it with an emphasis on what the equations mean rather than to engage in math for its own sake. Sampling circuits follow, and the volume closes with a brief mention of switched-capacitor circuits. While sampling is not covered to the extent that a digital-signal processing (DSP)

textbook would, the basic concepts are laid out as they apply to circuit design. From what is presented here, the reader should be better prepared to access both digital control theory and DSP literature.

Much of what is in this book must be credited in part to others from whom I acquired 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 was 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 Starič, 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

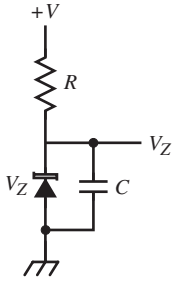
1

Waveform-Processing Circuits

Besides amplification and multiplication, various other waveform-processing functions are a part of the analog circuit design repertoire. Most of these are nonlinear. This chapter surveys a wide variety of waveform-processing circuits, where a *waveform* is an electrical function of time.

VOLTAGE REFERENCES

Stable and accurate voltage sources are needed as references for measurement circuits and power supplies. The Zener diode is a simple voltage-reference device. Although it has been in use a long time, it is still the most stable kind of reference available (other than reference standards such as temperature-controlled batteries or superconducting quantum-effect devices). A simple Zener-based reference is shown below.

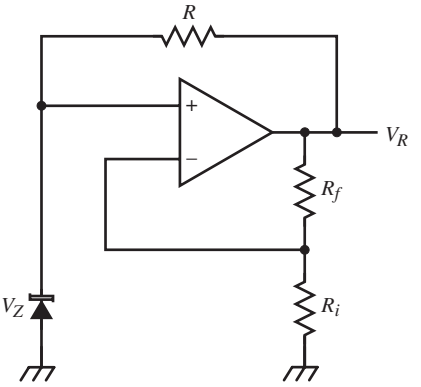


Zener diodes combine two mechanisms, tunneling and avalanche breakdown. Tunneling has a negative temperature coefficient (TC), and avalanche has a positive TC. At around 5 V the mechanism TCs cancel, but the tolerance for

5 V Zeners is not good, making selection necessary for low TC. The TC of Zeners increases reliably with Zener voltage V_Z above about 6 V at about 1 mV/°C per volt, or 0.1%/°C. At a V_Z of 5.6 V, the TC is that of a forward-biased diode, about −2 mV/°C. Placing a diode in series with a 5.6 V Zener results in a zero-TC 6.3 V Zener reference diode. Manufacturers' literature shows that low-TC diodes are around 6.3 V. Low-TC Zeners at higher voltages are also possible by stacking more diodes in series, but tracking makes repeatable manufacture of zero-TC devices more difficult.

Zener diodes are noisy, especially at low currents. Consequently, they are bypassed with a capacitor, as shown above. In integrated circuits (ICs), high-performance Zeners are built below the IC surface as *subsurface Zeners*, which are less noisy because surface effects are eliminated. Lateral ion-implanted Zeners have low-tolerance voltages (typically less than 1%) and are commonly used as references in IC circuits. With a substrate temperature controller on the same chip, monolithic references with 1 ppm/°C are commercially available.

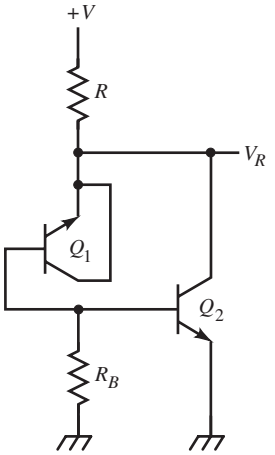
A minimum TC also depends on Zener current I_Z, typically 5 to 10 mA. The circuit shown above is subject to I_Z variation with the voltage supply. The resistor supplying I_Z can be bootstrapped with an operational amplifier (op-amp).



The op-amp output,

$$V_R = \left(\frac{R_f}{R_i} + 1\right) \cdot V_Z$$

supplies a stable Zener current of $(V_R - V_Z)/R$. This reference requires a starting circuit for the Zener, when power is first applied. A simpler bootstrap circuit, shown below, uses a transistor *b-e* junction as the Zener diode, for which V_Z is 6 to 7 V, with a TC of around 2 mV/°C.



The Zener Q_1 is in series with the forward-biased *b-e* junction of Q_2. The combination forms a reference Zener and has a low TC. Q_2 provides shunt regulation to reduce output resistance. Zener diodes have dynamic resistances of around 10 Ω, increasing with V_Z to 100 Ω. Thus, their load regulation is unacceptable for high stability and must be buffered. Another disadvantage to Zeners is that a smaller reference voltage is desired for monolithic 5 V regulators and other devices, such as analog-to-digital converters (ADCs) and digital-to-analog converters (DACs), which operate from 5 V.

A newer kind of voltage reference is based on the temperature characteristics of PN junctions themselves. Junction voltage has a negative TC of about −2 mV/°C. The differential voltage across two matched junctions (as for a differential amplifier [diff-amp]) is

$$\Delta V = V_T \cdot \ln \frac{I_2}{I_1} = \frac{kT}{q_e} \cdot \ln \frac{I_2}{I_1}$$

When the current ratio is held constant, ΔV has a positive, linear TC of

$$\mathrm{TC}(\Delta V) = \frac{k}{q_e} \cdot \ln \frac{I_2}{I_1} = \left(86.17 \frac{\mu\mathrm{V}}{^\circ\mathrm{C}} \right) \cdot \ln \frac{I_2}{I_1}$$

By scaling ΔV and adding it to junction voltage V, the output is

$$V_O = V + g \cdot (\Delta V) = V + g \cdot V_T \ln \frac{I_2}{I_1}$$

where g is the gain required to amplify ΔV so that its TC is opposite that of V. If the junction areas of Q_1 and Q_2 are not equal, the more general form of ΔV is

$$\Delta V = V_T \cdot \ln \frac{J_2}{J_1} = V_T \cdot \ln \left[\left(\frac{I_2}{I_1} \right) \cdot \left(\frac{A_1}{A_2} \right) \right]$$

where J is current density, and A is the *b-e* junction area: $J = I/A$. The TC of V_O is

$$\mathrm{TC}(V_O) = \frac{dV_O}{dT} = \frac{dV}{dT} + g \cdot \frac{d\Delta V}{dT}$$

We can substitute for the TC of V and ΔV, set $\mathrm{TC}(V_O)$ to zero, and solve for the gain g. To find $\mathrm{TC}(V)$, we first differentiate V:

$$\left. \frac{dV}{dT} \right|_I = \frac{d}{dt} \left(V_T \ln \frac{I}{I_S} \right) = -\left(\frac{1}{I_S} \cdot \frac{dI_S}{dT} \right) \cdot V_T + \frac{V}{T} \cong -2 \frac{\mathrm{mV}}{^\circ\mathrm{C}}$$

The expression

$$\mathrm{TC}\%(I_S) = \left(\frac{1}{I_S} \cdot \frac{dI_S}{dT} \right)$$

is the fractional $\mathrm{TC}(I_S)$.

From semiconductor physics, saturation current is

$$I_S = q_e \cdot A \cdot \left(\frac{D_h p_{no}}{L_h} + \frac{D_e n_{po}}{L_e} \right) = k \cdot T \cdot A \cdot n_i^2 \cdot \left(\frac{\mu_h}{L_h N_D} + \frac{\mu_e}{L_e N_A} \right)$$

where D are diffusion coefficients, L are diffusion length constants, p_{no} and n_{po} are the equilibrium minority hole and electron concentrations, respectively, N are the ion doping concentrations, μ are the carrier mobilities, and n_i is the intrinsic carrier concentration, where

$$n_o \cdot p_o = n_i^2$$

and n_o and p_o are the equilibrium electron and hole concentrations, respectively. In the I_S equation, μ and n_i^2 are temperature dependent; the other constants are fixed by geometry, doping, or materials properties. Delving deeper into solid-state physics,

$$n_i^2 \propto T^3 e^{-E_{go}/kT}$$

where E_{go} is the semiconductor bandgap energy, linearly extrapolated to 0 K. For silicon it is

$$E_{go}(\text{Si}) = 1.205 \text{ eV}$$

Because $V_T = kT/q_e$, the previous equation can be expressed as

$$n_i^2(T) \propto T^3 e^{-V_{go}/V_T}$$

I_S also depends on $\mu(T)$. For silicon, $\mu(T) \propto T^{-2.6}$. With this value, it follows from I_S that

$$I_S \propto T \cdot T^{-2.6} \cdot T^3 e^{-V_{go}/V_T} = T^{1.4} e^{-V_{go}/V_T}$$

By taking the derivative and dividing it by I_S,

$$\mathrm{TC\%}(I_S) = \frac{1}{I_S} \cdot \frac{dI_S}{dT} = \frac{V_{go}/V_T + 1.4}{T}$$

$\mathrm{TC\%}(I_S)$ can now be expressed, at 300 K,

$$\mathrm{TC\%}(I_S)\big|_{T=300\mathrm{K}} = 15.53\%/^{\circ}\mathrm{C} + 0.47\%/^{\circ}\mathrm{C} = 16\%/^{\circ}\mathrm{C}$$

The dominant effect on $\mathrm{TC\%}(I_S)$ with temperature is the first term, involving the bandgap. The mobility (second term) affects it only about 3%.

Returning to $dV/dT\big|_I$ and solving,

$$\left.\frac{dV}{dT}\right|_I = -\left(\frac{1}{I_S} \cdot \frac{dI_S}{dT}\right) \cdot V_T + \frac{V}{T} = \frac{V - V_{go} - 1.4 \cdot V_T}{T}$$

At 300 K and $I_S = 10^{-14}$ A, for $I = 1$ mA, $V = 0.655$ V, and

$$\frac{dV}{dT} = -1.95\,\frac{\mathrm{mV}}{^{\circ}\mathrm{C}} \cong -2.0\,\frac{\mathrm{mV}}{^{\circ}\mathrm{C}}$$

While we are calculating TCs, the fractional TC of I is

$$\mathrm{TC\%}(I)\big|_V = \frac{1}{I} \cdot \left.\frac{dI}{dT}\right|_V = \frac{I}{I_S} \cdot \frac{dI_S}{dT} - \frac{1}{T} \cdot \frac{V}{V_T}$$

At 300 K and $V = 0.655$ V, as before,

$$\mathrm{TC\%}(I)\big|_V = 16\%/^{\circ}\mathrm{C} - 8.44\%/^{\circ}\mathrm{C} \cong 8\%/^{\circ}\mathrm{C}$$

It is already established that

$$\mathrm{TC\%}(V_T) = \frac{1}{V_T} \cdot \frac{dV_T}{dT} = \frac{1}{T}$$

At 300 K, $\mathrm{TC\%}(V_T) = 0.33\%/^{\circ}\mathrm{C}$.

Returning to TC(V_O), we must yet find

$$\frac{d\Delta V}{dT} = \frac{V_T}{T} \cdot \ln \frac{J_2}{J_1}$$

Substituting $dV/dT|_I$ and the above equation into TC(V_O),

$$\left.\frac{dV_O}{dT}\right|_I = \frac{V}{T} - \left(\frac{1}{I_S} \cdot \frac{dI_S}{dT}\right) \cdot V_T + g \cdot \ln\left(\frac{J_2}{J_1}\right) \cdot \left(\frac{V_T}{T}\right)$$

From $dV/dT|_I$, this can also be expressed in V_{go} as

$$\left.\frac{dV_O}{dT}\right|_I = \frac{V_T \cdot (\ln(I/I_S) + g \cdot \ln(J_2/J_1) - 1.4) - \mathrm{V_{go}}}{T}$$

For TC(V_O) = 0, the required gain is

$$g = \frac{(1/I_S) \cdot (dI_S/dT) \cdot T - (V/V_T)}{\ln(J_2/J_1)} = \frac{V_{go}/T_T - (\ln(I/I_S) - 1.4)}{\ln(J_2/J_1)}$$

The constraint on achieving a zero TC is that I be held constant.

Example: Bandgap Reference Design

The CA3086 bipolar junction transistor (BJT) array has $I_S = 10^{-15}$ A at 300 K. Let I be 1 mA. Then $V = 0.715$ V, and from $dV_O/dT|_I$,

$$\frac{dV_O}{dT} \cong 2.38\,\frac{\mathrm{mV}}{^\circ\mathrm{C}} - 4.14\,\frac{\mathrm{mV}}{^\circ\mathrm{C}} + \left(86.17\,\frac{\mu\mathrm{V}}{^\circ\mathrm{C}}\right) \cdot g \cdot \ln\frac{J_2}{J_1}$$

The first two terms have a combined TC of −1.76 mV/°C. The gain required to null this TC is, from the formula for g,

$$g = \frac{(0.16/\mathrm{K}) \cdot (300\ \mathrm{K}) - (0.715\ \mathrm{V})/(25.87\ \mathrm{mV})}{\ln(J_2/J_1)} = \frac{20.4}{\ln(J_2/J_1)}$$

Finally, from V_O,

$$V_O = 0.175\,\text{V} + (20.4)\cdot V_T = 0.715\,\text{V} + 0.527\,\text{V} = 1.242\,\text{V}$$

The result for V_O is curiously close to V_{go}. Substituting g into V_O, then

$$V_O = V_{go} + 1.4V_T = 1.242\ \text{V}$$

A concise expression for V_O, in terms of g, follows from substituting ΔV into V_O:

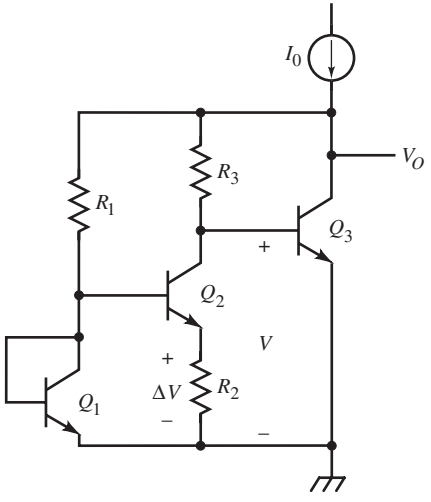
$$V_O = V_T \cdot \ln\left(\frac{I}{I_S}\right) + g \cdot V_T \cdot \ln\left(\frac{J_2}{J_1}\right) = V_T \cdot \ln\left[\left(\frac{I}{I_S}\right)\cdot\left(\frac{J_2}{J_1}\right)^g\right]$$

Substituting g, this reduces to the previous equation for V_O.

The gain formula of g can be expressed more simply using ΔV as

$$g = \frac{V_{go} + 1.4\cdot V_T - V}{\Delta V}$$

This simpler formula for g is expressed entirely in static circuit voltages.



This simple bandgap circuit is the *Widlar bandgap reference*, after Bob Widlar (pronounced "Wide-ler"), who invented the bandgap concept. R_1 sets current I_1 through Q_1. The current, or current density, of Q_1 must be larger than that of Q_2 to create a positive ΔV across R_2. Assuming $\alpha = 1$ for Q_2, the gain is $V_{R3}/\Delta V = R_3/R_2$. The application of the equation for V_O involves V_{R3} and V_{BE3}:

$$V_O = V_{BE3} + V_{R3} = V_T \cdot \ln \frac{I_{C3}}{I_{S3}} + \left(\frac{R_3}{R_2}\right) \cdot V_T \cdot \ln \frac{J_1}{J_2}$$

This must equal the previous expression for V_O above. The junction currents are kept constant by bootstrapping their sources from the stable output, supplied by I_0. Q_3 also shunt regulates the output. In this analysis, α error has been ignored, and biasing constrains $V_{BE3} > V_{BE1}$ to keep from saturating Q_2. Consequently, $I_2 < I_1 < I_3$ for equal areas. The topology places a limit on how large ΔV can be made.

Example: Widlar Bandgap Reference

Based on the topology of the Widler circuit shown above and the previous example, $I_{E3} = 1$ mA and $V_{BE3} = 0.715$ V. The output voltage is $V_O = 1.242$ V. Then

$$V_{R3} = V_O - V_{BE3} = 0.527 \text{ V}$$

We must set $I_1 < I_{E3}$ to reverse-bias the *b-c* junction of Q_2 and for $\Delta V = V_{R2} > 0$, $I_2 < I_1$. For maximum I_1/I_2, let $V_{CB2} = 0$ V, the same as V_{CB1}. At low currents, the drop across r_c' is negligible, and saturation of Q_2 is avoided. Then

$$V_1 = V_{BE3} \Rightarrow I_1 = I_{E3}$$

By choosing I_1, we choose ΔV as

$$\Delta V = V_T \cdot \ln \frac{I_{C3}}{I_2}$$

Let I_2 = 100 μA. Then

$$\Delta V = V_T \cdot \ln \frac{I_1}{I_2} = 59.6 \text{ mV}$$

and

$$g = \frac{V_O - V}{\Delta V} = 8.85$$

Furthermore,

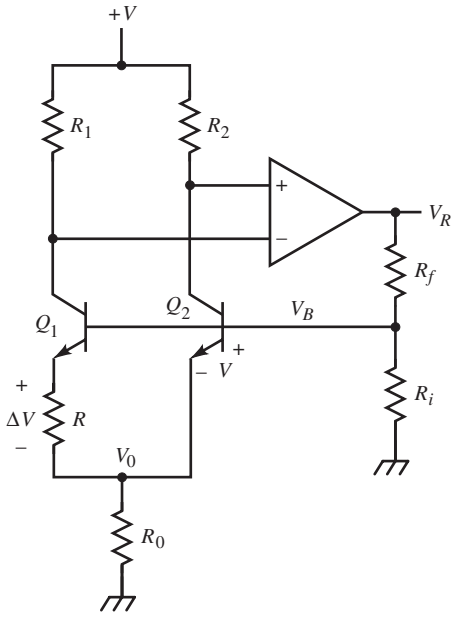
$$R_3 = \frac{V_O - V}{I_2} = \frac{0.527 \text{ V}}{100 \text{ μA}} = 5.27 \text{ kΩ}$$

$$R_2 = \frac{R_3}{g} = 595 \text{ Ω}$$

Then

$$R_1 = \frac{V_O - V_1}{I_1} = \frac{0.572 \text{ V}}{1 \text{ mA}} = 527 \text{ Ω}$$

Total current is 2.1 mA, considerably less than the zero-TC current of typical Zeners. Much lower currents are also feasible.



The Widlar reference has the disadvantage of a fixed output voltage that is not optimal for many applications. A bandgap reference with arbitrary output would be better.

The differential bandgap circuit shown above, invented by Paul Brokaw at Analog Devices, Inc., has output V_R that is set by R_f and R_i. Q_2 has a higher current density than Q_1, making ΔV the difference of the V_{BE} voltages;

$$\Delta V = I_1 \cdot R = V_{BE2} - V_{BE1} = V - V_T \cdot \ln \frac{I_1}{I_S} = V_T \cdot \ln \frac{J_2}{J_1}$$

The op-amp inputs are kept at the same voltage by feedback so that

$$I_1 \cdot R_1 = I_2 \cdot R_2$$

If J_2/J_1 and I_1 are chosen, then ΔV is determined by R. By choosing the *b-e* junction areas, we also determine I_2. Thus, R_1 and R_2 are determined. Then

$$V_0 = (I_1 + I_2) \cdot R_0$$

is set by R_0. The base voltage is then

$$V_B = V + V_0 = V_T \cdot \ln\left(\frac{I_2}{I_S}\right) + \left(\frac{R_0}{R}\right) \cdot \left(1 + \frac{I_2}{I_1}\right) \cdot V_T \cdot \ln\left(\frac{J_2}{J_1}\right)$$

Finally, the output is

$$V_R = V_B \cdot \left(1 + \frac{R_f}{R_i}\right)$$

For zero TC(V_R), we must have TC(V_B) = 0. For this circuit, V_B has the form of V_O, the basic bandgap-reference equation. V_B is V_O, V_0 corresponds to $g \cdot \Delta V$, and

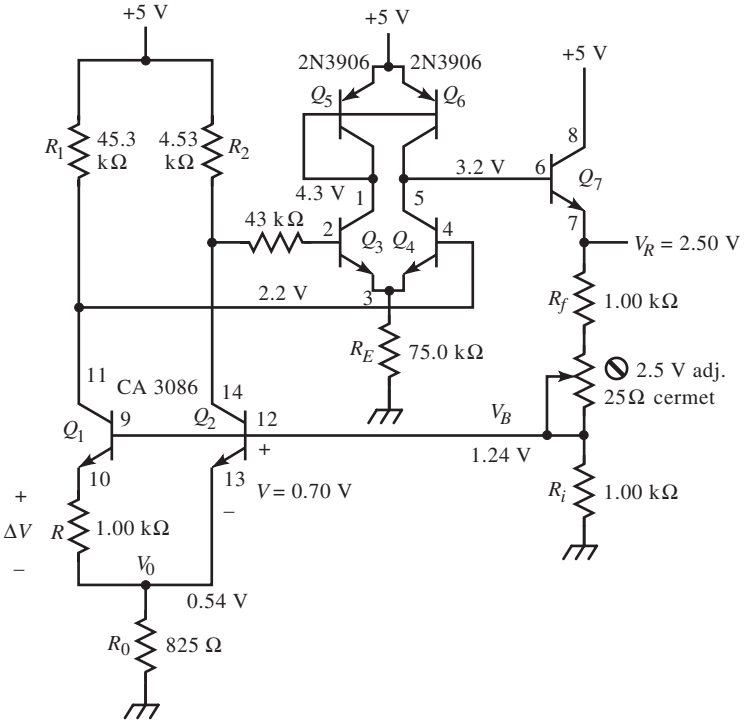
$$g = \frac{V_0}{\Delta V} = \left(\frac{R_0}{R}\right) \cdot \left(1 + \frac{I_2}{I_1}\right)$$

For TC(V_B) = 0, g must satisfy the g formula, or

$$V_0 = V_B - V = g \cdot \Delta V$$

The TC(V_0) > 0 and is linear with absolute temperature.

Example: Differential Bandgap Reference



The circuit sketched above is based on a five-transistor CA3086 array of matched transistors with equal areas. The goal is to design a 2.50 V reference. For CA3086 BJTs,

$$I_S = 10^{-15} \text{ A}$$

For this design, let

$$\frac{J_2}{J_1} = 10, \quad I_1 = 60\ \mu\text{A}, \quad I_2 = 600\ \mu\text{A}$$

Then

$$R = \frac{\Delta V}{I_1} = \frac{V_T \cdot \ln(10)}{60\mu\text{A}} = \frac{59.56\ \text{mV}}{60\mu\text{A}} = 992\ \Omega \Rightarrow 1.00\ \text{k}\Omega \pm 1\%$$

Now proceed to find R_0 by first calculating V:

$$V = V_{BE2} = V_T \cdot \ln\left(\frac{600\ \mu\text{A}}{10^{-15}\ \text{A}}\right) = 0.702\ \text{V}$$

Then, from Ohm's law and the expression for V_0,

$$V_0 = R_0 \cdot (660\ \mu\text{A}) = 1.242\ \text{V} - 0.702\ \text{V} = 0.541\ \text{V}$$

Solve for R_0:

$$R_0 = 819\ \Omega \Rightarrow 825\ \Omega \pm 1\%$$

Now calculate R_1 and R_2 from the *IR* equation. This is a ratio requiring another constraint to determine actual values. Note that V_{CB} of Q_1 and Q_2 are in series with that of Q_3 and Q_4. V_{C3} is a junction drop down from the supply of 5 V, or about 4.3 V. But V_{C4} is less; it is a junction drop up from the output of 2.5 V or 3.3 V. And $V_B = 1.242$ V. Split the voltage difference between the series *b-c* junctions so that

$$V_{B3} = V_{B4} = \frac{3.3\ \text{V} - 1.24\ \text{V}}{2} + 1.24\ \text{V} = 2.3\ \text{V}$$

Then

$$R_1 = \frac{5\ \text{V} - 2.3\ \text{V}}{60\ \mu\text{A}} = 45\ \text{k}\Omega \Rightarrow 45.3\ \text{k}\Omega \pm 1\%$$

and

$$R_2 = \frac{R_1}{10} = 4.5\ \text{k}\Omega \Rightarrow 4.53\ \text{k}\Omega \pm 1\%$$

To compensate the diff-amp for bias current, set the source resistances equal. This requires a base resistor for Q_3 of

$$R_{B3} = 45.3\ \text{k}\Omega - 4.53\ \text{k}\Omega = 40.8\ \text{k}\Omega \Rightarrow 43\ \text{k}\Omega$$

The diff-amp emitter bias current is set by R_E. If we choose it to be 20 μA, then base current for $\beta \cong 100$ is about 100 nA, a small fraction of 60 μA. The emitter voltage is a junction drop down from V_{B3}, or about 1.5 V. Then

$$R_E = \frac{1.5\ \text{V}}{20\ \mu\text{A}} = 75\ \text{k}\Omega$$

Finally, the feedback divider is

$$\frac{R_f}{R_i} + 1 = \frac{2.500\ \text{V}}{1.242\ \text{V}} = 2.013$$

Choose R_i = 1.00 kΩ, 1%, a convenient value. Then R_f = 1.013 kΩ. To allow adjustment of the output to correct for parts tolerances, place a trim-pot (screwdriver-adjusted potentiometer) in series with

$$R_f = 1.00\ \text{k}\Omega \pm 1\%$$

having twice the remaining resistance, to center the pot, or

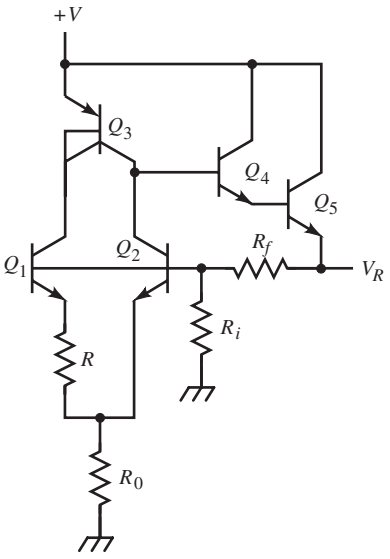
$$R_{adj} = 2 \cdot (1.013\ \text{k}\Omega - 1.00\ \text{k}\Omega) = 25.8\ \Omega \Rightarrow 25\ \Omega$$

A cermet pot has a low TC, required for the application, but such a low value may not be available in cermet, and R_f must be reduced to make the trim-pot larger.

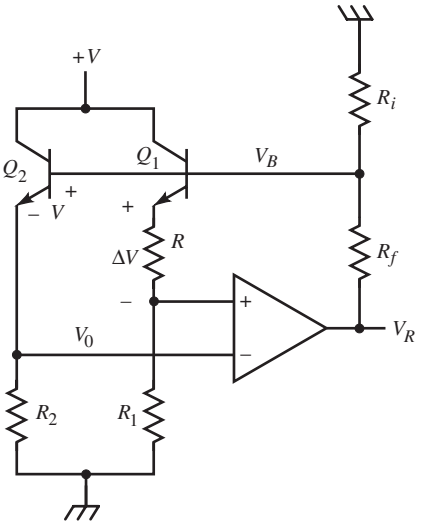
All parts values have now been determined, and the circuit can be "prototyped" to verify performance. A prototype was built with the following deviations:

$$R_1 = 49.9\ \text{k}\Omega \pm 1\%, \quad R_2 = 4.99\ \text{k}\Omega \pm 1\%, \quad R_0 = 820\ \Omega \pm 5\%$$

The supply measured 5.01 V, V_0 was 0.540 V, and V_B was 1.242 V. These measurements were taken on a warm spring evening in a building without air conditioning; the temperature was approximately 300 K. The CA3086 was then heated to about 50°C above ambient temperature with a soldering iron, and V_B became 1.237 V. The circuit was then cooled with circuit cooler (an aerosol); V_B was then 1.248 V. A rough calculation indicates that TC(V_B) is roughly 100 ppm, about the same TC as the metal-film 1% resistors. In a refined design, this discrete implementation should have all 1% metal-film resistors (no 5% composition resistors). Better yet, it should be an integrated circuit. But for the 30 minutes it took to build, it demonstrated the validity of the derived design equations.



In integrated form, a simple differential bandgap reference can have the topology shown above, in which an emitter-area ratio other than one is used, where $A_1 > A_2$. Q_3 is a current mirror, and Q_4 and Q_5 form a Darlington buffer to the output.



This third bandgap-reference topology uses an op-amp with inputs from the emitter (instead of collector) circuit of the bandgap cell, Q_1 and Q_2. The analysis is similar to the previous one. Equations for ΔV and V_R are valid. So are the *IR* equation and V_O, where V_0 equals $I_1R_1 = I_2R_2$. The equation for V_B is slightly modified:

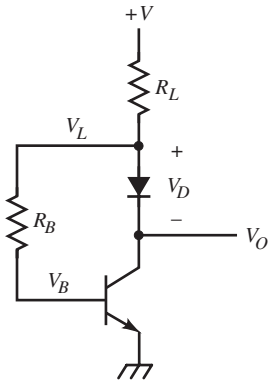
$$V_B = V + V_0 = V_T \cdot \ln\left(\frac{I_2}{I_S}\right) + I_1 \cdot R_1 = V_T \cdot \ln\left(\frac{I_2}{I_S}\right) + \left(\frac{R_1}{R}\right) \cdot V_T \cdot \ln\left(\frac{J_2}{J_1}\right)$$

By comparison,

$$g = \frac{V_0}{\Delta V} = \frac{R_1}{R}$$

and the previous expression for V_O also applies.

Besides these three popular bandgap circuits, various other topologies have been used in commercial ICs.



In some designs, a very simple voltage reference is needed that does not require a low TC. A shunt-feedback voltage source is shown above. We want V_O to be insensitive to the temperature and supply voltage V for good *power-supply rejection* (PSR).

This circuit has no closed-form static solution but can be designed without iteration, given V, V_O, and I_S. Assume that the diode and BJT are matched. The supply current is I_E; choose I_E. Then

$$V_B = V_T \cdot \ln \frac{\alpha \cdot I_E}{I_S}$$

By Kirchhoff's voltage law (KVL),

$$V_O = V - I_E \cdot R_L - V_D = V - I_E \cdot R_L - V_T \cdot \ln \frac{\alpha \cdot I_E}{I_S}$$

where V_D equals V_B. Then

$$R_L = \frac{V - (V_O + V_B)}{I_E}$$

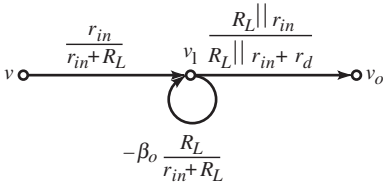
KVL is again applied to the base circuit:

$$\frac{V - I_E \cdot R_L - V_B}{R_B} = \frac{I_E}{\beta + 1}$$

Solving for R_B yields

$$R_B = (\beta + 1) \cdot \left[\left(\frac{V - V_B}{I_E} \right) - R_L \right]$$

The PSR is expressed in small-signal quantities as v_o/v.



The flow graph, shown above, reduces to

$$\frac{v_o}{v} = \frac{r_{in}}{r_{in} + (\beta + 1) \cdot R_L} \cdot \frac{R_L \| r_{in}}{R_L \| r_{in} + r_d} \cong \frac{r_{in}}{r_{in} + (\beta + 1) \cdot R_L}, \quad r_d \cong 0$$

$$\cong \frac{V_O}{V - V_B}, \quad r_e \cong 0$$

where

$$r_{in} = R_B + (\beta + 1) \cdot r_e$$

PSR is often expressed as the *PSR ratio* (PSRR):

$$\text{PSRR} \equiv 20 \cdot \log \frac{v}{v_o}$$

The TC of the shunt-feedback reference largely depends on TC(β). The output voltage is

$$V_O = V_B - V_D + I_B \cdot R_B = (V_B - V_D) + \frac{I_E}{\beta + 1} \cdot R_B$$

Because the junctions nearly cancel, the first term is negligible. In the second term, $I_B(T)$ varies with $\beta(T)$ at about 1%/°C. Less base current is required to sustain V_L with increasing β. Thus, V_O has a negative TC. For small changes in V_O and constant V, v_o varies inversely with β.

Feedback reduces the dynamic output resistance to

$$\begin{aligned} r_{out} &\cong \frac{r_{in} \| R_L}{1 + \beta \cdot R_L/(r_{in} + R_L)} = r_{in} \cdot \frac{R_L}{r_{in} + (\beta + 1) R_L} = \frac{r_{in}}{\beta + 1} \cdot \left[1 - \frac{v_o}{v} \right] \\ &\cong \frac{R_B}{\beta + 1} \cdot \left[\frac{V - V_L}{V - V_B} \right], \quad r_e \cong 0 \end{aligned}$$

Example: Shunt-Feedback Voltage Reference

The shunt-feedback reference circuit has the following design parameters:

$$I_S = 10^{-15} \text{ A}, \ \beta = 99, \quad \text{matched junctions}, \quad V = 5 \text{ V}, \ V_O = 2.5 \text{ V}$$

Let I_E = 1 mA. This value is chosen so that any load current is negligible in comparison. Applying V_B, R_L, and R_B yields

$$V_B = V_T \cdot \ln \frac{1 \text{ mA}}{10^{-15} \text{ A}} = 0.715 \text{ V}$$

$$R_L = \frac{5 \text{ V} - (2.5 \text{ V} + 0.715 \text{ V})}{1 \text{ mA}} = 1.79 \text{ k}\Omega \Rightarrow 1.8 \text{ k}\Omega$$

$$R_B = (100) \cdot \left[\frac{5 \text{ V} - 0.715 \text{ V}}{1 \text{ mA}} - 1.79 \text{ k}\Omega \right] = 250 \text{ k}\Omega \Rightarrow 240 \text{ k}\Omega$$

The PSR is calculated as follows:

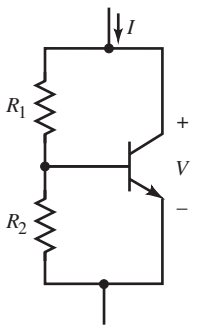
$$r_{in} = 240\ \text{k}\Omega + (100)\cdot(26\ \Omega) = 243\ \text{k}\Omega$$

$$\frac{v_o}{v} = \frac{243\ \text{k}\Omega}{243\ \text{k}\Omega + (100)\cdot(1.8\ \text{k}\Omega)} = 0.57$$

This does not amount to much power-supply rejection. As a PSRR, it is only 4.8 dB. The dynamic output resistance is reduced to

$$r_{out} = (243\ \text{k}\Omega)\cdot\left(4.26\times10^{-3}\right) = 1.03\ \text{k}\Omega$$

In this example, the shunt-feedback voltage reference r_{out} has only about a × 2 advantage over a resistive divider because R_L is not very large relative to r_{in}.



A simple voltage source that is easily floated is the V_{BE} *multiplier*, shown above. When driven by a current source, it behaves as a shunt-feedback amplifier with voltage,

$$V = R_1 \cdot I_B + V_{BE} \cdot \left(1 + \frac{R_1}{R_2}\right)$$

$$= \frac{I \cdot R_1}{\beta + 1} + V_{BE} \cdot \left(1 + \frac{R_1}{R_2}\right), R_1 + R_2 \to \infty$$

$$= V_{BE} \cdot \left(1 + \frac{R_1}{R_2}\right), \beta \to \infty$$

where I is the total current. Its main advantage over a current-driven resistor is its dynamic resistance,

$$r_{out} = (R_1 + R_2 \| r_\pi) \| r_m \cdot \left(\frac{R_1 + R_2 \| r_\pi}{R_2 \| r_\pi} \right)$$

where r_m is the BJT transresistance of r_e/α and $r_\pi = (\beta + 1)\, r_e$. The first shunt resistance is the divider resistance, and the second is the equivalent BJT resistance. If the resistive-divider loading is negligible, then

$$r_{out} \cong r_m \left(\frac{V}{V_{BE}} \right) = \frac{1}{\alpha} \cdot \frac{V_T}{I_E} \cdot \frac{V}{V_T \cdot \ln(I/I_S)} = \frac{V/I}{\alpha \cdot \ln(I/I_S)}$$

The numerator is the value of a current-driven resistor; the denominator is the improvement factor due to the BJT.

The voltage source driving this circuit is V_{BE}, which drifts with temperature. From V,

$$\frac{dV}{dT} = \frac{-I \cdot R_1}{\beta + 1} \cdot \alpha \cdot TC\%(\beta) + \left(1 + \frac{R_1}{R_2} \right) \cdot V_{BE} \cdot TC\%(V_{BE})$$

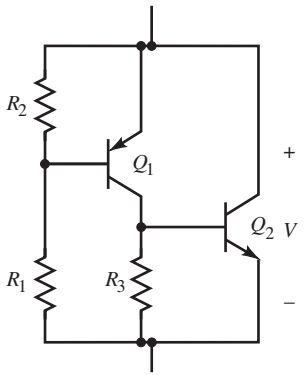
assuming TC(α) $\cong$ 0. The fractional TC of junction voltage with constant I, which is found for $V_{BE} = V$ on p. 6 from $dV/dT|_I$, is

$$TC\%(V_{BE})|_I = \frac{1}{V_{BE}} \cdot \left. \frac{dV_{BE}}{dT} \right|_I = \frac{1}{V_{BE}} \cdot \frac{V_{BE} - (V_{go} + 1.4 \cdot V_T)}{T} = \frac{1}{T} \cdot \left(1 - \frac{V_{go} + 1.4 \cdot V_T}{V_{BE}} \right)$$

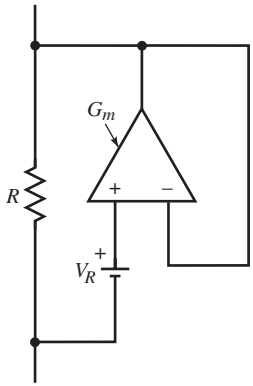
Then dV/dT, with a typical $V_{BE} = 0.7$ V, at 27 °C, becomes

$$\text{typical} \frac{dV}{dT} = -\frac{I \cdot R_1}{\beta + 1} \cdot (1\%/°\text{C}) - \left(1 + \frac{R_1}{R_2} \right) \cdot (0.7\,\text{V}) \cdot (0.26\%/°\text{C})$$

and the TC(V) < 0. This circuit is commonly used in the base circuit of complementary common-collector (CC) buffers, as an alternative to a voltage drop across a resistor driven by a current source. It can be designed so that its TC tracks the CC output BJTs in "Complementary Emitter-Follower Output Amplifier" in *Designing High-Performance Amplifiers.*



A V_{BE} multiplier with an additional gain stage is shown above. The circuit below is the general form of the V_{BE} multiplier.



The voltage source V_R drives the shunt transconductance amplifier across R. With a current of I, the voltage across R is

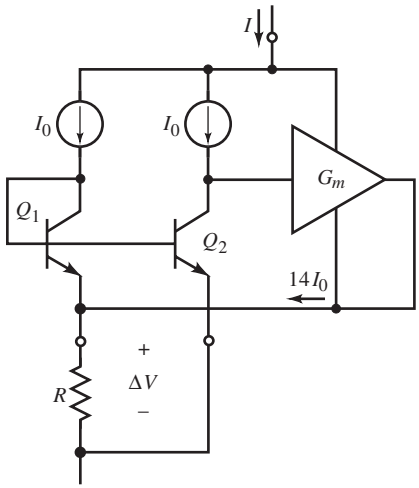
$$V = (I + V_R \cdot G_m) \cdot \left[R \left\| \frac{1}{G_m} \right. \right]$$

The incremental dynamic resistance is

$$r_{out} = R \left\| \frac{1}{G_m} \right.$$

Without the amplifier, it is R. For large G_m, it approaches zero.

CURRENT SOURCES



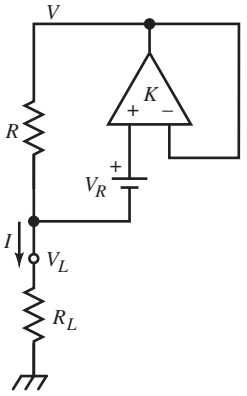
The three-terminal current-source IC, based on the National Semiconductor LM334, has a bandgap cell Q_1, Q_2, and a transconductance amplifier with an output of $14I_0$. Each BJT conducts I_0, and the area of Q_1 is 14 times that of Q_2, or $A_1 = 14A_2$. With the same current,

$$\frac{J_2}{J_1} = 14$$

The terminal current at 25°C is

$$I = 16 \cdot I_0 = 16 \cdot \left(\frac{\Delta V}{15R} \right) = \left(\frac{16}{15} \right) \cdot \frac{V_T \cdot \ln(14)}{R} = \frac{(1.067) \cdot (67.77\ \text{mV})}{R} = \frac{72.3\ \text{mV}}{R}$$

The data book specifies that the voltage across R is 67.7 mV. I varies with V_T and has a TC of 0.336%/°C at 25°C. A shunt RD combination in series with this part adds a negative TC. If the shunt R is chosen properly, the TC can be set to zero.



A current source based on the V_{BE}-multiplier concept is shown above. The amplifier has voltage gain K. Otherwise, the topology is the same as the V_{BE}-multiplier voltage source. The circuit equations are

$$V = K \cdot [(V_R + V_L) - V]$$

or

$$V = \left(\frac{K}{K+1} \right) \cdot (V_R + V_L)$$

Substituting for V,

$$V = I \cdot (R + R_L)$$

Also,

$$V_L = I \cdot R_L$$

Then solving for I,

$$I = \frac{\left(\frac{K}{K+1}\right) \cdot V_R}{R + R_L/(K+1)}$$

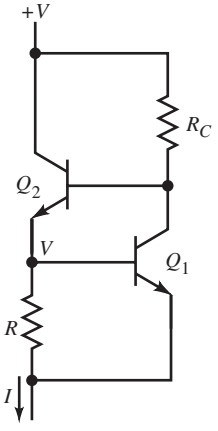
For an op-amp, $K \to \infty$, and

$$I|_{K\to\infty} = \frac{V_R}{R}$$

I is independent of the load, as desired. The dynamic output resistance, which ideally is infinite, is

$$r_{out} = -\frac{dV_L}{dI} = -\frac{dV_L}{dV} \cdot \frac{dV}{dI} = -\left(\frac{R_L}{R_L + R}\right) \cdot (-KR) = (K) \cdot (R \| R_L)$$

As $K \to \infty$, $r_{out} \to \infty$, as desired. R should be made as small as feasible to maintain high r_{out} for small R_L. R_L begins to affect I significantly as it approaches the value of R.



A BJT realization of the V_{BE}-multiplier current source is shown above. V_R is V_{BE} of Q_1, and K is the loop gain with V_{BE1} as input:

$$K = \frac{v}{v_{be1}} = \alpha \cdot \frac{R_C \| [(\beta+1)\cdot(r_{e2} + R \| r_{\pi 1} + R_L)]}{r_{e1}}$$

R should be made small for load insensitivity and R_C large for high K. $I \propto V_R = V_{BE1}$ and $\mathrm{TC}(V_{BE}) \cong -2$ mV/°C,

$$\mathrm{TC}\%(I) = \mathrm{TC}\%(V_{BE1})$$

In the BJT current source, the diff-amp input is the *b-e* junction of Q_1, and I_{E1} also contributes to I, or

$$I \cong I_{E1} + I_{C2}$$

If I_{C1} is chosen, then V_{BE1} is determined and R is calculated from

$$R = \frac{V_T \cdot \ln(I_{C1}/I_S)}{I - I_{E1}}$$

Next, R_C must be chosen to satisfy static constraints. Given R_L and V, and with BJT parameter I_S, then

$$I_{E2} = I - I_{E1} + \frac{I_{E1}}{\beta+1} = I - \alpha \cdot I_{E1}$$

and

$$V_{BE2} = V_T \cdot \ln \frac{I_{C2}}{I_S}$$

Then the current through R_C, corrected for I_{B2}, is

$$I_{RC} = \alpha \cdot I_{E1} + \frac{I_{E2}}{\beta+1} = \frac{[\beta^2/(\beta+1)] \cdot I_{E1} + I}{\beta+1}$$

With these calculated values,

$$R_C = \frac{V - (V_{BE2} + V_{BE1} + I \cdot R_L)}{I_{RC}}$$

Example: V_{BE}-Multiplier Current Source

The current source shown above has a 1 kΩ nominal load to ground and $V = 5$ V. Choose I to be 1 mA. The BJTs have $\beta = 99$ and $I_S = 10^{-15}$ A. Then, if we let

$$I_{C1} = 100\ \mu\text{A}$$

$$V_{BE1} = V_T \cdot \ln \frac{100\ \mu\text{A}}{10^{-15}\ \text{A}} = 0.655\ \text{V}$$

it follows that

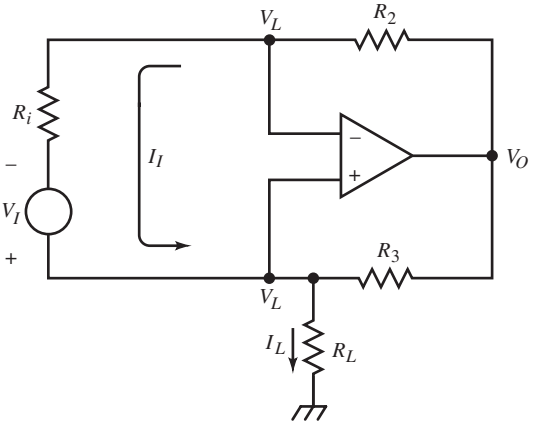
$$R = \frac{0.655\ \text{V}}{1\ \text{mA} - 100\ \mu\text{A}} = 728\ \Omega \Rightarrow 750\ \Omega$$

From the equation for I_{E2}, $I_{E2} = 0.901$ mA, and $V_{BE2} = 0.712$ V. $IR_L = 1$ V. Next, $I_{RC} = 10.80$ μA, and finally, from R_C,

$$R_C = \frac{5\ \text{V} - 2.367\ \text{V}}{10.80\ \mu\text{A}} = 24.4\ \text{k}\Omega \Rightarrow 24\ \text{k}\Omega$$

K is calculated to be 80. This makes $r_{out} = 342$ kΩ. The dynamic output resistance is positive, though the static resistance is negative because the output terminal is that of a source. Because $R \cong R_L$, this design could be improved by a choice of smaller I_{E1}, causing R to be smaller. Because I_{E1} is already a tenth I_{E2}, we are at a point of diminishing improvement. An op-amp realization would get around the lower limits on R.

This circuit is quite sensitive to the value of R. It was built using a 750 Ω, 5% resistor; the resulting I was about 7% low. With a trim-pot adjusted to 728 Ω, the error was about 0.2%. Therefore, a 1% value of 732 Ω would be a better choice for R.



An op-amp-based current source was invented in 1963 by Brad Howland at the Massachusetts Institute of Technology (MIT). It is the *Howland current source*, shown with a floating voltage source input. This circuit has positive feedback to the noninverting input. With a sufficiently large load resistance, the circuit becomes unstable. The positive feedback provides a bootstrap effect that keeps the load current I_L constant.

The op-amp keeps its inputs at the same voltage; both are at the load voltage V_L. The same voltage appears at both ends of the input branch through which flows the input current,

$$I_I = \frac{V_I}{R_i}$$

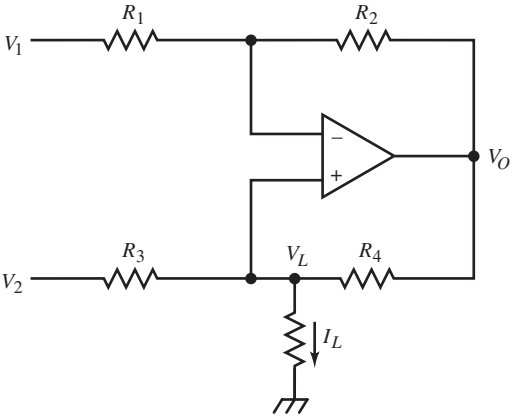
This current flows through R_2, causing

$$V_O = I_I \cdot R_2 + V_L = \frac{R_2}{R_i} \cdot V_I + V_L$$

V_O is thus established. It causes a current through R_3; applying Kirchhoff's current law (KCL) at the load node and substituting V_O,

$$I_L = I_I + \frac{V_O - V_L}{R_3} = \frac{V_I}{R_i} + \left(\frac{R_2}{R_i}\right) \cdot \left(\frac{1}{R_3}\right) \cdot V_I = \left(\frac{R_2 + R_3}{R_i R_3}\right) \cdot V_I$$

The cancellation of V_L in the numerator manifests bootstrapping. V_O tracks V_L, keeping I_L independent of V_L and hence the circuit behaves as a current source.



Floating voltage sources are usually inconvenient. A more general Howland circuit, shown above, has two voltage inputs, V_1 and V_2, with differential input

$$V_I = V_2 - V_1$$

What is different is that the currents in R_1 and R_3 can be different. The circuit is solved similar to the previous one. V_O from the inverting side is

$$V_O = V_L + I_2 \cdot R_2 = V_L + \frac{V_L - V_1}{R_1} \cdot R_2 = \left(-\frac{R_2}{R_1}\right) \cdot V_1 + \left(\frac{R_2}{R_1} + 1\right) \cdot V_L$$

On the noninverting side, applying KCL and substituting for V_O,

$$I_L = \frac{V_2 - V_L}{R_3} + \frac{V_O - V_L}{R_4} = \left(-\frac{R_2}{R_1 \cdot R_4}\right) \cdot V_1 + \frac{V_2}{R_3} + \left(\frac{R_2}{R_1 \cdot R_4} - \frac{1}{R_3}\right) \cdot V_L$$

This general expression for I_L is not independent of V_L, as required of a current source. The coefficient of V_L is set to zero under the condition

$$\frac{R_2}{R_1 \cdot R_4} = \frac{1}{R_3} \Rightarrow \frac{R_2}{R_4} = \frac{R_1}{R_3} \quad \text{or} \quad \frac{R_2}{R_1} = \frac{R_4}{R_3}$$

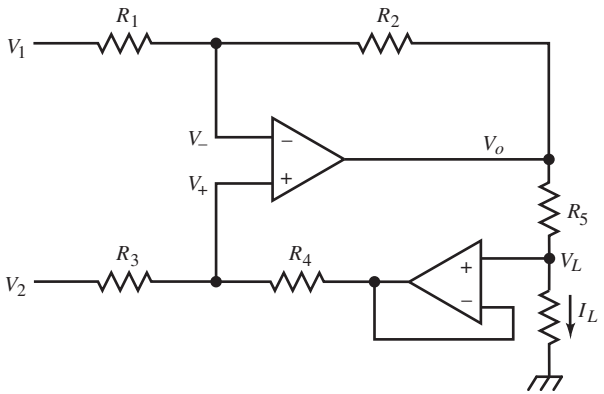
Under this condition, I_L reduces to

$$\text{current source } I_L = \frac{V_2 - V_1}{R_3} = \frac{V_I}{R_3}$$

The output resistance is found by regarding the static quantities of I_L as variables, and then differentiating and inverting

$$r_{out} = \frac{\partial V_L}{\partial I_L} = 1 \Big/ \left(\frac{R_2}{R_1 \cdot R_4} - \frac{1}{R_3}\right) = \frac{R_4}{R_2/R_1 - R_4/R_3}$$

Under the current-source conditions, r_{out} is infinite.



The modified Howland source above has an additional resistor R_5 and a buffer between the load and noninverting input. This increases compliance (load-voltage range) and load-current range because R_5 can be made small while R_4 satisfies the gain requirement of a current source. If $I_{R4} << I_L$, the buffer can be omitted and R_4 connected to R_5. For this circuit, assume finite op-amp gain K and apply superposition to the op-amp inputs:

$$V_- = \left(\frac{R_1}{R_1 + R_2}\right) \cdot V_O + \left(\frac{R_2}{R_1 + R_2}\right) \cdot V_1$$

$$V_+ = \left(\frac{R_3}{R_3 + R_4}\right) \cdot V_L + \left(\frac{R_4}{R_3 + R_4}\right) \cdot V_2$$

The op-amp output voltage is

$$V_O = K \cdot (V_+ - V_-) = \frac{K}{\left[(R_1 + R_2)/R_1\right] + K} \cdot \left(V_O|_{K \to \infty}\right)$$

when $K \to \infty$, V_O is

$$V_O|_{K \to \infty} = \left(\frac{R_1 + R_2}{R_1}\right) \cdot \left[\left(\frac{R_4}{R_3 + R_4}\right) \cdot V_2 - \left(\frac{R_2}{R_1 + R_2}\right) \cdot V_1 + \left(\frac{R_3}{R_3 + R_4}\right) \cdot V_L\right]$$

The load current then is

$$I_L = \frac{V_O - V_L}{R_5}$$

Substituting V_O for infinite K yields an expression in V_1, V_2, and V_L. When the coefficient of V_L is set to zero, the current-source condition results. Not surprisingly, it is the same as the previous conditions because the feedback topology is the same as the previous circuit. Then

$$I_L = \left(\frac{R_2}{R_1}\right) \cdot \frac{V_I}{R_5} = \left(\frac{R_4}{R_3}\right) \cdot \frac{V_I}{R_5}$$

where $V_I = V_2 - V_1$.

When the buffer is omitted (shorted), I_L is reduced by I_{R4}. The incremental gain of the noninverting loop is calculated as follows. For $\Delta I_L = 0$, the change in current through R_5 must equal that through R_3 and R_4, or

$$\frac{v_o - v_L}{R_5} = \frac{v_L}{R_3 + R_4}$$

or

$$\frac{v_L}{v_o} = \frac{R_3 + R_4}{R_3 + R_4 + R_5}$$

The noninverting loop gain to v_L must be that of a current source and is

$$\frac{v_+}{v_L} \cdot \frac{v_o}{v_+} \cdot \frac{v_L}{v_o} = \left(\frac{R_3}{R_3 + R_4}\right) \cdot \left(\frac{R_1 + R_2}{R_1}\right) \cdot \left(\frac{R_3 + R_4}{R_3 + R_4 + R_5}\right)$$

For the noninverting loop gain to be 1,

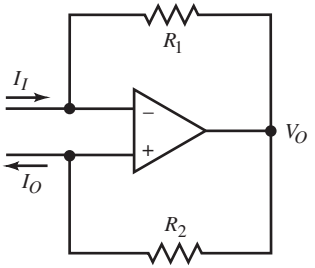
$$\frac{R_2}{R_1} = \frac{R_4 + R_5}{R_3}$$

Then the differential of the load current is

$$dI_L = \left(\frac{\partial I_L}{\partial V_L}\right) \cdot dV_L + \left(\frac{\partial I_L}{\partial V_o}\right) \cdot \left(\frac{\partial V_o}{\partial V_i}\right) \cdot dV_i$$

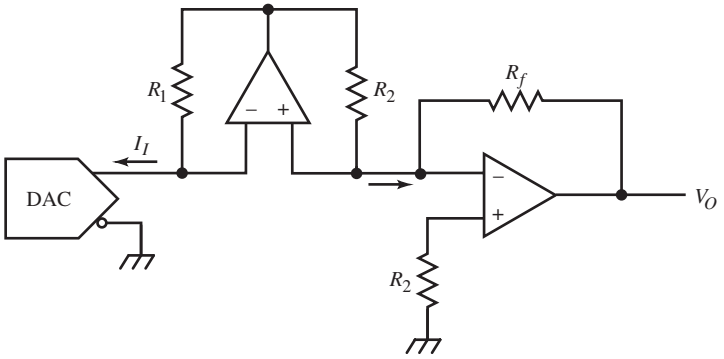
$$\begin{array}{ccc} \uparrow & \uparrow & \uparrow \\ 0 & 1/R_5 & A_v \end{array}$$

where $A_v \cdot dV_i = dV_o$. For a current source, only dV_i (not dV_L) can change I_L.

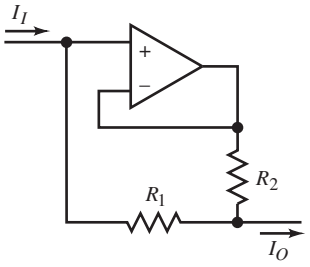


The figure above shows an inverting current-gain amplifier that uses positive feedback, a variation on the Howland topology. With the voltages at the op-amp inputs kept equal, R_1 and R_2 drop the same voltage. It then follows that

$$\frac{I_O}{I_I} = \frac{R_1}{R_2}$$

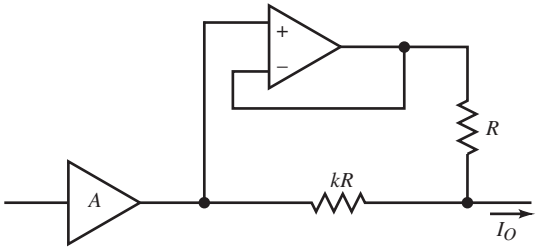


This application of the inverting op-amp current amplifier reverses the DAC output-current polarity and scales it for input to the inverting op-amp. The voltage output is negative. Of significance is the DAC output node, which is kept at the same voltage as the virtual ground (inverting input) of the op-amp, meeting the constraint of a limited-compliance DAC.

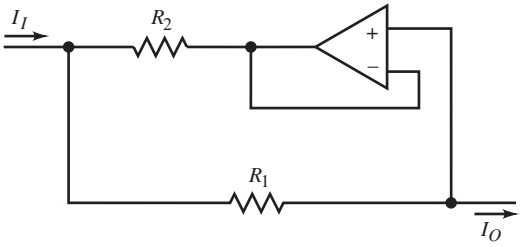


The noninverting current amplifier above applies to R_2, through the ×1 buffer, the same voltage that is across R_1. The current gain is

$$\frac{I_O}{I_I} = 1 + \frac{R_1}{R_2}$$



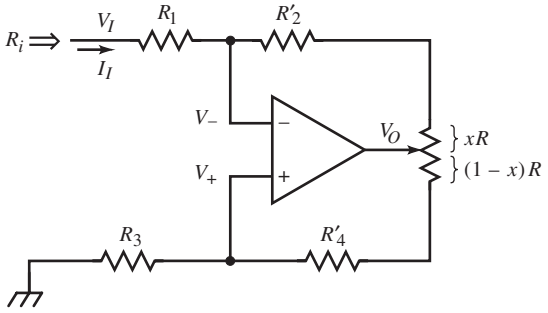
Shown above, this current amplifier is used to boost the output current I_O of an amplifier by k times.



By reversing the buffer, as shown above, current is attenuated instead, in this precision floating current shunt. This circuit is similar to a previous current source; V_R is removed, and R is driven by I_I instead. The current gain is the current divider formula,

$$\frac{I_O}{I_I} = \frac{R_2}{R_1 + R_2}$$

Example: Bipolar Simulated Resistance



A circuit with similar topology to the Howland current source provides a precision, adjustable, bipolar input resistance and can be used in the one-op-amp diff-amp in place of the grounded resistor. This is sometimes necessary due to

unavoidable parasitic resistance in the ground return path. Applying KCL twice gives

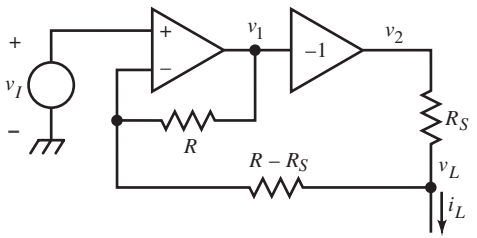
$$R_I = \frac{V_I}{I_I} = R_1 \cdot \left(1 + \frac{R_4}{R_3}\right) - R_2$$

where R_2 and R_4 include the trim-pot resistances. When $R_1 = R_3 = R$, then

$$R_i = R + (R_4 - R_2)$$

When the trim-pot is centered, $R_2 = R_4$, and $R_i = R$. If we set $R_2' = R_4'$, the trim-pot allows adjustment of R_i around R as center value.

Example: Inverting Howland Current Source



The goal in this example is to design an op-amp current source based on the bootstrapping behavior of the Howland source but with an inverted output. The ×(−1) amplifier can be an inverting op-amp. Then

$$v_2 = -v_1 = -\left[v_I - \left(\frac{R}{R - R_S}\right) \cdot (v_L - v_I)\right] = \left(\frac{R}{R - R_S}\right) \cdot v_L - \left(\frac{2 \cdot R - R_S}{R - R_S}\right) \cdot v_I$$

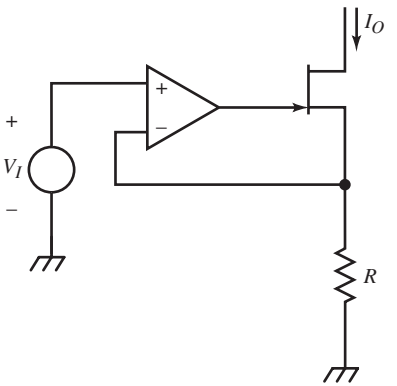
The load current is, by KCL,

$$i_L = \frac{v_2 - v_L}{R_S} - \frac{v_L - v_I}{R - R_S}$$

Substituting for v_2 and reducing, we find that the coefficient of v_L is zero, as required for a current source. Then i_L depends only on v_I:

$$i_L = -\frac{2v_I}{R_S}$$

This circuit therefore functions as a current source.

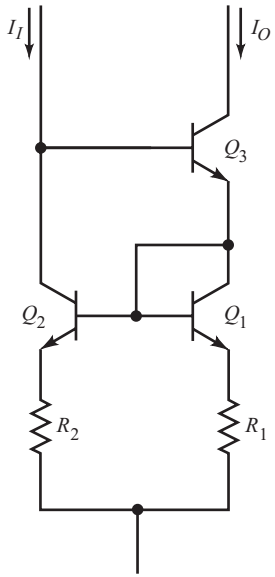


A common way to generate a current I_O from a given voltage V_I is to use an op-amp voltage-to-current (V/I) converter, shown above. The op-amp keeps V_I across R so that

$$I_O = \frac{V_I}{R}$$

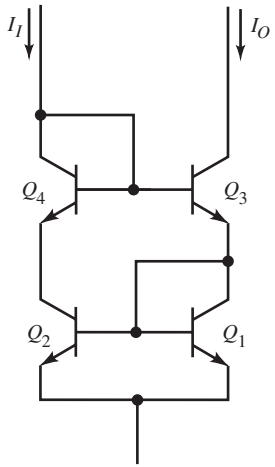
The field-effect transistor (FET) can be replaced by a BJT or Darlington, though it avoids error due to α loss. This circuit need not be grounded. Ground can be replaced by $-V_{EE}$ or, for the complementary V/I converter using the opposite polarity of transistor, by $+V_{CC}$.

The current mirrors in *Designing Amplifier Circuits*, "Current Mirrors", are current-gain amplifiers and can be used as current-driven current sources.



Wilson current mirror

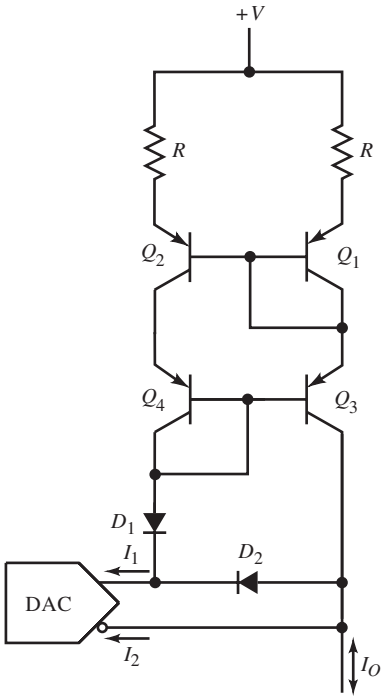
For high precision, the Wilson current mirror shown above should have a junction in the collector of Q_2 for electrical symmetry between Q_1 and Q_2, especially if R_1 and R_2 are not used, as shown below.



An application for the complementary form of this current mirror, shown below, is similar to that of the current-inverter circuit except that the output is a bipolar current. The total DAC output current is

$$I_{fs} = I_1 + I_2$$

D_1 keeps Q_3 out of saturation when D_2 conducts.

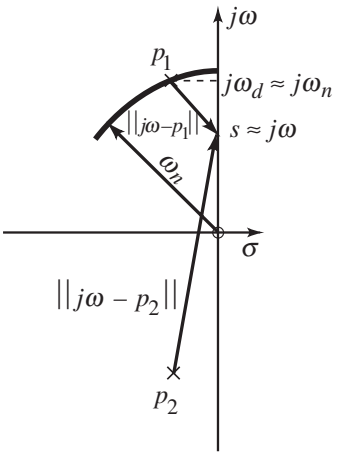


FILTERS

Filters are characterized generally by their transfer functions in the complex-frequency domain. As rational functions of *s*, numerator and denominator can be factored into first- and second-order factors. Higher-order filter polynomials are products of these lower-order factors. Filter responses can be categorized

broadly as low-pass (LP), high-pass (HP), or band-action filters, which are either band-pass (BP) or band-reject (or "notch") filters.

In *Designing Dynamic Circuit Response*, amplifier analysis assumed a low-pass response. In *Designing High-Performance Amplifiers*, some use of high-pass filters in composite amplifiers was demonstrated. In radio communications, highly resonant, or tuned, circuits are used as band-pass filters. These circuits have low ζ or high Q ($Q = 1/2 \cdot \zeta$). Their complex poles are very underdamped and have dominant imaginary components. That is, they lie near the $j\omega$-axis.



The conjugate pole-pair p_1 and p_2 is

$$p_{1,2} = -\alpha \pm j\omega_d = -\zeta \cdot \omega_n \pm j\omega_n \cdot \sqrt{1-\zeta^2} = -\frac{\omega_n}{2 \cdot Q} \pm j\omega_n \cdot \sqrt{1-\frac{1}{(2 \cdot Q)^2}}$$

For $Q >> 1$, the poles have imaginary component $\pm j\omega_d \cong \pm j\omega_n$. The steady-state sinusoidal (or $j\omega$-axis) response is found (as in "*s*-Plane Frequency Response" in *Dynamic Circuit Response*) from the zero-vector lengths divided by the pole-vector lengths. Note that $j\omega - p_1$ varies significantly in both magnitude and angle around $j\omega_n$, where peaking of the magnitude response occurs. At $j\omega_d \cong j\omega_n$, $\|j\omega - p_1\|$ is minimum and the band-pass transfer function is maximum. There, $\angle(j\omega - p_1)$ passes through 0°, an indication of resonance. From the geometry

of the above figure, variations in $j\omega$ around $j\omega_d$ have little effect on the length or orientation of the conjugate pole vector,

$$j\omega - p_2 \cong j\omega_n - (-j\omega_n) = 2j\omega_n, \quad j\omega \cong j\omega_n$$

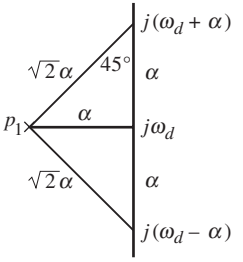
Similarly, the zero at the origin effects little change, and the net effect of the conjugate pole and zero is

$$\left.\frac{j\omega}{j\omega - p_2}\right|_{j\omega \cong j\omega_n} \cong \frac{1}{2}$$

These *narrowband approximations* assume that the poles are near $\pm j\omega_n$ and that the frequency range for $j\omega$ is around $j\omega_n$. The second-order resonant response is consequently reduced to a first-order approximation of the pass-band response:

$$\left.\frac{s\omega_n}{(s-p_1)\cdot(s-p_2)}\right|_{s\cong j\omega_n} \cong \frac{1}{2}\cdot\frac{\omega_n}{j\omega - p_1}$$

Application of the narrowband approximations effects a band-pass to low-pass filter transformation. The first-order result is the response centered around $j\omega_n$ instead of the s-plane origin.



A critical parameter of tuned circuits is their bandwidth relative to their resonant frequency. The less damped a resonant circuit is, the narrower its bandwidth and the more selective its response to a particular frequency channel. On the above plot, a closer view of the s-plane near p_1 is shown. As previously defined, bandwidth is the frequency at which the magnitude of the response

rolls off to $1/\sqrt{2}$ of its low-frequency value. In this case, two frequencies are centered about $j\omega_d$ where roll-off to $1/\sqrt{2}$ occurs. For bandpass response, bandwidth is defined by those frequencies. The magnitude of the passband response rolls off to bandwidth magnitude when $\|j\omega - p_2\| = \sqrt{2}$. At this vector length, the pole angles are 45°, and by geometry the bandwidth frequencies are $j(\omega_d + \alpha)$ and $j(\omega_d - \alpha)$. Under the narrowband approximation, the bandwidth frequencies are

$$j\omega \cong j(\omega_n \pm \alpha)$$

The bandwidth is consequently

$$\omega_{bw} = (\omega + \alpha) - \omega_n = \alpha$$

And the width of the passband is

$$\Delta\omega = (\omega_n + \alpha) - (\omega_n - \alpha) = 2 \cdot \alpha = 2 \cdot \omega_{bw}$$

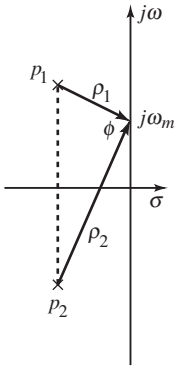
From the pole roots,

$$\alpha = \frac{\omega_n}{2 \cdot Q} \Rightarrow Q = \frac{\omega_n}{2 \cdot \alpha} = \frac{\omega_n}{\Delta\omega}$$

In this formula, the significance of expressing ζ as Q is made explicit; Q is the selectivity. The larger Q is, the narrower the bandpass width relative to the center frequency.

A geometric interpretation (Angelo 1969) in the *s*-plane also eases locating the frequency of maximum magnitude or gain ω_m for a complex pole-pair. When the pole vectors form a right angle at ϕ, the vertex on the $j\omega$ axis is at $j\omega_m$. Let the vectors be ρ_1 and ρ_2, as shown. Then the goal is to maximize the magnitude response,

$$\frac{\omega_n^2}{\|\rho_1\| \cdot \|\rho_2\|}$$



It is maximum when $\|\rho_1\| \cdot \|\rho_2\|$ is minimum. From geometry, the area of the triangle that is formed by ρ_1, ρ_2, and the vertical (dashed) line between the poles, of length $2 \cdot \omega_d$, is

$$\begin{aligned} A &= \frac{1}{2} \cdot \alpha \cdot (2 \cdot \omega_d) \\ &= \alpha \cdot \omega_d \\ &= \frac{1}{2} \|\rho_1\| \cdot \|\rho_2\| \cdot \sin\phi \end{aligned}$$

As the $j\omega$ vertex moves, the area remains constant. The magnitude response is thus

$$\frac{\omega_n^2}{2 \cdot \alpha \cdot \omega_d} \cdot \sin\phi \cong \frac{\omega_n}{2 \cdot \alpha} \cdot \sin\phi = Q \cdot \sin\phi$$

When $\phi = 90°$, sin ϕ is maximum as is the response. At ω_m, the peak magnitude is Q. From the Pythagorean theorem,

$$\omega_m^2 = \omega_d^2 - \alpha^2$$

The triangles themselves are not physically significant but are a mnemonic device for reasoning in the s-plane.

Cascaded stages of identical tuned circuits improve selectivity. This scheme of *synchronous tuning* has a bandwidth reduction factor previously derived in *Designing High-Performance Amplifiers*, "Mutliple-Stage Response." The factor, for n stages, is

$$\sqrt{2^{1/n}-1}$$

In this case, bandwidth reduction is desirable because it improves selectivity.

A shunt RLC is a parallel resonant circuit with an impedance of

$$Z_p = \frac{sL}{s^2LC + s(L/R) + 1} = \frac{s/C}{s^2 + s(1/RC) + 1/LC}$$

with parameters

$$\omega_n = 1/\sqrt{L \cdot C}, \quad Q = \frac{R \cdot C}{\sqrt{L \cdot C}} = \frac{R}{Z_n}, \quad Z_n = \sqrt{\frac{L}{C}}$$

The band-pass width is $\Delta\omega = 1/RC$ and is not affected by L. Thus L can be adjusted to tune the circuit without affecting $\Delta\omega$. These parameters describe a circuit in which Z_p is driven by a current source. For example, it can be a collector or drain load of a tuned amplifier stage.

A more accurate model of an LC parallel-resonant ("tank") circuit, commonly found in radios, includes the series resistance of the inductor, R_s. We then have three parallel branches with admittance,

$$Y = sC + \frac{1}{sL + R_s} + \frac{1}{R_p}$$

Solving for $Z = 1/Y$ gives

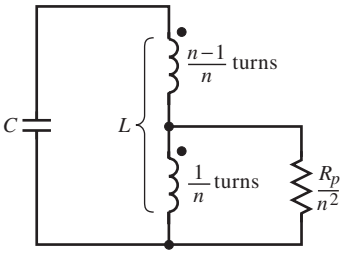
$$Z = (R_s \| R_p) \cdot \frac{s(L/R_s) + 1}{s^2(LC) + s[(L/R_p) + R_sC] + 1}$$

As usual, $\omega_n = 1/\sqrt{LC}$, but Q is now

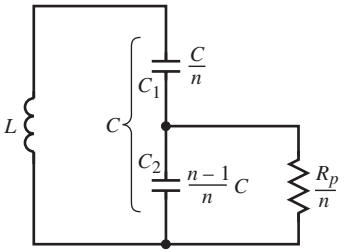
$$Q = \frac{1}{\sqrt{L/C}/R_p + R_s/\sqrt{L/C}}$$

Q is infinite when R_p is infinite and R_s is zero.

For large Q, the parallel-resonant circuit parameters suggest that L must be small or C large. Parasitic elements associated with components limit the practical range of values of L or C. Also, interstage loading resistance can be too small to allow high Q. In these cases, impedance transformation in the resonant circuit is often the solution.



A tapped inductance transforms load resistance R_p/n^2 to R_p across the shunt LC. The inductor is an autotransformer with a high coupling coefficient ($k \cong 1$). The mutual inductance causes the LC voltage to be n times that across the resistance, where n is the turns ratio of the total to bottom windings. The capacitor current is increased n times, causing R_p to appear $1/n^2$ times smaller across the lower winding.



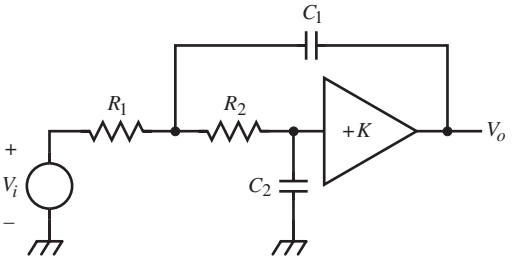
A capacitive divider is used in a similar way except that the capacitors do not have a mutually coupled field. The impedance across the inductor terminals is, for $\omega >> 1/(R_p/n)\cdot(C_1 + C_2)$,

$$Z = \frac{1}{sC_1} + \frac{1}{sC_2}\left\| \frac{R_p}{n} = \frac{1}{s(C_1 \| C_2)}\right\| \frac{R_p}{n}\cdot\left(\frac{C_1 + C_2}{C_1}\right)$$

The impedance shunting L is a series C_1C_2 branch shunting R_p. If the equivalent shunt LC resistance is R_p, as assumed, then n must be

$$n = \left(\frac{C_1 + C_2}{C_1}\right)$$

Two of the most popular op-amp second-order filters are the *Sallen-Key* and *multiple-feedback* filters.



The Sallen-Key LP filter, shown above, can be analyzed as a feedback amplifier for voltage gain or by application of KCL and divider formulas. The transfer function is

$$\frac{V_o}{V_i} = \frac{K}{s^2(R_1R_2C_1C_2) + s\{[1-K]\cdot R_1\cdot C_1 + (R_1 + R_2)\cdot C_2\} + 1}$$

Let the amplifier be a ×1 buffer. Then the transfer function collapses to

$$\frac{V_o}{V_i} = \frac{1}{s^2(R_1R_2C_1C_2) + s(R_1 + R_2)\cdot C_2 + 1}, \quad K = 1$$

An op-amp need not be used for the buffer; in some cases, an emitter-follower is good enough. The resonant frequency is at

$$\omega_n = \frac{1}{\sqrt{R_1R_2C_1C_2}}$$

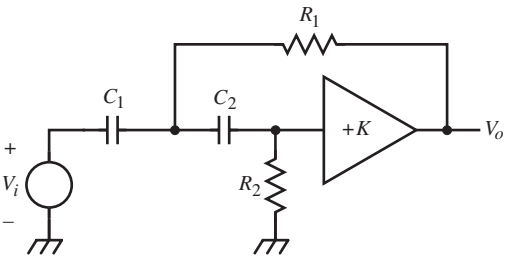
and

$$Q = \sqrt{\frac{(R_1 \| R_2)\cdot C_1}{(R_1 + R_2)\cdot C_2}}$$

For minimal waveform distortion, a Bessel or maximally flat envelope delay (MFED) response requires a Q corresponding to a pole angle of 30° ($\zeta = \sqrt{3}/2$) or

$$Q(\text{MFED}) = \frac{\sqrt{3}}{3} \cong 0.577$$

and, from ω_n and Q, $(R_1 \| R_2)\cdot C_1/(R_1 + R_2)\cdot C_2 = 1/3$.



This high-pass filter (above) has the general topological form of the low-pass, but with transfer function

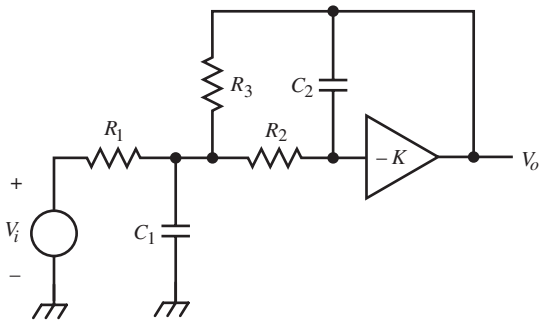
$$\frac{V_o}{V_i} = K \cdot \frac{s^2 R_1 R_2 C_1 C_2}{s^2(R_1 R_2 C_1 C_2) + s\{R_1 C_1 + [(K-1)/K] \cdot R_2 C_2\} + 1}$$

For $K = 1$,

$$\frac{V_o}{V_i} = \frac{s^2 R_1 R_2 C_1 C_2}{s^2 R_1 R_2 C_1 C_2 + s R_1 C_1 + 1}, \quad K = 1$$

Compared with the LP filter, ω_n is the same and

$$Q = \sqrt{\frac{R_2 C_2}{R_1 C_1}}$$

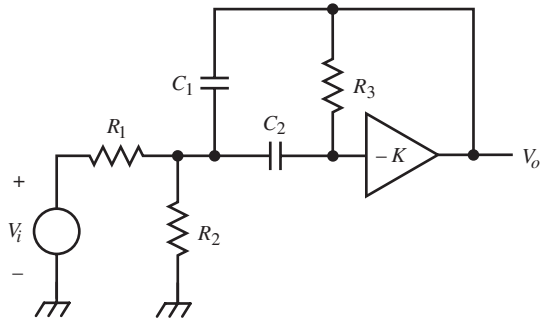


The multiple-feedback topology has two feedback paths, as in the LP filter shown above. For infinite op-amp gain, the LP filter transfer function is

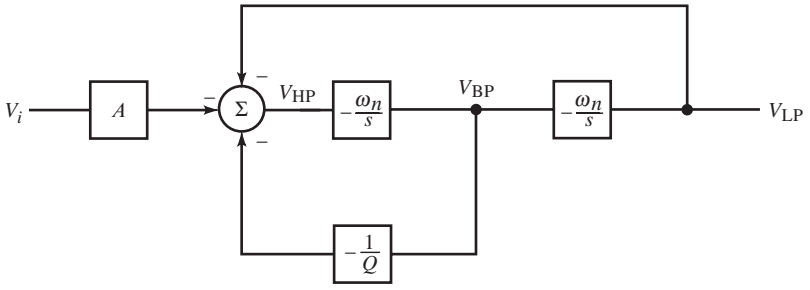
$$\frac{V_o}{V_i} = -\frac{R_3}{R_1} \cdot \frac{1}{s^2 R_2 R_3 C_1 C_2 + s\left\{R_3 C_1 + \left[\left(\frac{R_3}{R_1} + 1\right) \cdot R_2 + R_3\right] \cdot C_2\right\} + 1}$$

For the BP filter, shown below,

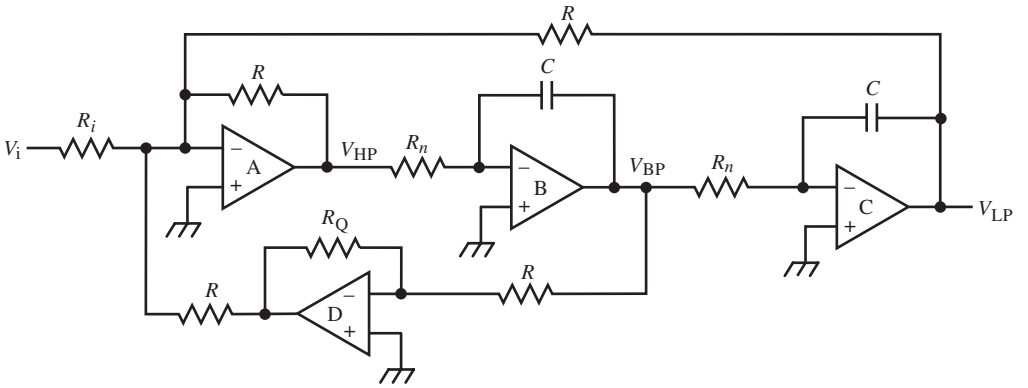
$$\frac{V_o}{V_i} = -\frac{s R_1 C_1^2 \cdot (R_1 \| R_2) \cdot R_3 C_2}{s^2 (R_1 \| R_2) \cdot R_3 C_1 C_2 + s (R_1 \| R_2) \cdot (C_1 + C_2) + 1}$$



These circuits cannot achieve high Q values without appreciable attenuation of the input signal, but they provide a simple second-order filter with good phase linearity. In the multiple-feedback topology, all elements affect both ω_n and Q. In practice, Q is limited to about 20.



The *state-variable* filter topology gets around the design difficulty of interacting filter parameters at the expense of additional circuitry. This scheme is that of the analog computer; cascaded integrators output state variables that are weighted, combined, and fed back or output. The block diagram shown above produces HP, BP, and LP outputs, with corresponding op-amp implementation below.



A quad op-amp IC suffices for gain blocks. Op-amp A is an input summing block, B and C are op-amp integrators, and D is a scaling feedback block. The transfer functions for the three filter outputs are

$$\frac{V_{\mathrm{LP}}}{V_i} = A_{vo} \cdot \frac{1}{(s/\omega_n)^2 + (1/Q\omega_n)s + 1}$$

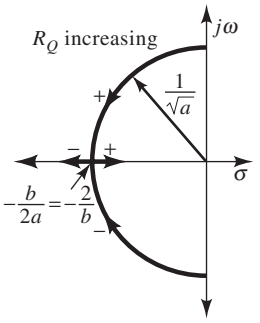
$$\frac{V_{\mathrm{BP}}}{V_i} = A_{vo} \cdot \frac{s/\omega_n}{(s/\omega_n)^2 + (1/Q\omega_n)s + 1}$$

$$\frac{V_{\mathrm{HP}}}{V_i} = A_{vo} \cdot \frac{(s/\omega_n)^2}{(s/\omega_n)^2 + (1/Q\omega_n)s + 1}$$

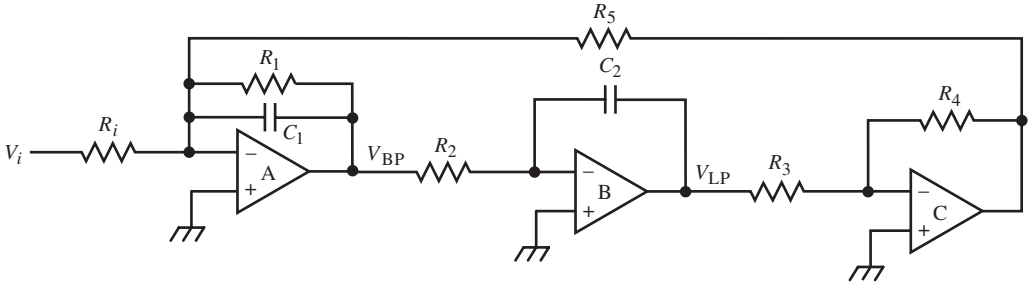
where

$$A_{vo} = -\frac{R}{R_i}, \quad \omega_n = \frac{1}{R_n C}, \quad Q = \frac{R}{R_Q}$$

In the characteristic equation, R_Q occurs in the linear term but not in the quadratic term, thus leaving it free for adjustment of Q independent of ω_n. Its adjustment has the locus of a semicircle centered at the origin.



State-variable filters can achieve a high Q relative to multiple-feedback filters.



A similar filter topology is the *biquad* filter (above), named for the biquadratic form of the transfer function

$$\frac{s^2+ds+e}{s^2+bs+c}$$

It is similar in form to state-variable filters, but damping is adjusted within the cascaded loop of blocks at A in the block diagram above. This topology, like the state-variable filter, has multiple filter outputs; both BP and LP are available. By weighting and combining outputs from two or three of the op-amps in a fourth op-amp output stage, additional filters can be realized.

For the biquad filter,

$$\frac{V_{BP}}{V_i} = -\frac{R_5}{R_i} \cdot \frac{s\left(\dfrac{R_2C_2}{R_4/R_3}\right)}{s^2\left(\dfrac{R_2C_2R_5C_1}{R_4/R_3}\right) + s\left(\dfrac{R_2C_2R_5}{(R_4/R_3)\cdot R_1}\right) + 1}$$

and

$$\frac{V_{LP}}{V_i} = \frac{R_5}{R_i} \cdot \frac{1}{R_4/R_3} \cdot \frac{1}{s^2\left(\dfrac{R_2C_2R_5C_1}{R_4/R_3}\right) + s\left(\dfrac{R_2C_2R_5}{(R_4/R_3)R_1}\right) + 1}$$

The filter parameters are then

$$\omega_n = \sqrt{\frac{R_4/R_3}{R_2R_5C_1C_2}}, \quad Q = R_1C_1\omega_n$$

The biquad filter has an advantage over previous filter topologies in that the band-pass $\Delta\omega$ is adjustable independent of center frequency ω_n. From the expression for Q above,

$$\Delta\omega = \frac{\omega_n}{Q} = \frac{1}{R_1C_1}$$

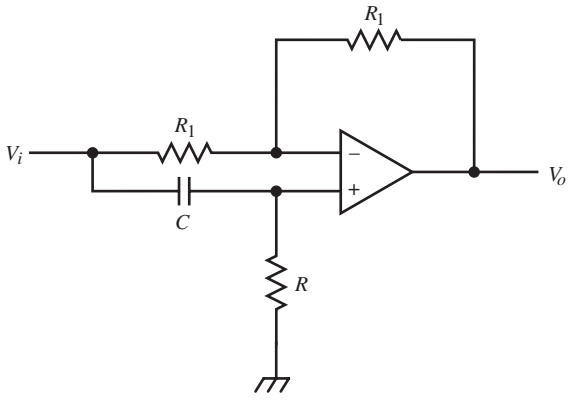
Because ω_n is independent of R_1, it independently adjusts $\Delta\omega$. The band-pass function can be simplified to

$$\frac{V_{BP}}{V_i} = -\frac{R_f}{R_i} \cdot \frac{sR_fC}{s^2(R_fC)^2 + s\left[R_fC\left(R_f/R_1\right)\right] + 1}$$

where

$$R_2 = R_5 = R_f, \quad R_4 = R_3, \quad C_1 = C_2 = C$$

Gain is independently set by R_i, and $\Delta\omega$ by R_1.



One other filter, shown above, is an all-pass filter that operates as a phase shifter. Its transfer function is

$$\frac{V_o}{V_i} = \frac{2sRC}{sRC+1} - 1 = \frac{sRC-1}{sRC+1} = -\frac{-sRC+1}{sRC+1}$$

The delay time through the filter is

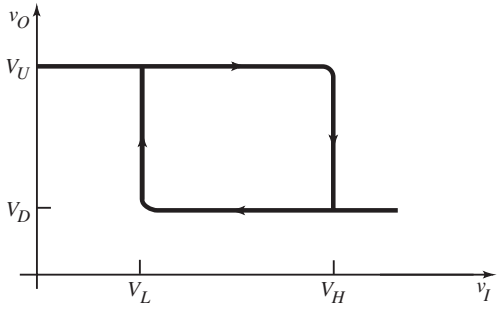
$$t_d = 2 \cdot R \cdot C$$

Phase can be adjusted by adjusting R. This can be done electronically using a CMOS DAC or FET.

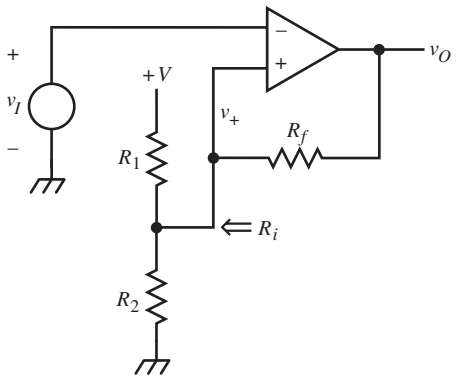
HYSTERETIC SWITCHES (SCHMITT TRIGGERS)

Comparators are usually inadequate for providing a single output transition when the input polarity changes. Slowly changing input signals with some noise cause the output to "dither": to alternate between the high and low states near the input threshold. This dithering is reduced or eliminated by an input deadzone or deadband, an input range around the threshold where no output change can occur. Furthermore, if the deadzone is state dependent,

the effect is called *hysteresis*. The figure below shows a characteristic square hysteresis loop.



The accompanying circuit is an inverting *hysteretic comparator*, or *Schmitt trigger* circuit. The state dependence is achieved by use of positive feedback. In effect, the Schmitt trigger is a bistable memory device.



Assume that the output is in the high state; then $v_O = V_U$. Thevenize the divider R_1, R_2 so that its Thevenin voltage is V_T, the threshold voltage, and its resistance is R_i:

$$V_T = \left(\frac{R_2}{R_1 + R_2}\right) \cdot V, \quad R_i = R_1 \| R_2$$

With v_O high, v_+ is set from the feedback divider R_f, R_i. When v_I increases to where the comparator inputs are equal, the output changes to a low state. This input voltage is V_H, the upper hysteresis threshold. By setting $v_+ = V_H$, by superposition,

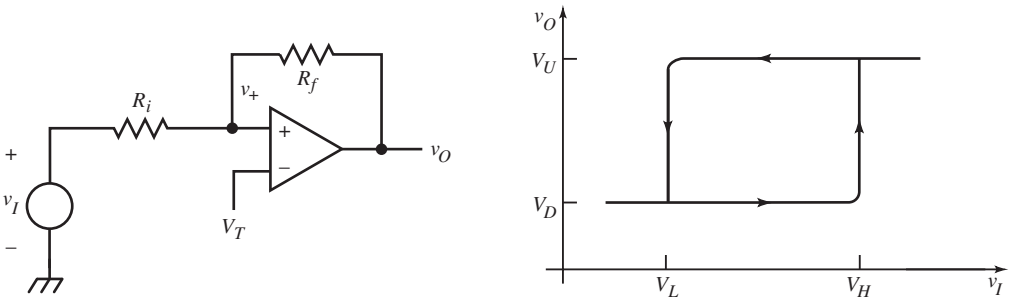
$$V_H = \left(\frac{R_i}{R_f + R_i}\right) \cdot V_U + \left(\frac{R_f}{R_f + R_i}\right) \cdot V_T$$

Once v_O is low, then v_I must decrease to V_L, the lower threshold, before v_O becomes high. By letting $v_+ = V_L$ and again applying superposition,

$$V_L = \left(\frac{R_i}{R_f + R_i}\right) \cdot V_D + \left(\frac{R_f}{R_f + R_i}\right) \cdot V_T$$

The deadzone size is the width of the hysteresis loop. This *hysteresis window* is

$$\text{inverting } \Delta v_I = V_H - V_L = \left(\frac{R_i}{R_f + R_i}\right) \cdot (V_U - V_D)$$



A noninverting form of hysteretic comparator, shown above, has a similar hysteresis loop, but it is traversed in the counterclockwise direction. The circuit has a positive feedback divider and can be analyzed by superposition. When v_O is either high or low, the threshold for v_+ is fixed at V_T.

Hence, we must solve for V_L and V_H after applying superposition. The results are

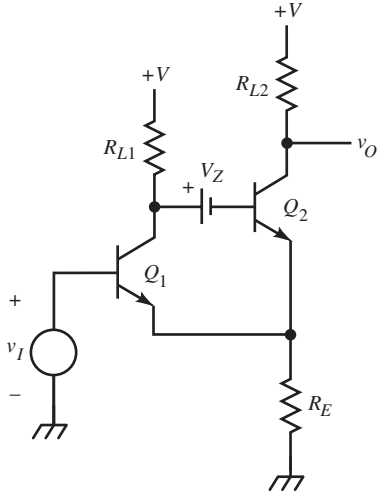
$$V_L = \left(\frac{R_f}{R_i} + 1\right) \cdot \left[V_T - \left(\frac{R_i}{R_f + R_i}\right) \cdot V_U\right] = -\left(\frac{R_i}{R_f}\right) \cdot V_U + \left(\frac{R_i}{R_f} + 1\right) \cdot V_T$$

and

$$V_H = -\left(\frac{R_i}{R_f}\right) \cdot V_D + \left(\frac{R_i}{R_f} + 1\right) \cdot V_T$$

with the following input hysteresis window:

$$\Delta v_I = V_H - V_L = \left(\frac{R_i}{R_f}\right) \cdot (V_U - V_D)$$



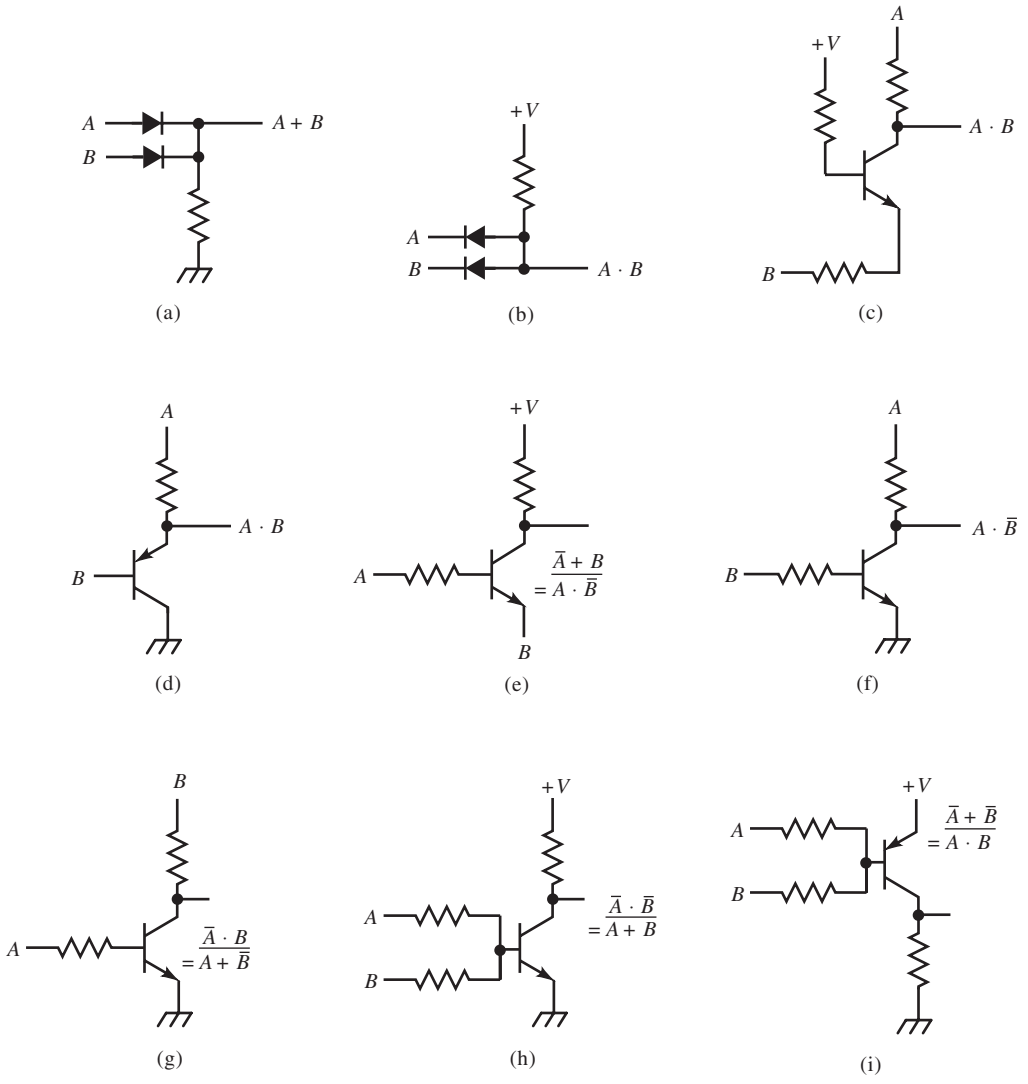
A two-transistor discrete realization of a hysteresis comparator is the *emitter-coupled* Schmitt trigger, shown above. This is a diff-amp with positive feedback from the collector of Q_1 through a voltage translator V_Z to the base of Q_2 returning to the emitter of Q_1. The hysteresis loop goes clockwise. The additional inversion at the collector of Q_2 reverses the direction of the loop at v_O. V_Z provides an additional degree of freedom in setting the thresholds.

This circuit introduces another aspect of hysteretic comparators: As the input approaches the threshold, the diff-amp transconductance increases. Away from the threshold, one of the diff-amp transistors is cut off, and diff-amp transconductance is low. As a result, loop gain is low. But when the threshold is approached, the cut-off transistor begins to conduct, and loop gain increases. When it reaches one, the positive-feedback loop becomes unstable and transitions to the other state. To find the threshold voltages accurately, an iterative solution is required since diff-amp gain changes with input voltage. In other words, a large-signal analysis is necessary using the diff-amp transconductance equation.

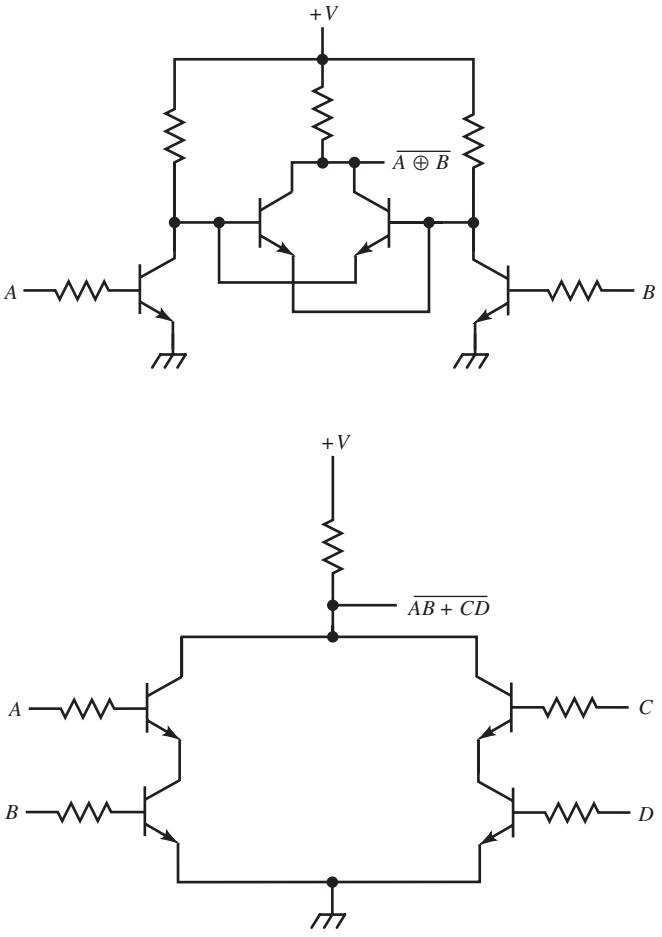
The hysteretic comparator is a positive-feedback amplifier and is inherently unstable. This instability must be controlled in the design; instability is allowed only for changing state. If the loop gain has peaking, then as the input voltage approaches the threshold, a loop gain of one is reached first at ω_m, the frequency at which the loop-gain magnitude peaks. Before the output changes, the loop oscillates with frequency ω_m. Therefore, otherwise stable loops require a loop gain without peaking.

Discrete Logic Circuits

Even analog designs are likely to require some logic functions. In discrete designs, it is often unnecessary to add logic ICs to perform simple logic functions. The circuits below are a variety of diode and BJT logic circuits using only two diodes or one transistor.



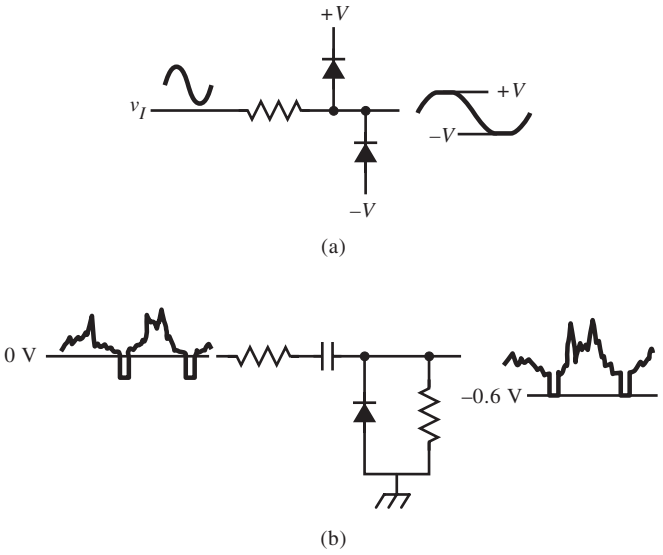
In the figure below, two more circuits are shown, using four transistors, to realize an exclusive-nor and an and-or-invert (AOI) gate below it.



These circuits have no particular merits other than their simplicity. Discrete realization of logic circuits is sometimes advantageous among analog circuitry when BJTs (or circuit-board space) are left over in an array and a gate or two is required.

Clamps and Limiters

Nonlinear circuits that modify waveforms in some manner involving limits are *clamps* or *limiters*. Depending on the particular application, they might have other names.

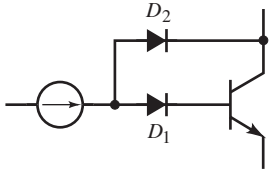


In (a), diodes are used to limit the range of v_I by "clipping" the signal outside the range of $\pm V$. This circuit is commonly used as an input protection circuit in MOS ICs and oscilloscope trigger inputs. It is sometimes called a *clipping circuit*. Figure (b) shows a type of clamp that establishes a waveform at a given voltage level. An application is as a *baseline restorer* in video signal processing. The negative extremes (sync tips) are established near ground by the clamp diode.

An important application of clamps is to keep transistors from saturating. Small-signal saturated transistors have excess charge in their base from being overdriven. This charge must be removed to turn the transistor off, and with limited base-current drive the base storage time increases. This causes a delay in turn-off. Falltime is not significantly affected.

In large-signal (power) transistors, although excess base charge is a storage-time factor, another effect dominates, causing fall times of unclamped transistors to be larger. With large collector currents, a BJT operates in the high-level injection region, where the collector minority-carrier concentration (majority carriers from the base) approaches that of the collector majority concentration.

Under high-level injection, the collector side of the *b-c* junction actually inverts in charge polarity due to a dominance of carriers from the base. This *Kirk effect* causes conductivity modulation of the base, causing the base width to effectively increase. This effect occurs at a v_{CE} just above saturation, in the quasisaturation region, and causes excess rounding in the collector family of curves at low v_{CE} (of a few volts). (A related effect, called *crowding*, is due to ohmic v_{BE} drop laterally across the base, which causes less conduction in the center of the base than in the outer ring closer to the base contact.) Conductivity modulation also affects the fall time. As excess charge is swept out of the base, the excess base width begins to decrease, and turn-off commences. By decreasing excess base drive, conductivity modulation also decreases, along with both storage and fall times.



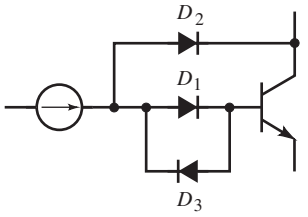
The *Baker clamp* prevents saturation by adding two diodes to a BJT. The *b-c* junction is a diode in series with D_1. Together, they conduct with a voltage drop of two junctions. D_2 shunts them and conducts at a lower voltage, with one junction drop. Thus, D_2 clamps the D_1-*b-c* path and keeps it from conducting.

Quantitatively, by KVL,

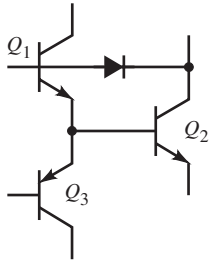
$$v_{CE} \cong v_{BE} + v_{D1} - v_{D2} \cong v_{BE}$$

For approximately equal diode voltage drops, v_{CE} is clamped at v_{BE} or with $v_{CB} \approx 0$ V. Transistors with appreciable r'_c or large collector current may require an additional diode in series with D_1 to clamp v_{CB} at about a junction drop.

The disadvantage of the Baker clamp is the higher on-state v_{CE}, typically a half volt. The *Schottky clamp* has reduced v_{CE}(on). It is a variation of the Baker clamp and is very simple: A Schottky diode, with forward voltage of about 0.4 V, shunts the *b-c* junction. When v_C decreases to 0.4 V below the base voltage, the Schottky diode turns on, clamping v_{CE} at about 0.3 V. In Schottky logic output stages, this clamp prevents hard saturation while allowing lower v_{CE}(on) than a standard Baker clamp.



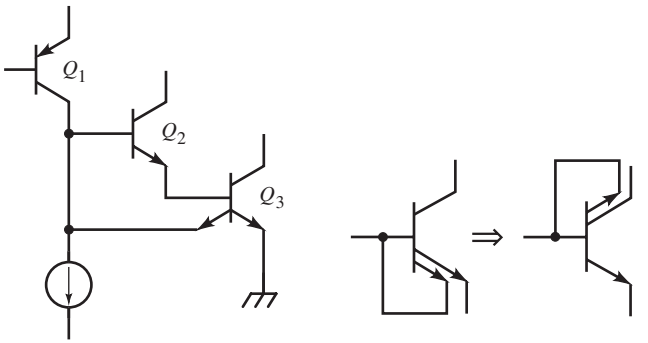
As D_1 is in series with the *b-e* junction, no turn-off path exists without a shunt *b-e* resistor or, as shown above, a diode shunting D_1 and reversed (or antiparallel) relative to D_1.



If a complementary CC driver is used to drive the clamped transistor, as shown above, only one added diode is required for the Baker clamp. The *b-e* junction of Q_1 takes the place of D_1 in the previous circuit and the *b-e* junction of Q_3 for

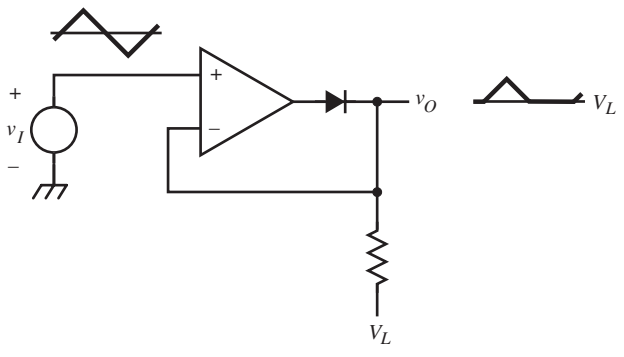
D_3. The diodes themselves also must have suitably fast turn-off recovery capability.

The feedback clamp scheme shown below, used in ICs, is a form of the *Murphy clamp*.

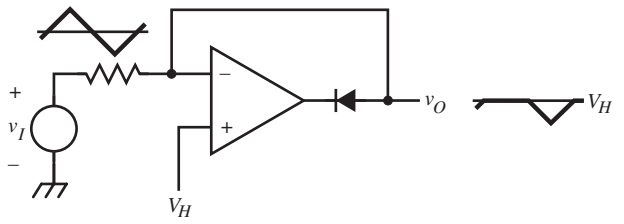


A second emitter is added to the transistor to operate as a clamp. This anti-saturation emitter is connected back to the base. Because the *b-c* junction, as a diode, also points outward (as does the emitter arrow), the second emitter can be regarded as a second collector, as shown. It is more heavily doped than the collector and shunts current from it at low v_{BC}. When the collector voltage decreases to near saturation, the second emitter dominates; for the same reverse bias, it has more minority carriers to inject into the base. As a result, the *b-c* junction does not conduct heavily in the forward direction, avoiding hard saturation.

In the Murphy clamp, Q_3 avoids saturation by its second emitter. This emitter is reverse-biased by v_{BE2}, allowing v_{CB3} to decrease to v_{BE2} before the anti-saturation emitter takes effect. In addition, the shunted collector current is taken from the input drive current. The current gain of Q_2 amplifies this limiting action, resulting in a "sharp" limiting response.

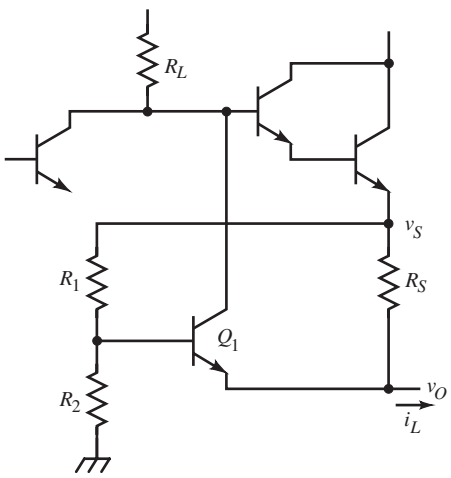


(a)



(b)

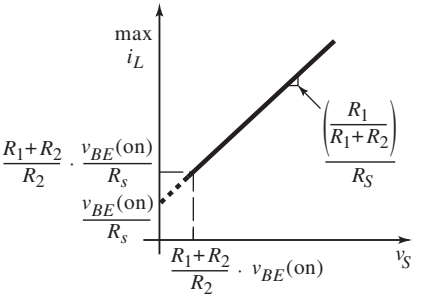
Diodes and op-amps are combined in these precision clamps. They use the op-amp input as a comparator when the loop is open and diode nonconducting: in (a), when $v_I < V_L$, and in (b), when $v_I > V_H$. For the input ranges where the diodes conduct, the op-amps operate as ×1 buffers. In effect, these clamps are half-wave rectifiers with programmable limiting voltages.



A common limiter in power supply circuits is the *foldback current limit*, with one realization of it shown above. As load current i_L increases, the voltage drop across R_S increases until the *b-e* junction of Q_1 is turned on. The transistor conducts, diverting base drive from the series regulator Darlington. The current is limited to

$$\max i_L = \frac{v_{BE}(\text{on}) + v_S \cdot [R_1/(R_1 + R_2)]}{R_S}$$

where $v_{BE}(\text{on})$ is the v_{BE1} value for which the loop gain is barely sufficient to sustain the limiting value of i_L. The maximum current depends on v_S, graphed below.

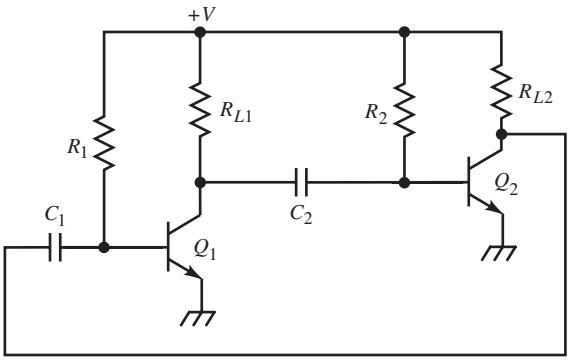


The current limit decreases or "folds back" with reduced voltage so that a short conducts reduced current and thus power dissipation in the Darlington BJTs. As the short "clears" (load resistance increases), more current is allowed.

Multivibrators and Timing Circuits

Multivibrators (MVs) are positive-feedback (or *regenerative*) switching circuits with analog timing of switching behavior. They can be *bistable*, having two stable states (as Schmitt trigger circuits have); *monostable*, having one stable state; or *astable*, having no stable states. Monostable multivibrators (MMVs) are also called "one-shots"; they change output state upon input of a trigger signal. This quasistable state lasts for a timed interval (until the MMV "times out"), at which time it

reverts to its stable state. MMVs are used to generate a triggered pulse of a given duration. Astable MVs are digital oscillators, or *clock generator* circuits, and sometimes are called "free-running" MVs.



The two-BJT astable MV topology consists of two capacitively coupled common-emitter (CE) stages. When v_{C2} goes low, the negative transition is coupled through C_1, cutting off Q_1. R_1, C_1 form an *RC* differentiator, and v_{B1} begins to increase exponentially as C_1 charges through R_1. (Assume that the *b-e* junction of Q_1 is not in reverse breakdown.) When v_B reaches v_{BE1}(on), Q_1 conducts, causing its collector to transition to near ground. This cuts off Q_2 for the second half-cycle of oscillation in the same way Q_1 was cut off. For identical CE stages, the duty-ratio,

$$D = \frac{t_{on}}{t_{on} + t_{off}} = \frac{t_{on}}{T}$$

is 50%. When Q_1 turns on and drives Q_2 off, v_{C2} goes high (to $+V$), and C_1 is charged in the other direction through R_{L2} and the *b-e* junction of Q_1. This recharging time constant must be shorter (at least by 1/5) than R_1C_1 to fully recharge the capacitor. Therefore, R_1, R_2 must be at least five times R_{L1}, R_{L2}.

MV frequency is increased by reducing coupling capacitor *C*. To analyze the effect of C_{be} on timing, assume that the base timing resistor *R* is a current source

I instead. This assumption is valid for off-times that are much less than the time constant RC when the exponential base voltage can be approximated as linear. Then C and C_{be} form a capacitive voltage divider so that a negative input step of $-V$ causes v_B to step down to

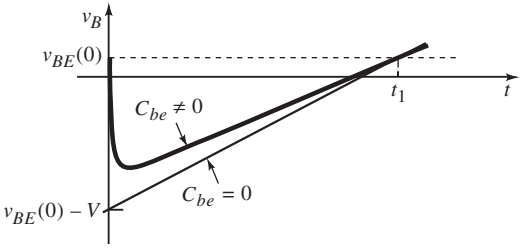
$$v_B(0^+) = -V \cdot \left(\frac{C}{C + C_{be}} \right) + v_{BE}(0)$$

Without the effects of C_{be}, the timing ramp of v_B has a slope of I/C and times out at

$$t_1 = \frac{C}{I} \cdot V$$

With C_{be},

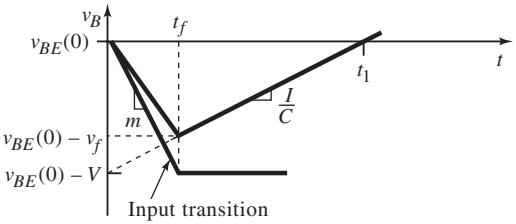
$$v_{BE}(t) = \left(\frac{I}{C + C_{be}} \right) \cdot t - v_B(0^+), \quad t > 0$$



The offset of $v_{BE}(0)$ does not affect timing since it is also the voltage threshold for determining t_1. Set $v_{BE}(t_1) = v_{BE}(0)$ and solve for t_1; it is the same as the previous t_1. Therefore, C_{be} does not affect timing.

The previous analysis assumed zero fall time of the negative input step. To examine this assumption, consider the effect of switching time on timing by

approximating the negative transition at the collector with a linear approximation to the waveshape and its average slope m.



Let the transition time be t_f. Then, from the above v-t graph, the slower the transition, the smaller the negative excursion of the timing voltage, $-v_f$; thus,

$$t_1 - t_f = \frac{C}{I} \cdot v_f$$

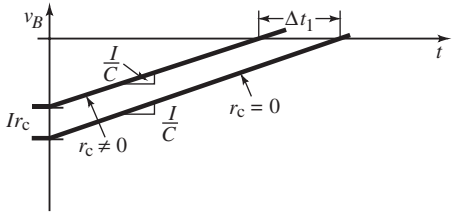
and

$$v_f = -\left(\frac{I}{C} + m\right) \cdot t_f$$

with $m = -V/t_f$. Substituting and solving for t_1,

$$t_1 = \frac{C}{I} \cdot \left(\frac{-I}{C} + \frac{V}{t_f}\right) \cdot t_f + t_f = \frac{C}{I} \cdot V$$

Therefore, when $t_f < t_1$, the falltime has no effect on timing.

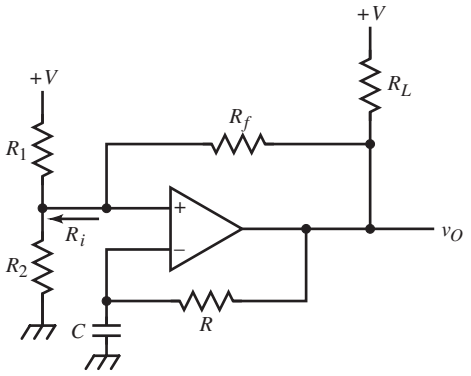


The final timing analysis determines the effect of collector resistance r_c of the conducting transistor. From the above graph, although the slope of $v_B(t)$ is independent of r_c, the initial step size is not. With r_c, I causes an opposing step of $I \cdot r_c$ that translates into a time error of

$$\Delta t_1 = \frac{I \cdot r_c}{I/C} = r_c \cdot C$$

The fractional error is $I \cdot r_c / V$. Consequently, to minimize r_c, the BJTs must be driven well into saturation by R_1 and R_2.

The two-BJT astable MV timing is independent for each half-cycle. This makes adjustment of half-cycle timing easy, but since the period depends on both half-cycles, each can contribute to period error. Two separate timing circuits are not necessary.



The astable MV above, based on an open-collector comparator such as the LM393, has only one timing capacitor. This single-supply clock generator uses regenerative feedback through R_f to effect a hysteretic comparator while the timing is done by R and C. When $v_O = V_U$ (high state), the timing waveform

at v_- is increasing and crosses V_H at the v_+ input. Then v_O goes low (to V_D), and the timing waveform decreases toward it while $v_+ = V_L$. When it reaches V_L, v_O switches high again.

For single-supply operation, R_1 and R_2 set a voltage around which the input hysteresis limits of V_L and V_H are chosen. When $v_O = V_U$, the timing resistance is

$$R_U = R + R_L \| \left(R_f + R_i \right), \quad R_i = R_1 \| R_2$$

V_U and V_H are calculated from the divider formed by R_L, R_f, and R_i. The timing voltage v_- begins at V_L and heads for V_U. The high-state time duration is

$$t_H = R_U \cdot C \cdot \ln\left(\frac{V_U - V_L}{V_U - V_H} \right)$$

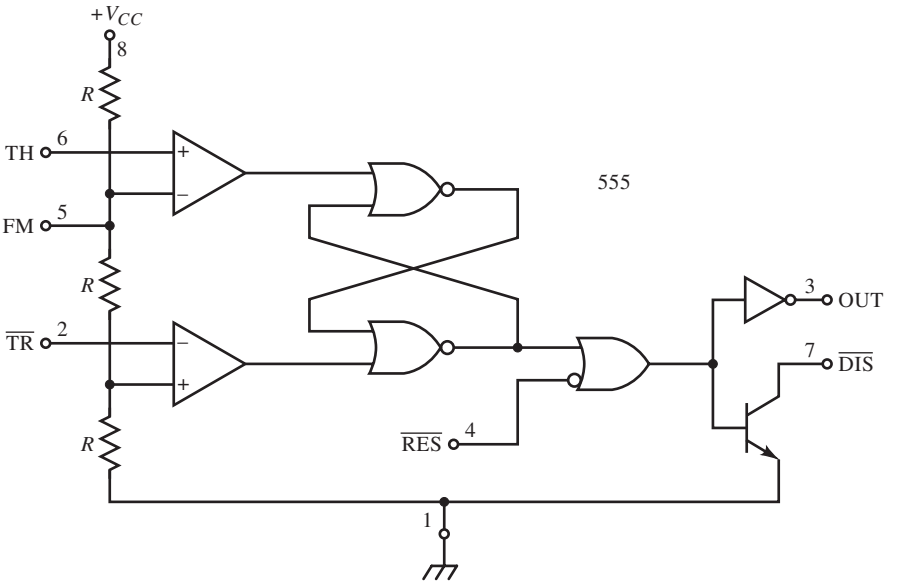
During the low output state, $v_O = V_D$ and $v_+ = V_L$. The comparator BJT saturates, and $V_D \cong 0$ V. Then v_- decreases from V_H toward V_D, and

$$t_L = R \cdot C \cdot \ln\left(\frac{V_H - V_D}{V_L - V_D} \right) \cong R \cdot C \cdot \ln\left(\frac{V_H}{V_L} \right)$$

Finally, the output period is:

$$T = t_H + t_L$$

One of the most versatile MV circuits is the *timer*, notably the 555 bipolar and 7555 CMOS ICs, with block diagram shown below. This timer is mainly applied as a MMV or clock generator. It consists of two comparators with trigger (/TR) and threshold (TH) inputs. The other inputs are taken from a resistive divider from the supply, V_{CC}. For the bipolar version, $R = 5$ kΩ; for the CMOS version it is $R > 100$ kΩ.



The threshold-comparator input threshold voltage is

$$V_H = \frac{2}{3} \cdot V_{CC}$$

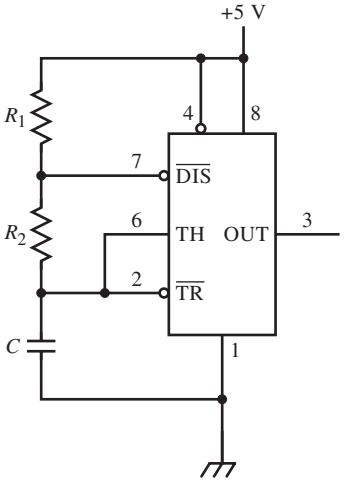
The trigger-comparator voltage is

$$V_L = \frac{1}{3} \cdot V_{CC}$$

The comparator outputs drive a NOR-gate RS flip-flop (FF). The trigger input overrides the threshold input for control of the output. When /TR is asserted [v(/TR) < V_L], the output is forced high. The reset input (/RES), however, overrides all other inputs. A separate transistor (BJT in 555, MOSFET in 7555) with open collector (or drain) marked /DIS (for discharge) is an alternative output for MV control.

An astable MV circuit using the timer is shown below. Initially, /TR is low and the output is high. The discharge transistor is off. C charges through R_1 and R_2

until the timing voltage crosses V_H. The output goes low, /DIS sinks current, and C discharges to ground through R_2. When the timing voltage crosses V_L, the cycle repeats.



The time duration for a high output is

$$t_H = (R_1 + R_2) \cdot C \cdot \ln\left(\frac{V_{CC} - V_L}{V_{CC} - V_H}\right) = (R_1 + R_2) \cdot C \cdot \ln 2 \cong (0.693) \cdot (R_1 + R_2) \cdot C$$

For a low output,

$$t_L = R_2 \cdot C \cdot \ln\left(\frac{V_H}{V_L}\right) = R_2 \cdot C \cdot \ln 2 \cong (0.693) \cdot R_2 \cdot C$$

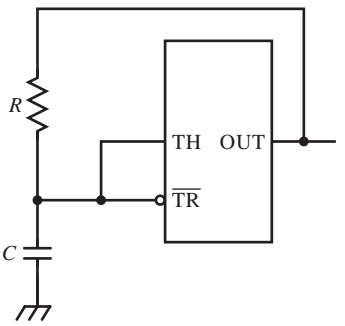
The period is therefore

$$T = t_H + t_L = (R_1 + 2R_2) \cdot C \cdot \ln 2 \cong (0.693) \cdot (R_1 + 2R_2) \cdot C$$

with duty-ratio

$$D = \frac{R_1 + R_2}{R_1 + 2 \cdot R_2}$$

Given $t_L < t_H$, a duty-ratio of $D < 0.5$ is not possible with this circuit. An alternative circuit achieves a longer t_L by placing an additional resistor in series with /DIS. A CMOS alternative with accurate 50% duty-ratio is shown below.



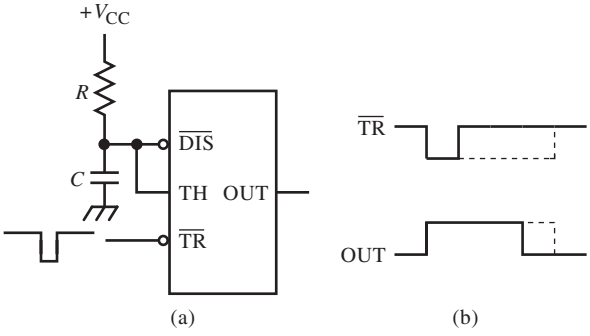
The output is used instead of /DIS to control timing, and the timing elements are the same for both half-cycles. The period is

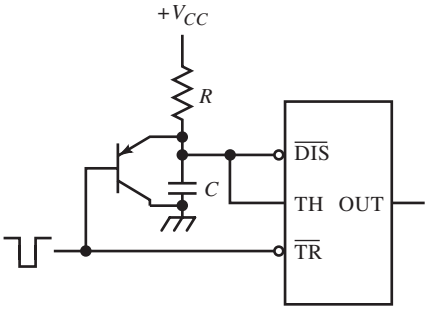
$$T = 2 \cdot R \cdot C \cdot \ln 2 = 1.386 \cdot RC$$

A timer-based MMV, (a) below, is triggered by a negative-going pulse. It sets the output high, and C begins to charge through R. The initial voltage on C is 0 V, and it charges to V_H. The time-out is thus

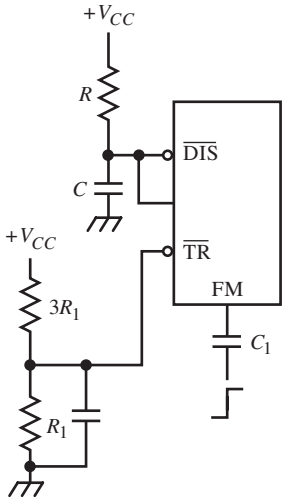
$$t_H = -R \cdot C \cdot \ln\left(1 - \frac{V_H}{V_{CC}}\right) = R \cdot C \cdot \ln(3) \cong 1.1 \cdot R \cdot C$$

If the trigger pulse duration exceeds t_H, the output is kept high until the trigger goes high, as in (b).



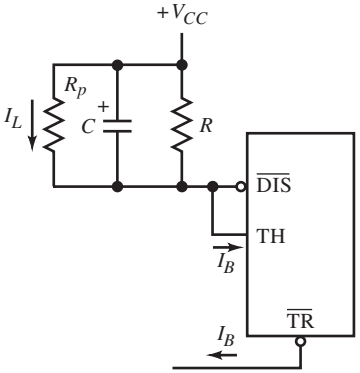


A MMV that begins its time-out after the trigger pulse goes high is shown above. The additional PNP transistor keeps C discharged until the trigger releases. This MMV is retriggerable in that the output remains high as long as trigger pulses continue to occur before time-out. Each new pulse resets the timing and retriggers the MMV.



The FM terminal (pin 5) allows control of the comparator thresholds, where a positive transition on the FM terminal triggers the MMV. The /TR input is biased at $V_{CC}/4$. The trigger-comparator divider voltage is raised by the positive step of the trigger to where the step exceeds $V_{CC}/2$ and starts the time-out.

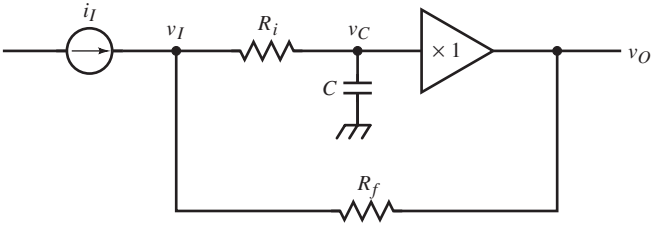
For long time-outs, a large R, a large C, or both are needed. For a large R, timer-comparator bias currents can cause timing error. The bipolar timer has an NPN diff-amp input stage in its threshold comparator and a PNP stage for the trigger comparator. The bias currents I_B place a limit on the minimum charging current. For large C, electrolytic capacitors are likely to be used. Their limitation is their leakage current. Aluminum electrolytics typically leak at 20 nA/μF; tantalum leaks at 5 nA/μF. If C is placed as shown below, its leakage current I_L path is modeled by a shunt resistance R_p, shunting the timing resistor R. The maximum limit on R is thus set by R_p. This circuit has less timing noise (*jitter*) because the target voltage of C is 0 V instead of V_{CC}, the same voltage to which R_p would discharge C. When C is ground-based, R_p opposes charging instead of aiding it.



Capacitance and Resistance Multipliers

Timing circuits with long time-outs often require large capacitors. For accurate timing, these capacitors are plastic. Large-value plastic capacitors are volumetri-

cally large and expensive. The *capacitance multiplier* is a circuit that uses gain to make a small capacitor appear electrically large. One realization is shown below.



Here, a current source drives the C multiplier to generate a ramp waveform. The ×1 buffer causes R_i and R_f to form a current divider because it keeps the voltage across R_f the same as across R_i. The charging current is a fraction of the input current, or

$$\frac{i_C}{i_I} = \frac{R_f}{R_f + R_i}$$

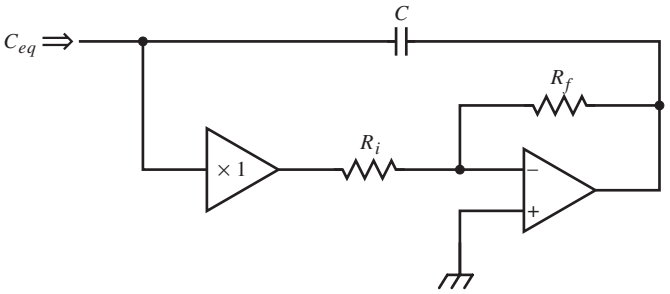
The equivalent capacitance C_{eq} is based on the relation

$$i_I = C_{eq} \cdot \frac{dv_O}{dt} \Rightarrow \frac{i_I}{C_{eq}} = \frac{dv_O}{dt} = \frac{i_C}{C}$$

The last equality follows because $v_O = v_C$. Applying the current fraction,

$$C_{eq} = \left(\frac{i_I}{i_C}\right) \cdot C = \left(\frac{R_i}{R_f} + 1\right) \cdot C$$

This is a "transcapacitance" multiplier because v_O is not across the same terminals that i_I flows through. The input is C_{eq} in series with $R_f \| R_i$.

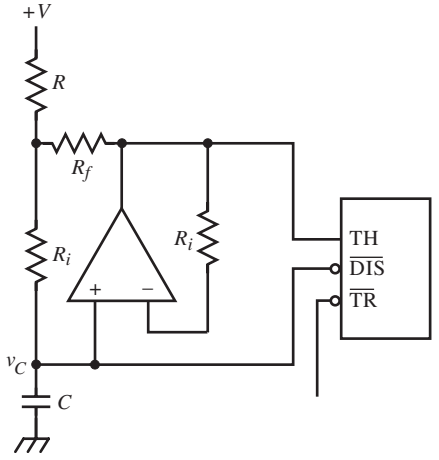


A true capacitance multiplier (above) can be based on the Miller effect. The ×1 buffer drives an inverting op-amp with a gain of $-R_f/R_i$. Applying Miller's theorem,

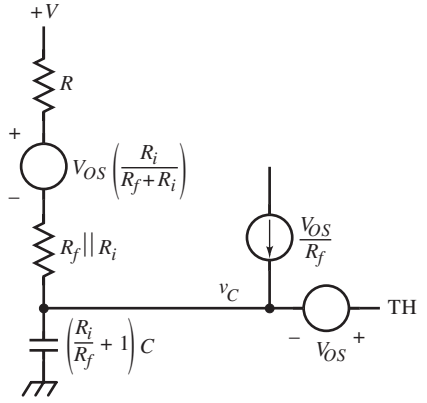
$$C_{eq} = \left(\frac{R_f}{R_i} + 1\right) \cdot C$$

Above the bandwidth of the amplifier branch, the input is no longer purely capacitive but also has a shunt *RL* in series with C_{eq}.

Example: Timer with Capacitance Multiplier



The threshold-terminal bias current of a 555 timer limits its useful timing range as an MMV. A current-divider capacitance multiplier is used to extend the time-out t_H by connecting it as shown above. The TH input is now driven by the op-amp output, eliminating bias-current error from TH but introducing op-amp offset current and voltage error. An equivalent circuit shown below is derived as follows.



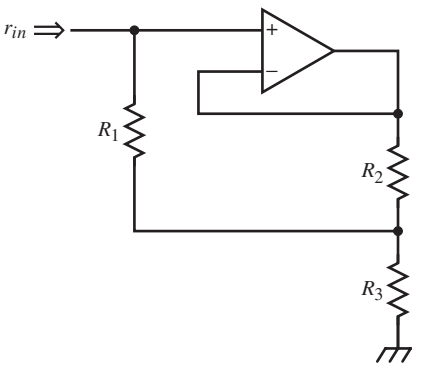
The op-amp circuit with R_f and R_i is Thevenized and floated on v_C. The op-amp offset voltage V_{OS} is divided by the resistors so that its Thevenin voltage is

$$V_{OS} \cdot \left(\frac{R_i}{R_f + R_i} \right)$$

in series with $R_f \| R_i$. Because the op-amp has ×1 gain, V_{OS} also is in series with the TH input. In addition, V_{OS} at the op-amp output contributes V_{OS}/R_f to the timing current. From this model, timing error is calculated.

For significant multiplication of C, $R_i >> R_f$. The series Thevenin voltage source is then about V_{OS}, and the I_{OS} term in the error current dominates.

Capacitance multiplication is achieved in the preceding circuits by applying bootstrapping to a current divider. The idea can be extended to resistance



multiplication. A basic instance is the bootstrapped CC or FET CD stage (in "The Effect of Base-Emitter Shunt Resistance" of *Designing Amplifier Circuits*).

A more deliberate and precise resistance multiplier, shown above, uses an op-amp buffer instead of a CC or common source (CS) amplifier. Solving the circuit for r_{in},

$$r_{in} = R_1 \cdot \left(1 + \frac{R_3}{R_2}\right) + R_3$$

Trigger Generators

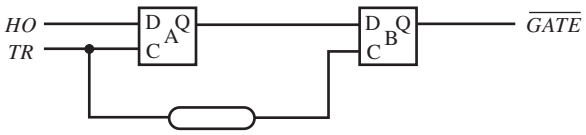
A trigger generator is a kind of precision synchronizer. In television deflection systems, horizontal and vertical scans or *sweeps* of the CRT are synchronized to the video signal by perturbing a free-running oscillator with the synchronization (or *sync*) signal, forcing it to lock to the sync frequency. The sync pulse also corrects the phase each cycle.

In synchronous digital systems, asynchronous events must be synchronized to the system clock. This is usually done using a flip-flop (or *flop*) clocked by the system clock; the asynchronous pulse is the data input to the flop. The problem with this scheme is that if the data pulse changes state too soon before the active edge of the clock, the flop setup time is insufficient and the output state is indefinitely indeterminate. A second flop cascaded with the first can reduce the indeterminate time skew at the expense of one clock period of delay. Of course,

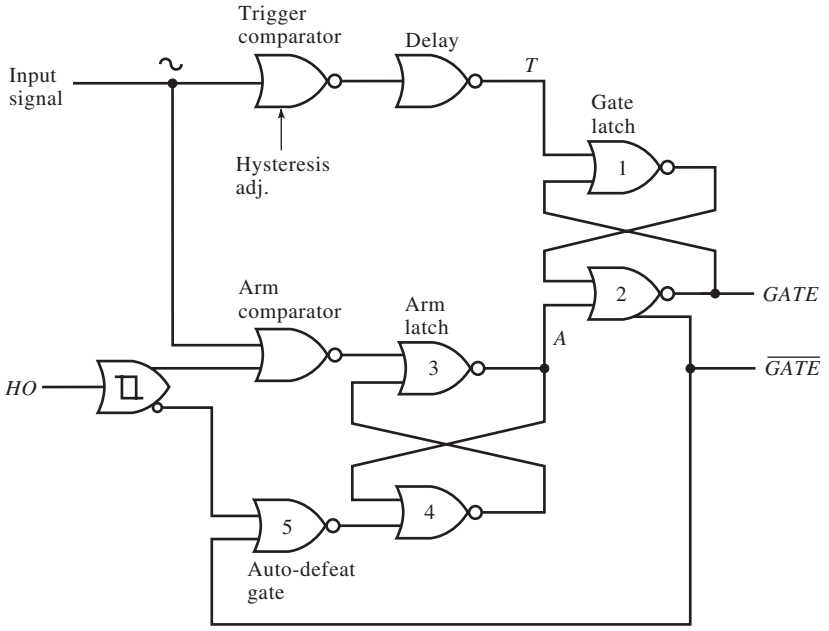
the first flop output may remain indeterminate for longer than one clock period, but the probability diminishes rapidly with time.

This synchronizing problem is especially acute in oscilloscopes. The vertical waveform is fed to an event processor. On a selected slope and at the voltage of the trigger-level control, a comparator generates an *event*, an output transition that is used to start the sweep. When the sweep reaches the right end of the CRT screen, the CRT beam is turned off, and the beam retraces back to the left side, where it settles to the same starting position.

During sweep retrace, a hold-off pulse keeps input waveforms from firing the sweep until it is settled in its starting position. This hold-off pulse is asynchronous with the trigger events. If an event occurs while hold-off is releasing, time skew occurs in starting the sweep; this *trigger jitter* causes successive traces to be horizontally misaligned; the trace appears fuzzy, and multiple traces can be observed.

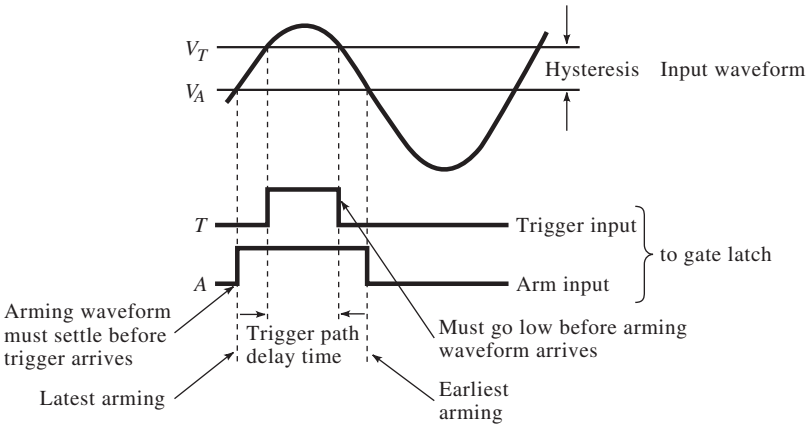


The trigger generator above reduces jitter by synchronizing the trigger events, *TR*, and hold-off pulses, *HO*. The circuit consists of two D flip-flops and a delay device, such as a digital delay-line or the propagation delay of some logic devices. The flops are clocked on positive (rising) edges by *TR*. The sweep gate is asserted low (as */GATE*) and enables the sweep generator. The negative (falling) edge of */GATE* must occur consistently at a fixed delay time relative to the rising edge of *TR*. Assume *HO* is low; *TR* clocks flop A. As its Q output settles, *TR* is delayed and then clocks flop B. If the delay is long enough, flop A output becomes valid and sets up the D input of flop B for its setup time. Then when the delayed *TR* clocks B, the output edge time is determinate. Most of the time, *HO* does not violate flop A D input setup time, but when it does, flop B is required to synchronize its release with the trigger event. This



trigger generator is adequate for oscilloscopes of up to about 50 MHz bandwidth.

For higher performance, the faster trigger generator shown above is implemented as an IC with ECL logic. The comparators are ECL gates. The D flops of the previous scheme are replaced with faster RS flops or latches: a gate latch (corresponding to flop B) and an arm latch (corresponding to flop A). The trigger waveform drives both trigger and arm comparators. The trigger event out of the trigger comparator sets the gate latch high (*GATE* is high) if arm latch output *A* was low. If so, the sweep runs, and its end is detected, causing *HO* to assert. *HO* forces reset of the arm latch; *A* is forced high. This forces *GATE* low and resets the gate latch. When *HO* goes low, before *GATE* can assert again, *A* must be asserted low by setting the arm latch from the arm comparator. Then the gate latch is "armed" and can be set by the trigger comparator.



The detailed sequence of events is shown above. *HO* has become low. When the input waveform goes below the arm-comparator threshold V_A, the arm latch is set (*A* is low) through gate 3. The range of time when arming can occur is shown as the negative half-cycle of the input waveform. It crosses the trigger-comparator threshold V_T, is delayed, and attempts to set the gate latch by asserting *T* high. If *A* has been low long enough, *GATE* goes high without jitter.

The hysteresis is the difference in comparator thresholds. Because the waveform has a finite slope, a time delay is generated between the latest possible arming *A* and earliest trigger *T*. The additional gate delay in the trigger path lets *A* settle to a valid level at the gate latch input.

The relative time delay of the trigger and arm paths is critical to proper synchronization. Otherwise, two trigger anomalies can occur: trigger jitter and double triggering.

Consider first the case where the trigger-path delay is *too short* relative to the arm path. A low logic level at *A* must be established through the arm path preceding trigger-path assertion at *T* (of a high level). With insufficient delay in the trigger path, *T* could assert before *A* is settled. This occurs when the input, on its positive slope, is crossing V_A just as *HO* releases at the latest possible arming time. Gate 3 of the arm latch will be driven by a quick pulse, barely enough to cause the arm latch to change state after some time. This leaves *A*

indeterminate. When the trigger-path input to gate 2 goes low, *A* causes an uncertain starting edge at *GATE*. Trigger jitter is the result.

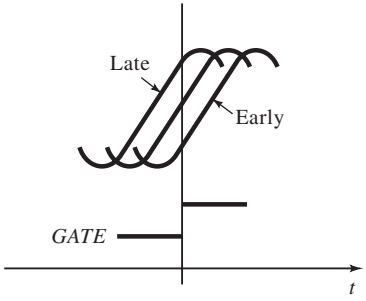
Now consider the case in which the trigger-path delay is *too long* relative to the arm path. When the negative slope of the input signal decreases through V_T, the trigger path propagates a low level at *T*. Meanwhile, the arm path also propagates, through the arm latch, a low level at *A*. This arming of the gate latch must be preceded by a low level at *T*. If the trigger-path delay is excessive, the release of the gate latch at *A* occurs while *T* is still high, causing *GATE* to go high. This produces an extra *GATE* or double trigger. Under correct operation, only the positive slope asserts *GATE*. Double triggering occurs when *HO* releases just as the input signal crosses the arm comparator threshold at the earliest possible arming. On the screen, the waveform appears to be triggering on both slopes.

The optimal delay between the paths is somewhere between these two extremes. Delay-time tolerance is provided by the comparator threshold hysteresis and finite slope (or slew rate) of the input waveform. Because the trigger path to gate 2 is one gate longer than the arm path, the hysteresis delay on the negative slope of the waveform must be at least one gate delay, t_{pd}, to avoid double triggering. An approximate maximum t_{pd} can be calculated by assuming a maximum amplitude sinusoidal input of frequency f_{max}. The gate propagation delay must be

$$t_{pd} < \frac{\text{hysteresis}}{\text{slew rate}} = \frac{V_T - V_A}{2\pi \cdot V_{fs} \cdot f_{max}}$$

A hysteresis adjustment of the comparator threshold on the trigger comparator allows the trigger generator to be adjusted for maximum-frequency fault-free triggering.

Trigger jitter can be observed on an oscilloscope screen. Spurious traces to the left of the main trace are caused by late triggering. A waveform that is advanced in phase (hence, late) appears shifted to the left. This may seem counterintuitive since later time is to the right on the screen. But it is *GATE* that is late relative to the waveform. If the sweep gate had started on time, the waveform would not have advanced as far in phase.



Slew-rate limiting of the input to the trigger system causes it to shift in time on the screen as the trigger level is adjusted. The limited waveform is time distorted, and its phase error varies with amplitude relative to the vertical signal. A time-domain test of trigger-generator performance is to let the input be a pulse of varying width. The minimum width that achieves a stable trigger is an index of generator speed capability.

In the trigger-generator circuit diagram, the function of gate 5 has not yet been described. The *auto trigger* mode causes the sweep to run at a low rate (typically <50 Hz) to display a trace on the screen when no signal is present (or when the trigger controls are not adjusted properly). An auto-mode retriggerable MMV is driven by *GATE*. If the sweep has not run for a while, the auto-MMV times out and gates the sweep on directly. When a triggered gate occurs, the MMV is reset and the free-run mode turned off. Now, if *HO* were the input to gate 4, at high sweep speeds, it would be difficult to get out of the free-run mode without gate 5.

At high sweep rates, *HO* has a large duty-ratio. That is, it takes much longer to retrace and recover from a sweep than the sweep time. If a slow trigger waveform is applied with a period much greater than the sweep time (when *HO* is unasserted), the probability is low that during the sweep the signal would cross the hysteresis window. To escape the free-run mode, a *GATE* pulse is needed to reset the auto-MMV. The auto-defeat gate (gate 5) is added to provide the pulse.

When *HO* unasserts during free run of the sweep, if the trigger signal is below V_A, the arm-latch output *A* goes low. Since the sweep was run by the auto-gate waveform, *GATE* is low, */GATE* is high, and gate 5 thus blocks *HO* from resetting the arm latch. This allows the input time to cross V_T and set the gate latch on

its next half-cycle. If it does while *HO* is low, a triggered gate is asserted and the auto-MMV is reset. More likely, if *HO* is high, gate 5 allows /HO to reset the arm latch, and *A* goes high, resetting the gate latch low. *GATE* is high for the propagation time around the loop of gates 2-5-4-3-2. This is a few nanoseconds, enough time to reset the auto-MMV.

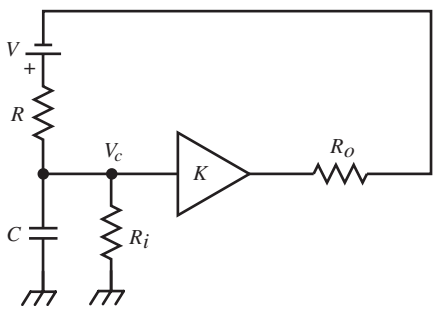
Another approach to auto triggering, *autolevel*, is to generate a triangle wave with a counter and DAC and sum it with the trigger-level control output. The triangle wave scans the level through the input range. When it intersects the input waveform, a trigger event is generated, and triggering then occurs. If no input is present, it autogenerates a trigger when the set trigger level is crossed.

An alternative approach to automatic triggering, *peak-to-peak auto*, uses positive and negative peak detectors to generate voltages at the extrema of the input signal. The trigger-level potentiometer is then placed between these peak voltages so that its control range is always within the signal range.

Ramp and Sweep Generators

Oscilloscope time-base systems consist of a trigger generator followed by a sweep generator. The sweep generator is a gated ramp or sawtooth generator that drives the horizontal deflection amplifier. Ramp generators are also used in magnetically deflected CRT display systems to generate deflection-coil currents, in mass spectrometers, and in pulse-width modulators.

A *bootstrap ramp generator* uses the bootstrapping technique to maintain a constant voltage across a timing resistor *R*. As *C* charges, the top end of *R* follows it, driven by a buffer. The result is a linear ramp output.



With floating voltage source V, the ramp slope is

$$\frac{dv}{dt} = \frac{V/R}{C} = \frac{V}{R \cdot C}$$

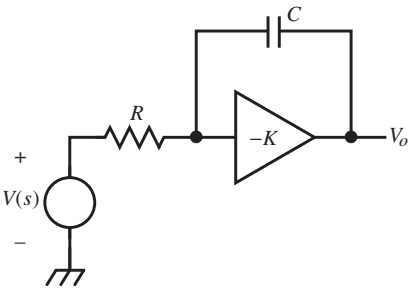
This perfect scheme is spoiled by shunt resistance R_i (of the capacitor and buffer input), buffer output resistance, R_o, and buffer gain deviation from unity. Let V be gated on at $t = 0$ as

$$v(t) = V \cdot u(t) \Rightarrow V(s) = V/s$$

Then solving the circuit in s,

$$\frac{V_c(s)}{V(s)} = \frac{R_i}{R + R_o + (1-K) \cdot R_i} \cdot \frac{1}{s\left[\left(\frac{R + R_o}{1-K}\right) \| R_i\right] \cdot C + 1}$$

where K is a buffer voltage gain near one. The bootstrap effect appears as the increase in effective resistance of $R + R_o$ by $1/(1 - K)$ times. With very large R_i, this increases the time constant by the same factor. In effect, the ramp is generated as the initial segment of a long exponential curve. K, R_o and R_i all contribute to the time-constant deviation from RC.



Another approach to ramp generation is to use an op-amp integrator. This *Miller ramp generator* has a transfer function of

$$\frac{V_o(s)}{V(s)} = -K \cdot \frac{1}{s(1+K) \cdot R \cdot C + 1}$$

For an op-amp, $K \to \infty$, and the transfer function approaches $-1/sRC$, an ideal integrator. With finite gain, the output is an exponential with time constant multiplied by $(1 + K)$; the early part of the curve is approximately linear. The fractional deviation, ε, from linear response is

$$\varepsilon \cong \frac{t}{2 \cdot \tau} = \frac{t}{2 \cdot R \cdot C \cdot (1+K)}$$

where τ is the effective time constant. The fractional nonlinearity was derived by series-expanding the response exponential to the quadratic term, subtracting the (ideal) linear and constant terms, and dividing by the linear term. Error grows with time as the exponential becomes increasingly sublinear. This error formula applies to both bootstrap and Miller ramp generators; the responses of both are exponentials.

Finite gain also causes a slope error in the Miller integrator. In the transfer function, let $V(s)$ be that of V/s. Then the transfer function corresponds to the normalized time-domain response of

$$\frac{dv_o(t)/V}{dt}$$

Applying the initial value theorem,

$$\left.\frac{dv_0(t)/V}{dt}\right|_{t=0} = \lim_{s \to \infty} s\left(-K \cdot \frac{1/s}{s(1+K) \cdot R \cdot C + 1}\right) = -\left(\frac{K}{1+K}\right) \cdot \frac{1}{RC}$$

The initial slope of the ideal ramp response, $-1/sRC$, is $-1/RC$. Thus,

$$\text{fractional slope error} = \frac{K}{1+K}$$

Fast Miller ramp generators are driven by an input current source instead of a voltage source and R. This produces a more linear response and reduces the effect of input impedance. The ideal response of $-1/sC$ is only approximate. More precisely,

$$\frac{V_o(s)}{I_i(s)} = \frac{1}{sC \cdot \left(\dfrac{1+K(s)}{K(s)} \right)} = \left(\frac{K(s)}{1+K(s)} \right) \cdot \frac{1}{sC}$$

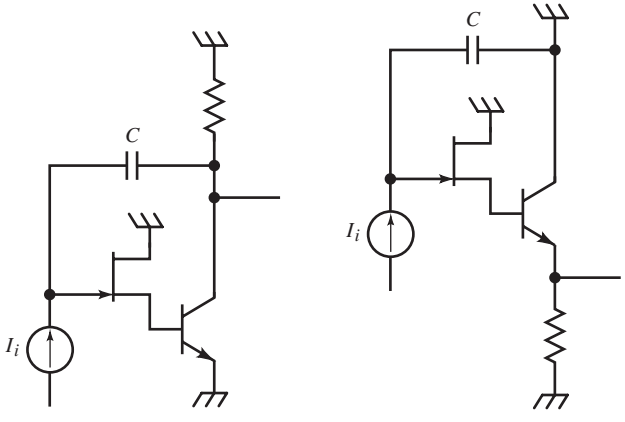
The ideal response is multiplied by the s-domain equivalent of the fractional slope error. To approach the ideal, the op-amp must maintain high gain at high frequencies. This is a major limitation, especially since the op-amp output impedance gyrates inductively. For

$$K(s) = \frac{K_o}{s\tau_{bw} + 1}$$

$$\frac{K(s)}{1+K(s)} = \left(\frac{K_o}{1+K_o} \right) \cdot \frac{1}{s\left[\tau_{bw}/(1+K_o)\right]+1} \cong \left(\frac{K_o}{1+K_o} \right) \cdot \frac{1}{s\tau_T + 1}$$

The op-amp adds an additional pole at its unity-gain frequency f_T.

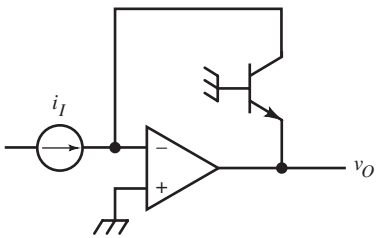
An improved ramp generator is based on the simplicity of a gated current source charging a capacitor, followed by a buffer amplifier. Feedback loops are avoided, and the step response is faster than for the previous two schemes.



(a) (b)

Bruce Hofer recognized the topological equivalence of this current-source ramp generator (b) with the Miller generator (a). The difference is in where ground is placed. In the Miller, the BJT is a CE; in the current-source, it is a CC. In the Miller, any anomalous switching voltage at the FET gate is coupled through C to the output; in the current-source generator, it is bypassed by C to ground.

Logarithmic and Exponential Amplifiers



Logarithmic amplifiers are useful for compressing a wide dynamic-range waveform, for multiplying, and for function generation. This simple log-amp is based on the BJT *b-e* (diode) junction *v-i* relationship:

$$i = I_S \cdot e^{v/V_T}, \quad i >> I_S$$

The input current i_I is the BJT collector current and the output voltage,

$$v_O = -v_{BE} = -V_T \cdot \ln\left(\frac{i_C}{I_S}\right) = -V_T \cdot \ln\left(\frac{i_I}{I_S}\right)$$

This log-amp reflects the temperature sensitivity of the *b-e* junction and drifts due to both V_T and I_S.

Detailed expressions for i_C and i_E (Gibbons and Horn 1964) indicate that several error terms vanish when v_{CB} is zero. The logarithmic relation extends to lower currents for i_C (because $v_{CB} = 0$ and $v_{BE} \neq 0$) and is accurate over about

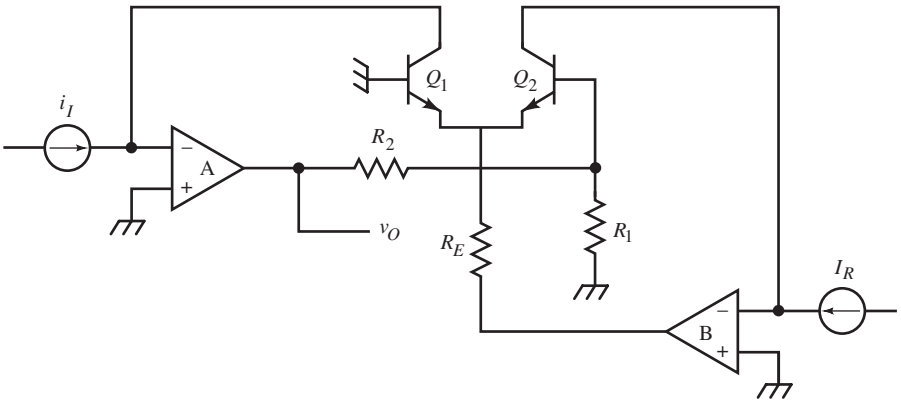
nine decades. The high end (typically 10 mA) is limited by ohmic voltage drops in series with v_{BE} and at the low end (typically 100 pA) by I_S. From v_O,

$$v_{BE} = V_T \cdot \ln i_C + V_T \cdot \ln I_S$$

The slope of v_{BE} versus ln i_C is linear with a slope of

$$\frac{dv_{BE}}{d(\log i_C)} = \frac{dv_{BE}}{d(\ln i_C)} \cdot \frac{d \ln i_C}{d \log i_C} = V_T \cdot \frac{1}{\log e} = 59.56 \frac{\text{mV}}{\text{dec}}$$

For each decade of change in i_C, v_{BE} changes by about 60 mV at 300 K over typically nine decades of i_C.



The effect of I_S can be eliminated with a matched pair of BJTs, as in the *log-ratio amplifier* above. Assuming equal I_S,

$$\Delta v_{BE} = v_{B2} = v_{BE2} - v_{BE1} = V_T \cdot \ln\left(\frac{I_R}{i_I}\right)$$

The output is v_{B2} scaled by the divider, or

$$v_O = \left(\frac{R_2}{R_1} + 1\right) \cdot v_{B2} = -\left(\frac{R_2}{R_1} + 1\right) \cdot V_T \cdot \ln\left(\frac{i_I}{i_R}\right)$$

The second op-amp B functions like A in keeping $i_{C2} = I_R$, a reference current. For $i_I < I_R$, $v_O > 0$. Op-amp B controls the emitter currents $i_I + I_R$, whereas op-amp A controls v_{B2} to keep $i_{C1} = i_I$. R_E is selected so that the output voltage range of B can span the range of $i_I + I_R$.

This circuit eliminates dependence on I_S but is dependent on V_T. The 0.33%/°C TC of v_{B2} is sometimes compensated by making R_1 a positive TC (PTC) thermistor. Its TC is found by differentiating v_O with respect to T, setting it to zero, and solving for dR_1/dT. The fractional TC equation is

$$\frac{1}{v_O}\cdot\frac{dv_O}{dT} = \mathrm{TC}\%(v_O) = \frac{1}{T} - \frac{1}{R_1}\cdot\frac{dR_1}{dT}\cdot\left(\frac{R_2}{R_1+R_2}\right) = \frac{1}{T} - \mathrm{TC}\%(R_1)\cdot\left(\frac{R_2}{R_1+R_2}\right)$$

When $\mathrm{TC}\%(v_O) = 0$,

$$\mathrm{TC}\%(R_1) = \left(\frac{1}{T}\right)\cdot\left(\frac{R_1+R_2}{R_2}\right) \cong (0.33\%/°\mathrm{C})\cdot\left(\frac{R_1+R_2}{R_2}\right) \text{ at } 300 \text{ K}$$

Example: Log-Amp Design

A log-amp based on the log-ratio circuit requires an input range of 0.1 μA to 1 mA with minimal temperature drift. The output must be 0 V at an input of 1 mA (fs) and 4 V at 0.1 μA (zs).

The scaling is 4 V per four decades or 1 V/dec of i_I. The divider must be

$$\left(\frac{R_2}{R_1}+1\right) = \frac{1\ \mathrm{V/dec}}{60\ \mathrm{mV/dec}} = 16.8$$

PTC thermistors with 1 kΩ values are available, so let $R_1 = 1$ kΩ. Then from the divider ratio, $R_2 = 15.8$ kΩ. The TC of R_1 must be

$$\mathrm{TC}\%(R_1) = (0.33\%/°\mathrm{C})\cdot\left(\frac{16.8}{15.8}\right) = (0.35\%/°\mathrm{C})$$

For $v_O = 0$ V at 1 mA input, $I_R = 1$ mA. This can be supplied from a voltage reference of 5 V through a 4.99 kΩ, 1% resistor. When i_I is supplied from v_I through a 100 kΩ, 1% resistor, the input range is 10 mV to 100 V. Finally, for a minimum op-amp B output voltage of about −4.7 V, $R_E = 4\text{ V}/2\text{ mA} = 2\text{ k}\Omega$.

Op-amp input capacitance C_i can destabilize the log-amp. A feedback capacitor C_f forms a pole with collector resistance r_c. The loop gain is

$$GH = \frac{K}{s\tau_{bw} + 1} \cdot \frac{r_c}{r_M} \cdot \frac{sr_M \cdot C_f + 1}{sr_c \cdot (C_f + C_i) + 1}$$

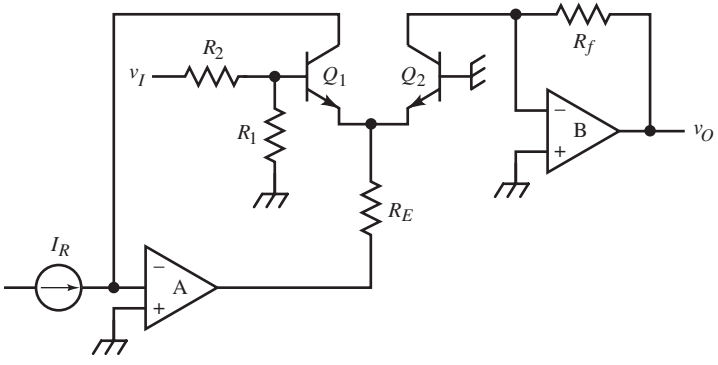
where K and τ_{bw} are op-amp open-loop gain and bandwidth, and r_c is collector resistance:

$$r_c = (\beta + 1) \cdot r_o$$

Also,

$$r_M = \frac{r_e + r_e'}{\alpha} = r_m + \frac{r_e'}{\alpha}$$

The zero depends on r_M, which varies with i_C. For stability, the poles must be separated and the zero placed for lead compensation near a loop gain of one.



The inverse function, exponentiation, is achieved by modifying the log-amp. In this exp-amp (or antilog-amp), op-amp A maintains a constant $i_{C1} = I_R$. Then

$$v_{B1} = \left(\frac{R_1}{R_1 + R_2}\right) \cdot v_I = v_{BE1} - v_{BE2} = V_T \cdot \ln\left(\frac{I_R}{i_{C2}}\right)$$

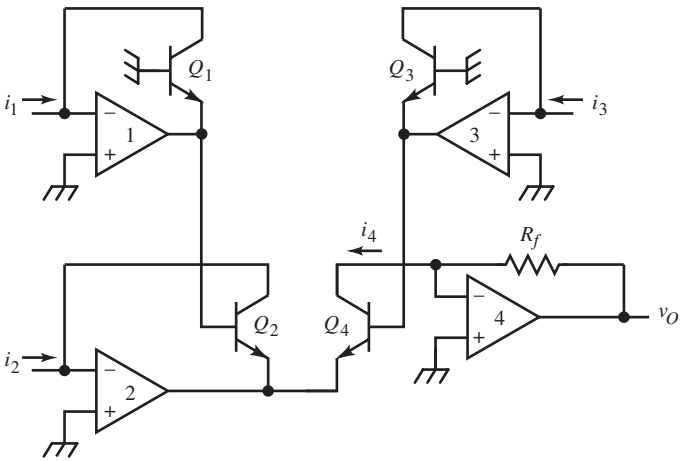
and at the output,

$$v_O = R_f \cdot i_{C2}$$

Solving for i_{C2} in v_{B1} and substituting yields

$$v_O = R_f \cdot I_R \cdot \exp\left[\left(\frac{R_1}{R_1 + R_2}\right) \cdot \left(\frac{v_I}{V_T}\right)\right]$$

Temperature compensation is also required for V_T.



Log-amps, exp-amps, and amplifiers combine to form function-generating circuits. Four op-amps and matched BJTs combine in the circuit diagram above to form a multiplier or divider based on the relation

$$xy = \log^{-1}(\log x + \log y) = \exp(\log x + \log y)$$

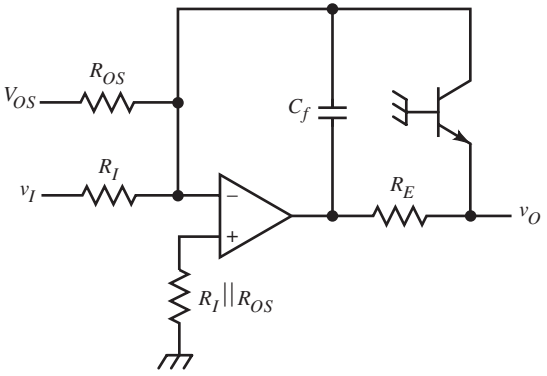
The BJTs are connected with *b-e* junctions in series so that

$$V_T \cdot \ln\left(\frac{i_1}{I_S}\right) + V_T \cdot \ln\left(\frac{i_2}{I_S}\right) = V_T \cdot \ln\left(\frac{i_3}{I_S}\right) + V_T \cdot \ln\left(\frac{i_4}{I_S}\right)$$

or

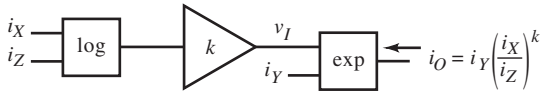
$$i_4 = \frac{i_1 \cdot i_2}{i_3}$$

In practice, the basic log-amp is stabilized, and voltage input and offset (for bipolar operation) are provided, as shown below.



V_T can be temperature compensated with a thermistor, but a more exacting approach uses two log-ratio amps and a divider. The first log-amp has inputs i_I and I_R, and the second is a temperature compensator; it has inputs of I_R and $k \cdot I_R$, where k is a scale factor. The log output is the quotient of the two log-ratio amp outputs:

$$\frac{V_T \cdot \ln(i_I / I_R)}{V_T \cdot \ln(kI_R / I_R)} = \frac{\ln(i_I / I_R)}{\ln k}$$



A log-ratio amplifier with output,

$$v_O = V_T \cdot \ln\left(\frac{i_X}{i_Z}\right)$$

and an exp-amp with output,

$$i_O = i_Y \cdot e^{(V_I/V_T)} = i_Y \cdot \exp\left(\frac{v_I}{V_T}\right)$$

are combined with an amplifier of gain k in the figure above. The output is

$$i_O = i_Y \cdot \left(\frac{i_X}{i_Z}\right)^k$$

For $k < 1$, the amplifier is replaced by a voltage divider. This function is quite versatile; scaled powers of ratios include squares, cubes, square roots, and truncated power series of transcendental functions. For example,

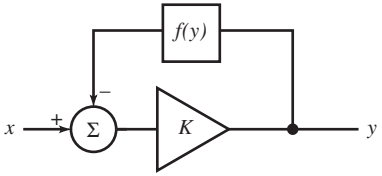
$$\sin x \cong x - \frac{x^3}{6.79}, \quad 0 \le x \le \frac{\pi}{2} \quad \pm 1.35\% \text{ error}$$

With fractional powers, a better approximation is (see Sheingold 1974)

$$\sin x \cong x - \frac{x^{2.827}}{6.28}, \quad 0 \le x \le \frac{\pi}{2} \quad \pm 0.25\% \text{ error}$$

Also,

$$\tan^{-1} x \cong \frac{\pi}{2} \cdot \frac{x^{1.2125}}{1 + x^{1.2125}}$$



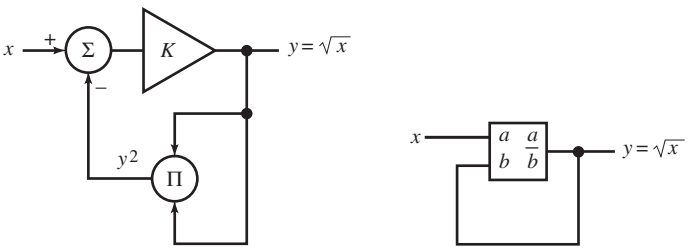
The inverse function $f^{-1}(x)$, of a function circuit $f(y)$, can be realized by placing $f(y)$ in the feedback path of an op-amp. The output,

$$y = K \cdot (x - f(y))$$

or

$$x = \frac{y}{K} + f(y)$$

For infinite K, $x = f(y)$, or $y = f^{-1}(x)$. This is useful for generating functions implicitly.



In (a), a multiplier (indicated by the Π symbol) is in the feedback path of a noninverting op-amp. The circuit outputs the square root of the input:

$$y = K \cdot (x - y^2) \Rightarrow \frac{y}{K} = x - y^2$$

For infinite K, $x = y^2$ or $y = \sqrt{x}$. Implicit function generation, as in (b), is often more accurate for the same complexity of function blocks because the range over which intermediate variables must maintain accuracy is reduced.

This is especially true of the commonly used root sum of squares (rss) function, the *vector magnitude* function or Pythagorean formula:

$$y = \sqrt{\sum_{k=1}^{n} x_k^2}$$

The range of x_k^2 can be large but need not be if computed implicitly. An implicit formula is derived from y. Squaring each side, adding $y \cdot x_n$, factoring, and dividing,

$$y = \frac{\sum_1^n x_k^2 + yx_n}{y + x_n} = \frac{\sum_1^{n-1} x_k^2 + x_n(y + x_n)}{y + x_n} = \frac{\sum_1^{n-1} x_k^2}{y + x_n} + x_n$$

In this formulation of y, it is fed back to divide x_k^2, thus reducing the required input dynamic range.

A function describing the output of a bridge circuit with a sensor in one branch is

$$y = \frac{x}{1 + x} \Rightarrow y = x \cdot (1 - y)$$

More generally,

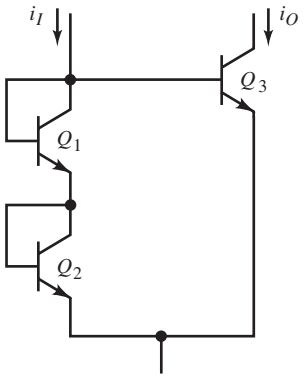
$$y = \frac{A \cdot x^k}{1 + A \cdot x^k} = (A - y) \cdot x^k$$

Finally, a two-term power series with $y(0) = 0$ is

$$y = A \cdot x + x \cdot (B \cdot x + C \cdot y) = \frac{A \cdot x + B \cdot x^2}{1 - C \cdot x}$$

FUNCTION GENERATION

Function generation by log-amp circuits is based on the logarithmic nature of BJT junctions. The translinear cell is also a basis for function generation. The basic cell can be generalized by placing an arbitrary number of junctions in series on either the input or output side of the cell. For example, the current squarer shown below has two junctions in series on the input side.

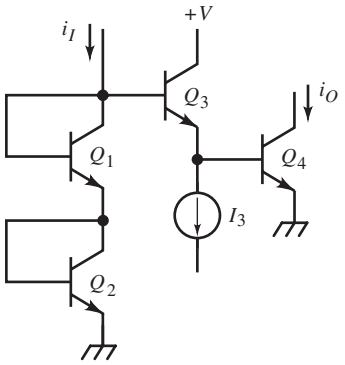


As in translinear-cell analysis, assume that β is infinite. The output current is derived from the circuit equations

$$v_{BE3} = 2 \cdot V_T \cdot \ln\left(\frac{i_I}{I_S}\right)$$

$$i_O = I_S \cdot \exp\left(\frac{v_{BE3}}{V_T}\right) = \frac{i_I^2}{I_S}$$

More generally, m series input junctions results in i_I^m output. The translinear cell is not dependent on I_S as this circuit is.



In the above circuit, the improved current squarer adds Q_3, biased at a constant current I_3. Its additional junction drop provides the needed correction for I_S:

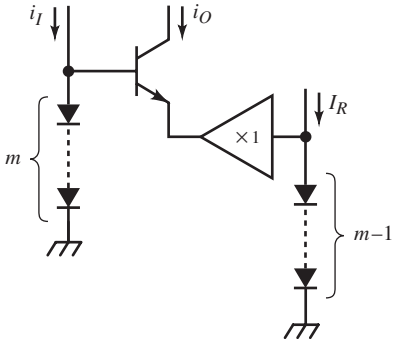
$$v_{B4} = V_T \cdot \ln\left(\frac{i_O}{I_S}\right) = v_{B3} - v_{BE3} = 2 \cdot V_T \cdot \ln\left(\frac{i_I}{I_S}\right) - V_T \cdot \ln\left(\frac{I_3}{I_S}\right)$$

or

$$i_O = \frac{i_I^2}{I_3}$$

The squarer circuit concept can be extended to m input diodes and $m-1$ stages like Q_3 with currents I_1 to I_{m-1}. The output current is then

$$i_O = \frac{i_I^m}{I_1 \cdot I_2 \cdots I_{m-1}}$$

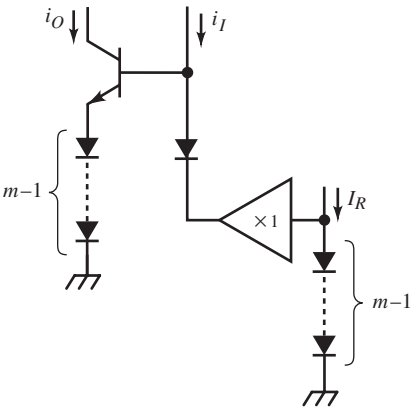


A more general scheme is shown above. Here, two stacks of diode junctions drive the BJT. The stack biased by I_R is buffered to keep BJT current separate from I_R. Applying KVL to the loop and eliminating $V_T \cdot \ln$,

$$\frac{i_O}{I_S} = \left(\frac{i_I}{I_S}\right)^m \left(\frac{I_S}{I_R}\right)^{m-1} \Rightarrow \frac{i_O}{I_R} = \left(\frac{i_I}{I_R}\right)^m$$

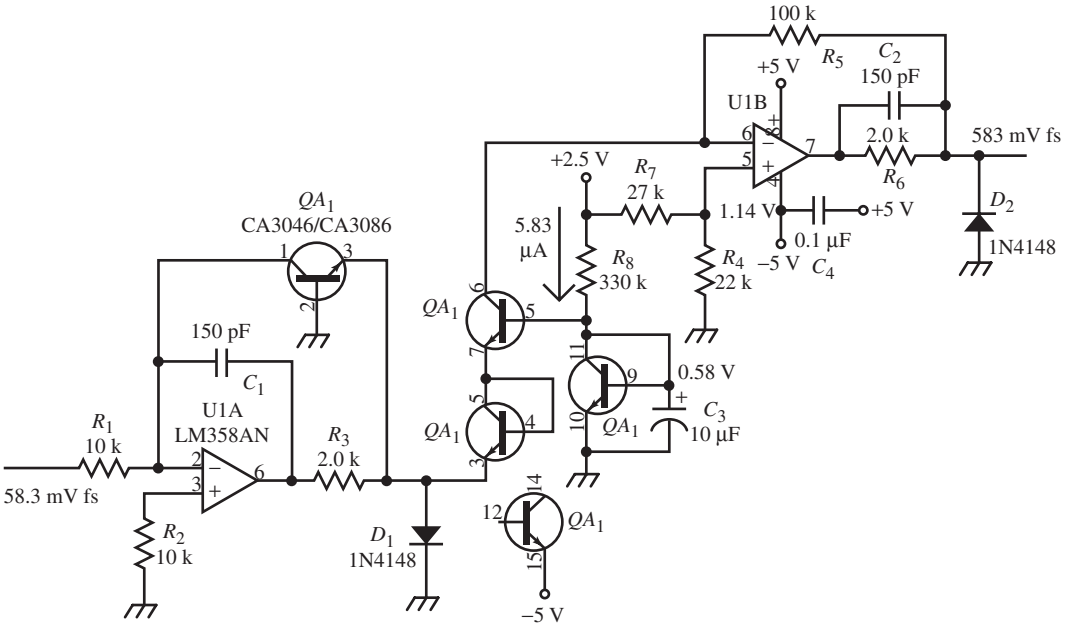
Fractional powers result by exchanging i_I and i_O, as shown below, with appropriate modifications. The BJT is moved to the former input string, and its first diode is now at the new input. Applying KVL and discarding $V_T \cdot \ln$,

$$\left(\frac{i_O}{I_S}\right)^m = \left(\frac{i_I}{I_S}\right) \cdot \left(\frac{I_R}{I_S}\right)^{m-1} \Rightarrow \frac{i_O}{I_R} = \left(\frac{i_I}{I_R}\right)^{1/m}$$



Example: Square-Root Circuit

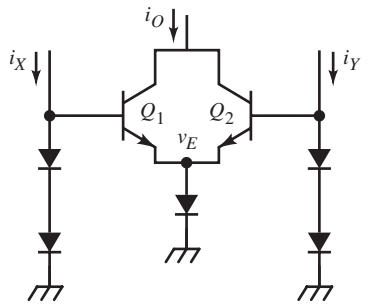
An example of a square-root circuit used in a medical laser is shown below to compensate for the quadratic (square-law) behavior of the flashlamp. A transistor array is used to provide matched BJTs.



The BJT with base pin 9 corresponds in function to Q_3 of the current-squarer. The circuit function is

$$\frac{v_o^2}{v_i} = \left(\frac{R_f}{R_i}\right) \cdot R_f \cdot I_R = 5.832 \text{ V} \Rightarrow \frac{(v_o/V_R)^2}{(v_i/V_R)} = \frac{R_f}{R_i}, V_R = R_f \cdot I_R$$

where R_f = R5, R_i = R1, and I_R = 5.83 µA is set by R8. The circuit has a ×10 gain at full scale. The transistor array has $I_S = 10^{-15}$ A.



A rss circuit, shown above, based on a generalized translinear cell can be extended to three or more inputs. The voltage at the emitter, across the common diode, is

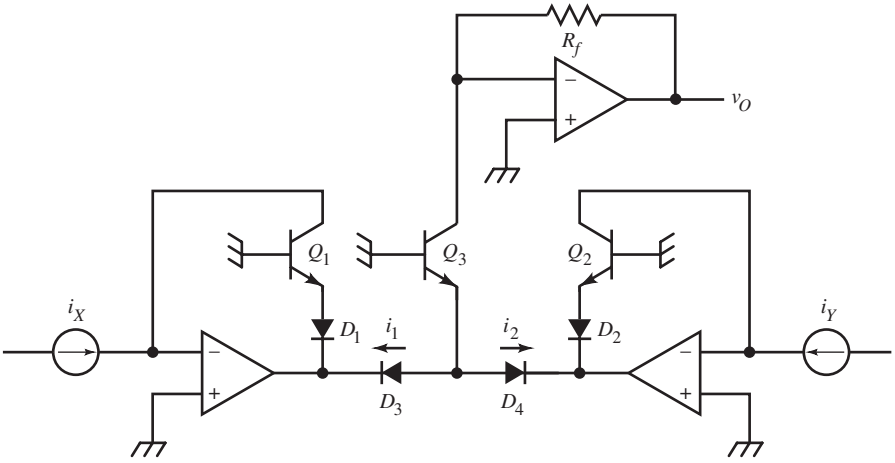
$$v_E = V_T \cdot \ln\left(\frac{i_1 + i_2}{I_S}\right)$$

Applying KVL and removing $V_T \cdot \ln$ results in

$$\frac{i_1}{I_S} + \frac{i_2}{I_S} = \left(\frac{i_X}{I_S}\right)^2 \cdot \left(\frac{I_S}{i_1 + i_2}\right) + \left(\frac{i_Y}{I_S}\right)^2 \cdot \left(\frac{I_S}{i_1 + i_2}\right) = \frac{i_X^2 + i_Y^2}{i_1 + i_2} \cdot \frac{1}{I_S}$$

The output current is

$$i_O = i_1 + i_2 = \sqrt{i_X^2 + i_Y^2}$$



A log-antilog scheme for an rss circuit (above) has common output current,

$$i_O = i_1 + i_2$$

and two loops to which KVL is applied:

$$\left(\frac{i_X}{I_S}\right)^2 = \left(\frac{i_1}{I_S}\right)\cdot\left(\frac{i_O}{I_S}\right) \quad \text{and} \quad \left(\frac{i_Y}{I_S}\right)^2 = \left(\frac{i_2}{I_S}\right)\cdot\left(\frac{i_O}{I_S}\right)$$

These equations reduce to

$$i_X^2 = i_1 \cdot i_O \quad \text{and} \quad i_Y^2 = i_2 \cdot i_O$$

Adding them results in

$$i_O \cdot (i_1 + i_2) = i_O^2 = i_X^2 + i_Y^2 \Rightarrow i_O = \sqrt{i_X^2 + i_Y^2}$$

An rms circuit similar to these rss circuits consists of three cascaded blocks: a squarer, averager (integrator), and square-root block. The rms circuit requires the intermediate averaging function.

The Analog Devices AD534 cleverly implements the rms function implicitly, using a current-squarer circuit and making use of its inherent divider capability. By feeding back the output as a divisor, the circuit implements the rms function,

$$v_o = \sqrt{avg\left(v_i^2\right)}$$

or

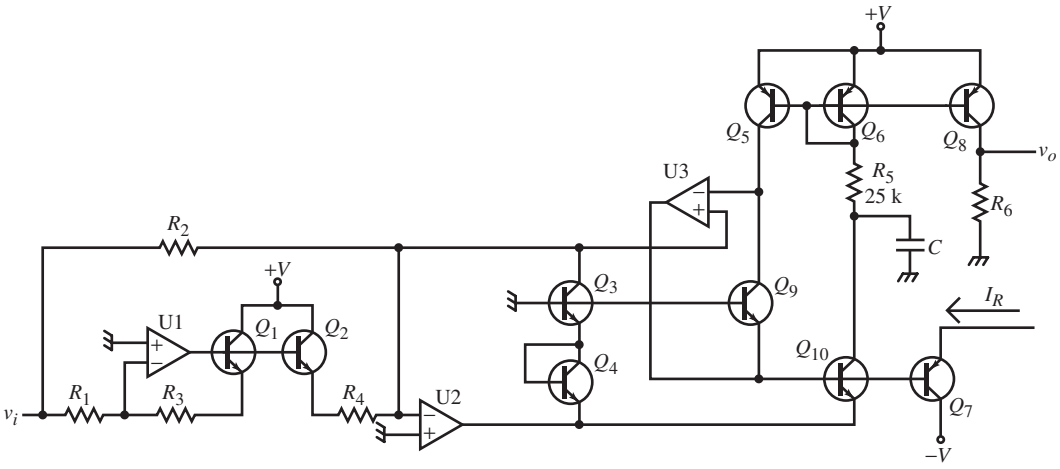
$$v_o^2 = \left|avg\left(v_i^2\right)\right|$$

by implementing the equivalent expression,

$$v_o = \frac{avg\left(\left|v_i\right|^2\right)}{v_o}$$

The scheme is shown below. The first stage is an absolute-value circuit that drives the squarer-divider with i_{C3}. The Q_3 collector voltage is $|v_i|/R_2$. An ampli-

fier, U3, is used to drive the emitter of Q_9 with the voltage required for it to conduct i_{C5}.



Writing the loop equation for the squarer,

$$2 \cdot V_T \cdot \ln\left(\frac{i_{C3}}{I_S}\right) = V_T \cdot \ln\left(\frac{i_{C9}}{I_S}\right) + V_T \cdot \ln\left(\frac{i_{C10}}{I_S}\right)$$

This reduces quickly to

$$i_{C10} = \frac{i_{C3}^2}{i_{C9}}$$

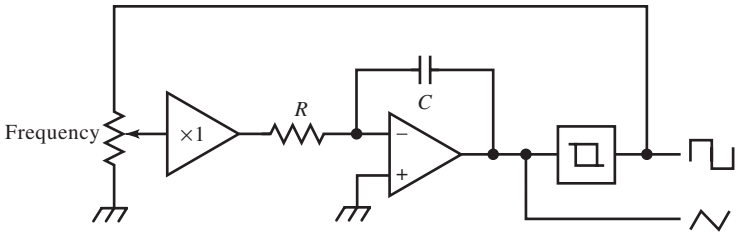
The current mirror (Q_5, Q_6, and Q_8) replicates the squarer output current, i_{C10}, as rms output current, i_{C8}, and feeds a copy of it back to the divider as i_{C9}. The squarer output current is averaged by R_5, C. Q_7 is driven with a reference current, I_R. Then its emitter voltage is related to i_{C9}, the output current, by

$$v_{e7} = V_T \cdot \ln(i_o / I_R)$$

With proper scaling of v_i and I_R, v_{e7} is a log-rms (or dB-rms) output.

TRIANGLE-WAVE GENERATORS

Function generators are a low-cost and versatile signal source. Their oscillator is a kind of multivibrator combined with a dual-slope ramp generator. This *triangle-wave generator* (TWG) outputs both triangle and square waves. TWGs are voltage-controlled oscillators (VCOs); their frequency can be accurately controlled over several decades by an input voltage, designated voltage-controlled frequency (VCF) or voltage-controlled generator (VCG) on commercial function generator (FG) instruments. This makes them useful for frequency sweeping or modulation. An external phase-locked loop (PLL) can make them accurate frequency sources as well.



An early approach to triangle-wave generation coupled a Miller ramp generator with a bistable MV or Schmitt trigger. The bipolar square-wave from the switch is integrated by the op-amp, producing a triangle wave. For a symmetric (50% duty-ratio) output, the magnitudes of the square-wave levels must be equal for equal triangle-wave slopes. The frequency can be adjusted, as shown, by varying the amplitude of the square-wave input to the integrator. Like ramp generators using an op-amp integrator, this approach is both speed- and precision-limited. For a VCF range of three decades and a full-scale square-wave amplitude of 10 V, at zero scale (the low-frequency end) the square-wave amplitude is 10 mV. For 1% waveform symmetry, the integrator input offset error must be less than 100 μV.

Speed-wise, for a high-frequency limit of 1 MHz, triangle-wave amplitude of 10 V, and triangle nonlinearity of less than 1%, the slope magnitude is

$$\frac{dv_{TW}}{dt} = \frac{V_{SQ}}{RC} = \frac{V_{TW}}{T/4} = \frac{10\text{ V}}{250\text{ ns}}$$

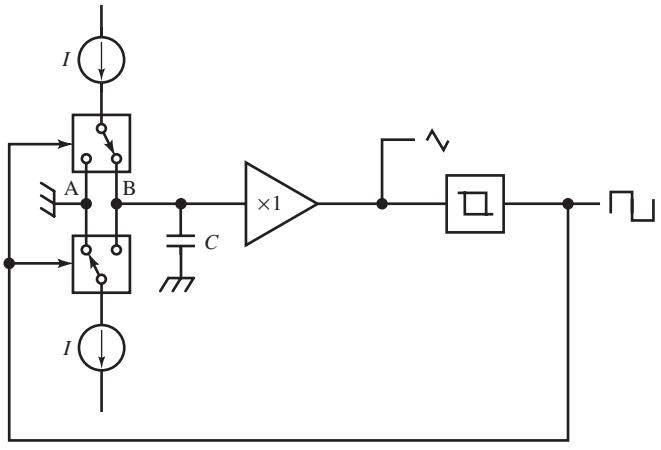
where V_{SQ} and V_{TW} are square-wave and triangle-wave amplitudes and T their period. V_{SQ} is also 10 V, and $RC = 250$ ns. The integrator op-amp gain requirement for the specified nonlinearity is calculated from the ramp fractional nonlinearity, which is

$$\varepsilon \cong \frac{t}{2\cdot\tau} - \frac{t}{2\cdot R\cdot C\cdot(1+K)}$$

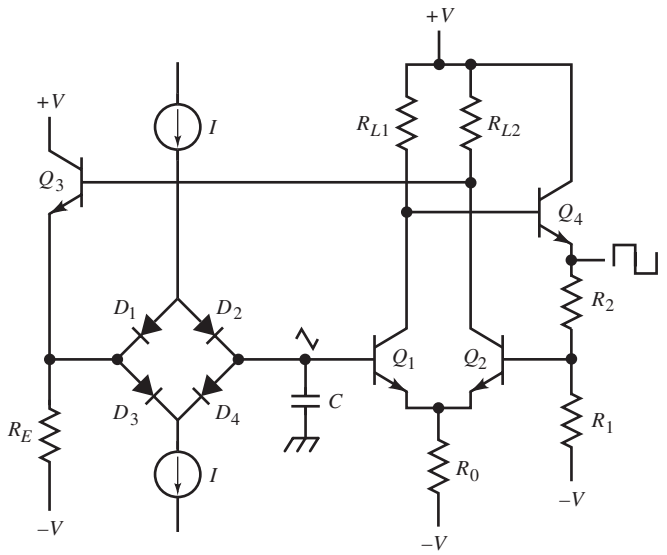
The integrator fractional nonlinearity and required gain are

$$\varepsilon = \frac{T/2}{2\cdot(1+K)\cdot R\cdot C} \Rightarrow K = \frac{T}{\varepsilon\cdot 4\cdot R\cdot C} - 1$$

K must be at least 999 at 1 MHz, an almost 1 GHz unity-gain bandwidth.

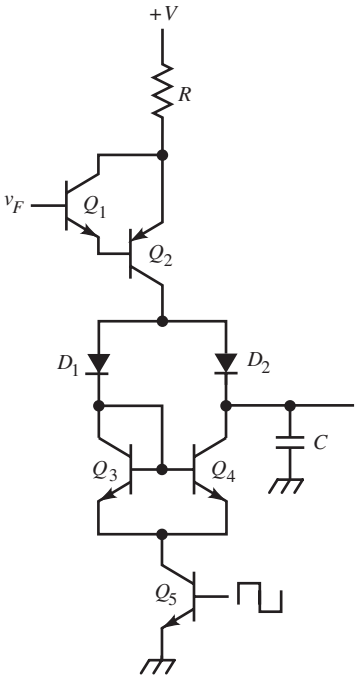


To achieve better performance, current-source ramp generation is used with a bipolar current source. Two matched current sources are used to achieve bipolar charging of *C*. The positive (source) and negative (sink) currents are alternately switched into *C* each half cycle. This scheme is much faster and more precise than the first one but requires matching of current sources for time symmetry and symmetric hysteresis thresholds for voltage symmetry.

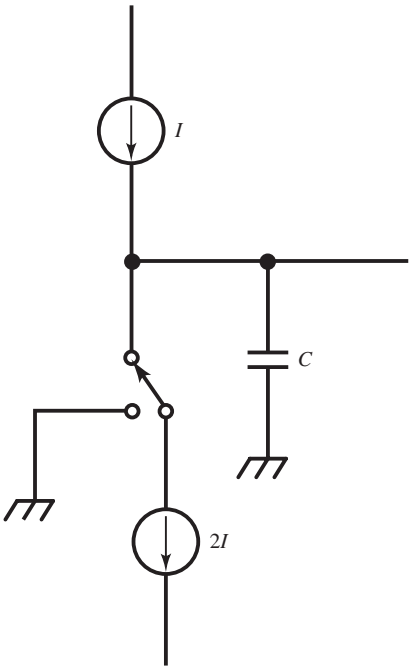


A more concrete circuit realization (above) is a fast TWG, capable of over 100 MHz triangle-wave frequency. The hysteresis-switch and current-switch driver Q_1–Q_4 can be implemented with a single ECL NOR gate. The diode bridge is switched by Q_3 and is limited mainly by diode shunt capacitance. Sometimes a Faraday shield is placed between the left and right halves of the diode bridge to decouple the square-wave node v_{E3} from the triangle-wave node at *C*. The input resistance of the Schmitt trigger is not high, resulting in appreciable

nonlinearity. By adding a good buffer, its additional delay slows the loop. Speed and precision are, as usual, in conflict.

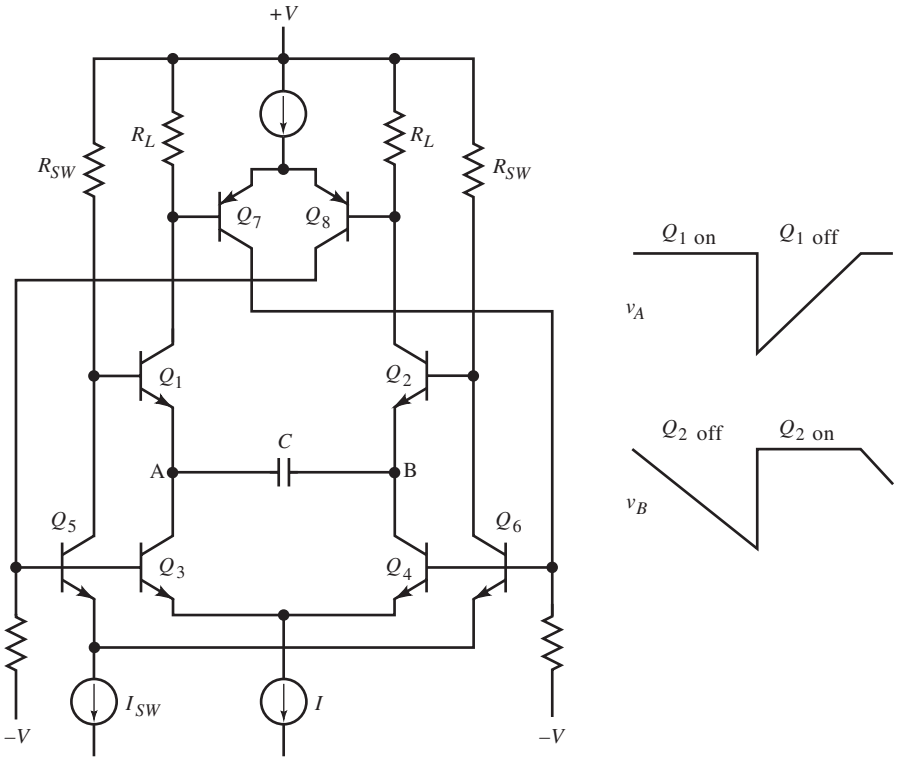


A variation on the bipolar current supply uses only one current source for charging C. This eliminates the matching problem between sources. The Signetics NE565 FG and NE566 PLL both have TWGs of the above design. Q_1 and Q_2 form a current source or V/I converter with input v_F, the VCF input. It generates I through R, the timing resistor. Switching of I is controlled by Q_5. When it is off, D_2 conducts I, and C charges (positive). When Q_5 is on, $v_{C3} < v_{C4}$, and D_1 conducts I through Q_3, where it is replicated in Q_4 as $-I$. Thus, C is discharged with the same magnitude of current as it is charged. Symmetry depends on current-mirror matching.

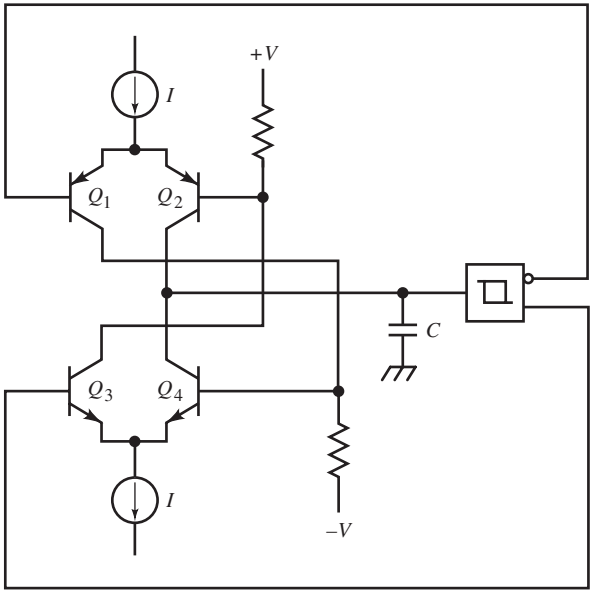


It is not necessary to switch both current sources. To simplify switching, the *I*-2*I* scheme is used. Only one current switch is required, but now the *I* and 2*I* sources must be matched at an accurate ratio of 2.

A more recent approach to better symmetry is to control one current source based on the triangle-wave slope generated by the other. In the differential TWG, instead of connecting triangle-wave node A (page 107) to ground, it charges a second capacitor of the same value from $-I$ while the node B capacitor charges from $+I$. Its triangle wave is inverted relative to the one at node B. The two waveforms are then summed. Ideally, the sum is zero, but any difference is an error voltage that is applied to one of the current sources to correct the slope of the waveform it is generating to match the other. This scheme increases symmetry enough at low currents to extend the VCF range an extra decade.



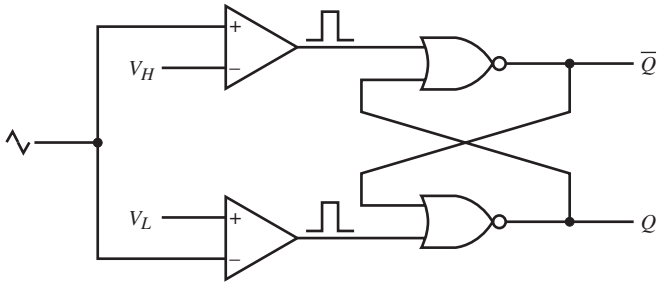
A conceptually similar scheme is used in the Exar XR2206 FG IC, shown in simplified form above. The triangle wave is developed differentially across C and requires a diff-amp pick-off. Q_1–Q_4 are the timing-current switches, but there is only one timing-current source. Q_1, Q_2, Q_7, and Q_8 are the hysteresis switch. The *b-e* junctions of Q_1–Q_2 sense the waveform and switch on when v_E is reduced sufficiently. On one half-cycle, Q_1 and Q_4 conduct; on the other half-cycle, Q_2 and Q_3 conduct. I_{SW} sets the hysteresis thresholds. If the R_{SW} are matched, the levels are symmetric. A disadvantage in this scheme is that the switching voltage is added to the capacitor ramp voltage at nodes A and B, as shown. It is difficult for a diff-amp to common-mode reject these fast switching edges, and some "glitches" appear in the triangle-wave output.



The performance of a TWG loop depends on the subsystems within the loop. The timing capacitor must be of sufficient quality (plastic), and for multiple frequency ranges, a matched set is usually required. The triangle-wave buffer must be fast, have low input-voltage offset and bias current, and high input resistance. The current switches must be fast and have low current leakage when off. Transistors are generally superior to diodes in both leakage and switching characteristics and are used as switches as shown in the TWG loop above. A minimum of two switches is required in a balanced two-source scheme, as shown. The switching scheme here consists of complementary diff-amps. Two BJTs switch the current-switch BJTs.

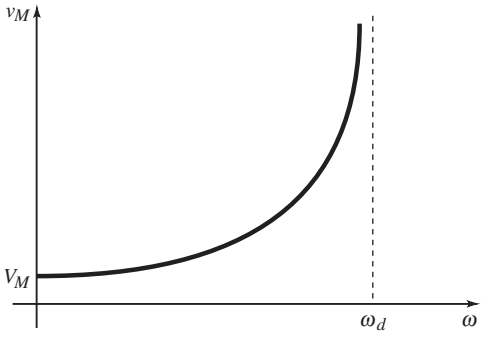
The timing-current generators also must be capable of operating accurately over as many decades of current as the VCF range because output frequency is proportional to timing current. Complementary *V/I* converters are commonly used to supply both polarities of current. These converters must be driven by precision circuitry that establishes symmetric voltages at their inputs. These circuits need only the bandwidth required of the FM VCF signal. Sometimes BJTs are used as the current-source transistors in *V/I* converters instead of FETs for their higher output impedance. When a source is supplying the timing current

through a diode bridge, the triangle wave is at its output. Whatever parasitic output capacitance the current source has affects the timing. The additional isolation offered by a BJT switch over a diode minimizes this problem.



Within the TWG loop is the hysteresis switch. Schmitt triggers of the regenerative MV kind are sometimes used, but their thresholds are often not accurate enough. The dual-level comparator circuit (above) is more accurate. The comparator outputs are asserted (high) only momentarily as the ramp crosses a threshold at a peak. The RS flop, made of NOR gates, is set to alternating states. This is not the fastest circuit since it consists of several stages of processing and can be used up to about 20 MHz.

Not only must the thresholds be symmetric; a hysteresis switch must also have little delay because the total loop delay determines the maximum operating frequency. As loop delay time t_d becomes an appreciable fraction of the period T, the ramp increases in magnitude beyond the threshold before the slope changes. The triangle-wave amplitude thus begins to increase with frequency.



The effect is graphed. V_M is the low-frequency triangle amplitude and hysteresis threshold. Because of delay, it increases to $v_M(f)$. For a ramp slope of m,

$$v_M = V_M + m \cdot t_d$$

where the slope is

$$m = \frac{v_M}{T/4} = \frac{4v_M}{T}$$

Substituting into v_M,

$$v_M = V_M + \frac{4v_M}{T} \cdot t_d = \frac{V_M}{1-(4t_d/T)} = \frac{V_M}{1-(\omega/\omega_d)}, \quad \omega_d = \frac{\pi}{2t_d}$$

Using the approximation

$$\frac{1}{1-x} \cong 1+x, \quad x << 1$$

then

$$v_M \cong V_M \cdot \left(1+\frac{\omega}{\omega_d}\right), \quad \omega << \omega_d$$

This equation also results from assuming the low-frequency slope of

$$m = \frac{4V_M}{T}$$

in v_M. The period increases over its low-frequency value by $4 \cdot t_d$. From the v_M approximation, ω_d is like a zero break frequency for v_M, but from the exact expression, v_M is vertically asymptotic at ω_d. As ω approaches ω_d, t_d dominates T. A t_d of 10 ns breaks at

$$f_d = \frac{1}{4t_d}$$

or 25 MHz.

Because $v_M(\omega)$ is nonlinear, a nonlinear compensator is required to make $v_M(\omega) = V_M$. Transfer-function compensation is based on linear analysis, requiring another approach. This is an adaptive control problem; a parameter V_M must be varied to achieve ideal compensation. The amplitude error is

$$\Delta v_M = v_M - V_M = \frac{V_M \cdot (\omega/\omega_d)}{1-(\omega/\omega_d)} = v_M \cdot \left(\frac{\omega}{\omega_d}\right) = v_M \cdot \left(\frac{f}{f_d}\right)$$

If V_M is replaced by $V_M - \Delta v_M$ or if the triangle wave itself, v_{TW}, is modified to $v_{TW} + \Delta v_M$, then v_M becomes a constant V_M. Because the TWG frequency is determined by the VCF voltage v_F, frequency information can be derived from it to set the comparator thresholds. The slope of v_{TW} is I/C. If I is generated by a V/I converter by VCF voltage v_F across timing resistor R, then,

$$m = \frac{v_F}{RC}$$

For $\omega << \omega_d$, the low-frequency slope applies. Equating to m above,

$$f = \frac{v_F}{4 \cdot R \cdot C \cdot V_M}$$

and the adaptive V_M is

$$\text{adaptive } V_M \rightarrow V_M - \Delta v_M = V_M - \frac{V_M}{(V_M/v_F)(RC/t_d)-1} \cong V_M - v_F \cdot \left(\frac{t_d}{RC}\right)$$

This is an instance of model-reference adaptive control and is based on a priori knowledge of the circuit-model f. When R or C are switched to change frequency ranges, the model changes. Switching of RC in the adaptive V_M is required for the model to represent the circuit.

Another approach to TWG-loop compensation is to place in the loop a time advance to cancel t_d. In the s-domain, the delay is e^{-st_d} and a compensating advance is e^{st_d}. Time advances are not realizable in physical (causal) systems, and this form of compensation can be only approximated. The power-series expansion of the time delay is

$$e^{-st_d} = 1 - st_d + \frac{s^2 t_d^2}{2} - \frac{s^3 t_d^3}{6} + \cdots \cong 1 - st_d \cong \frac{1}{st_d + 1}$$

The single-pole approximation to a time delay suggests that a phase-lead compensator with zero at $1/t_d$ and higher-frequency pole provides approximate compensation.

A rational approximation to the time advance is the Padé approximation. A first-order approximation is

$$e^{st_d} \cong \frac{st_d/2 + 1}{-st_d/2 + 1}$$

It is a nonminimum-phase transfer function. A second-order Padé approximation is

$$e^{st_d} \cong \frac{s^2 t_d^2 + 6st_d + 12}{s^2 t_d^2 - 6st_d + 12}$$

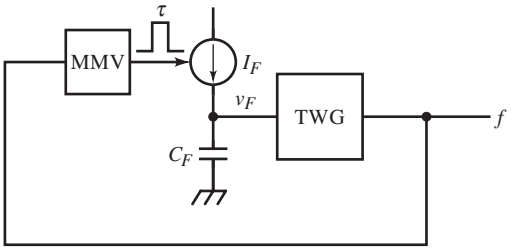
It can be factored, which results in

$$z_{1,2} = \left(\frac{1}{t_d}\right) \cdot \left(3 \pm j\sqrt{3}\right), \quad p_{1,2} = \left(\frac{1}{t_d}\right) \cdot \left(-3 \pm j\sqrt{3}\right)$$

The pole-zero placement is symmetric about the origin and represents an all-pass filter. This is consistent with the delay function in that it effects only a shift in time with no amplitude change. The pole and zero pairs have an angle of

$$\phi = \tan^{-1} \frac{\sqrt{3}}{3} \cong 30°$$

Besides being rational approximations, these delay compensators are linear whereas the circuit is nonlinear. Frequency-domain analysis assumes sinusoids, not triangle waves. For $\omega << \omega_d$, the nonlinearity is not too severe and can be approximated as linear, and the phase error in the rational approximations is minimal.



The versatility of the FG is due partly to its wide VCF range. For frequency-response or Bode magnitude plots, a logarithmically swept VCF directly produces a log-frequency plot on an oscilloscope display. A simple way of producing a logarithmic sweep is shown above. On each cycle of the FG output, the MMV is triggered. It gates on a current source for a fixed time, transferring a fixed charge to the VCF capacitor C_F. This increments v_F by a fixed amount, causing the output frequency to increment. As the frequency increases, the rate of increase of v_F increases along with it and is exponential. The resulting frequency sweep, when displayed, is stepwise logarithmic.

Quantitatively, let the MMV time-out be $\tau < T = 1/f$. During a given time interval Δt, n periods of the output occur. Then

$$\frac{\Delta v_F}{\Delta t} = \frac{n \cdot (I_F \tau / C_F)}{n \cdot T} = f \cdot \left(\frac{I_F \cdot \tau}{C_F} \right)$$

In the limit, for small Δv_F or infinite n,

$$\lim_{n \to \infty} \frac{\Delta v_F}{\Delta t} = \frac{dv_F}{dt} = \left(\frac{I_F \tau}{C_F} \right) \cdot f$$

As a voltage-to-frequency (V/F) converter (VFC),

$$f = k_F \cdot v_F$$

Substituting into the limit equation and solving,

$$v_F = v_F(0) \cdot e^{(I_F \cdot \tau / C_F) \cdot k_F \cdot t}$$

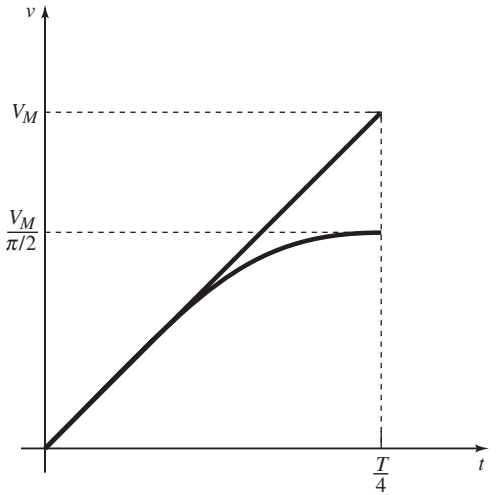
The sweep-rate constant is $(I_F \cdot \tau / C_F) \cdot k_F$, and the sweep time is

$$t_{SWP} = \frac{\ln(f/f(0))}{(I_F \cdot \tau / C_F) \cdot k_F}$$

where $f(0)$ is the starting frequency and f the ending frequency. The VFC constant k_F can be found from the TWG parameters. For triangle-wave amplitude of V_M, the slope is

$$\frac{V_M}{T/4} = \frac{I}{C} = \frac{v_F/R}{C} \Rightarrow f = \left(\frac{1}{4RC} \cdot \frac{1}{V_M}\right) \cdot v_F \Rightarrow k_F = \frac{1}{4RC} \cdot \frac{1}{V_M}$$

Besides VCF, the TWG current sources can be individually controlled for variable symmetry or duty-ratio. A *voltage-controlled symmetry* (VCS) input allows pulse-width modulation (PWM) of the output waveforms.



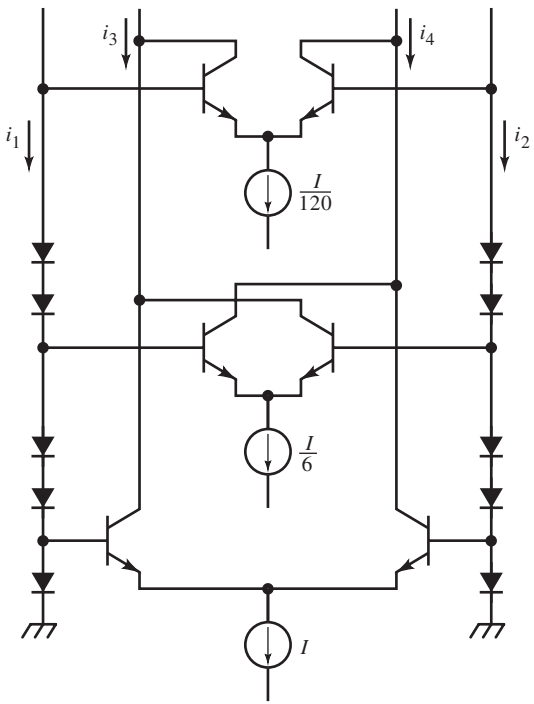
A third waveform available on FGs is the sine wave. This function is not generated by the TWG loop but by a sine converter circuit. Most commonly, this is a multiple-diode clamp that performs a piecewise-linear waveform shaping of

the triangle wave. Commonly used shapers use diode bridges for symmetry and diode drift cancellation. The initial part of the triangle and sine waves has the same slope. For sine amplitude of V_A, the slopes are equated:

$$\frac{V_M}{T/4} = V_A \cdot \omega \Rightarrow \frac{V_M}{V_A} = \frac{\pi}{2} \cong 1.57$$

The most significant anomaly of these sine shapers is the triangle-wave peaks in the sinusoid. At the peaks, the derivative of the triangle wave is discontinuous and is difficult to remove entirely from the sine wave. Typically, three-breakpoint sine shapers produce less than 0.25% total harmonic distortion (THD) in the audio range.

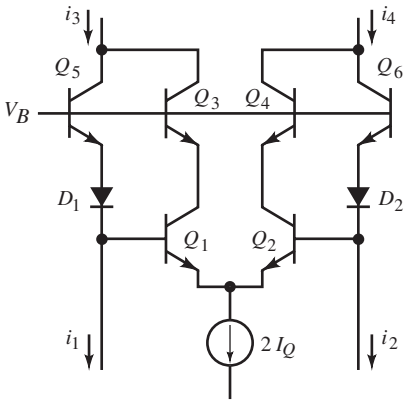
Another scheme uses the hyperbolic tangent function of the BJT diff-amp as an approximate sine converter. This results in somewhat more distortion than the multiple-clamp circuit. MOS diff-amps have a quadratic transfer function and also approximate a sine output with adjusted parameters.



A more elegant approach is to return to the translinear cell concept. By stacking diodes with diff-amp pick-offs, arbitrary power series expansions can be realized, as in the above circuit for three terms of a sine expansion. The Taylor-series expansion is

$$\sin\frac{i_1}{i_2} = \sin x = x - \frac{x^3}{6} + \frac{x^5}{120}$$

The truncation error is less than 0.07%. For two terms, the error is 0.14%, still favorable relative to the other schemes. This sine converter operates over an input range of x from 0.5 to 2.



The two-term sine shaper sketched above has a topology similar to a Gilbert gain cell. It passes the linear term through Q_5–Q_6 while Q_1–Q_2 develop the second term of the output. Together,

$$i_O = i_I \cdot \left[1 - \frac{2I_Q/I_I}{1+\left(i_I/I_I\right)^2}\right]$$

where

$$i_I = i_1 - i_2, \quad i_O = i_3 - i_4, \quad I_I = i_1 + i_2, \quad I_Q = i_{C1} + i_{C2}$$

The ratio $2I_Q/I_I$ is adjusted for minimum THD of less than 0.1% when the ratio is about 1.5.

Absolute-Value (Precision Rectifier) Circuits

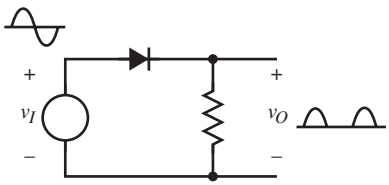
A full-wave rectifier performs the absolute-value function

$$|x| = \begin{cases} x, & x \geq 0 \\ -x, & x < 0 \end{cases}$$

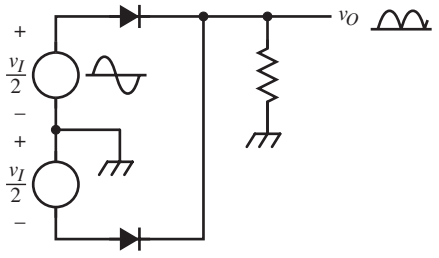
A half-wave rectifier performs the function

$$f(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

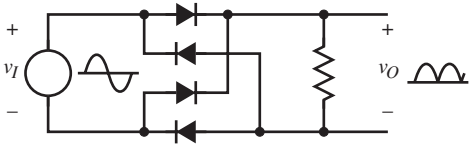
Half-wave rectification is mainly applied as a bipolar to unipolar (ac-to-dc) conversion technique in low-power power supplies or AM demodulators. Rectifiers can be designed to output either polarity; the positive outputs of $|x|$ or $f(x)$ could instead be $-|x|$ or $-f(x)$. For waveform processing involving mathematical functions, $|x|$ is more commonly needed.



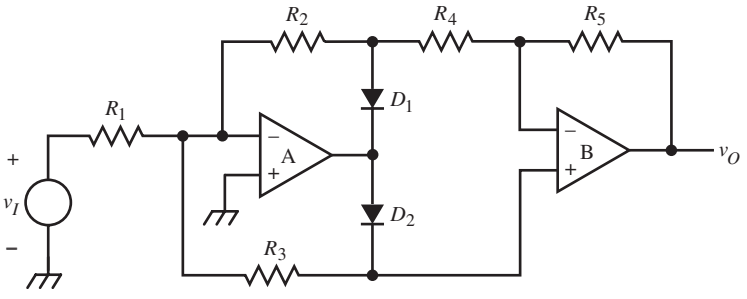
A single diode can half-wave rectify (above).



Two diodes are used for full-wave rectification with a ground-based differential input (above) and a diode bridge for a floating input (below).



A more recent addition to rectifier circuits is the *synchronous rectifier*. It uses active devices, usually MOSFETs, that are switched by the input itself. These circuits are associated with power conversion and are commonly found in power supplies, both linear and switched. The precision clamps of "Clamps and Limiters" are half-wave rectifiers.



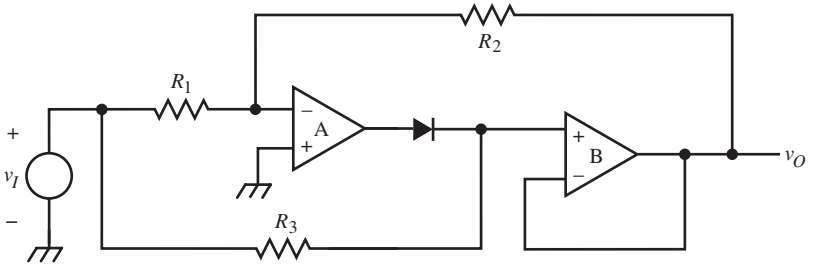
Absolute-value circuits can be designed many ways. The common constraint is that the gain magnitude for $v_I < 0$ be the same as for $v_I \geq 0$, so that $-A_{v-} = A_{v+}$. A common absolute-value circuit is shown above. It has similar frequency responses for positive and for negative inputs. On the positive half-cycle, D_1 conducts, and op-amp A operates in the inverting configuration. Op-amp B converts this output to a single-ended signal v_O. The gain expressions are

$$A_{v+} = \left(-\frac{R_2}{R_1}\right)\cdot\left(-\frac{R_5}{R_4}\right); \quad A_{v-} = \left(-\frac{R_3 \| (R_2 + R_4)}{R_1}\right)\cdot\left(\frac{R_5}{R_2 + R_4} + 1\right)$$

Equating A_{v+} to $-A_{v-}$, the constraint on resistor values is

$$R_3 = R_5, \quad R_2 = R_4$$

Because R_1 is unconstrained, it can set the gain. Op-amp B must have a fast large-signal response to follow the discontinuities in the waveform it amplifies.



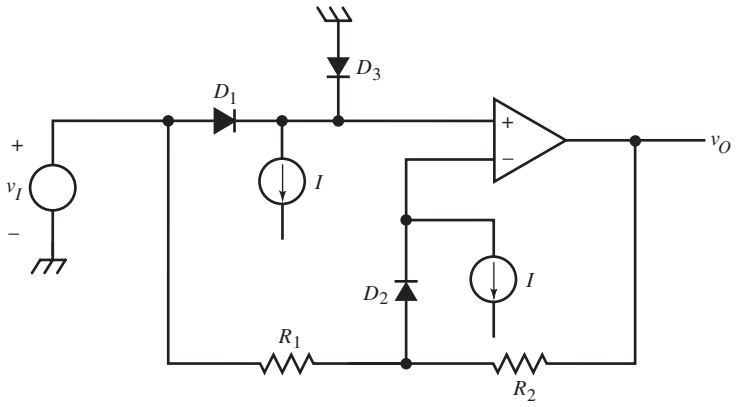
This circuit uses only one diode. Op-amp A functions for negative inputs as the first stage of gain in the forward path, cascaded with op-amp B, a ×1 buffer. The gain is

$$A_{v-} = -\frac{R_2}{R_1}$$

For positive inputs, op-amp A is disconnected from B, and v_I is applied directly to the input of B through R_3. The value of R_3 does not affect gain; its purpose is to limit current from op-amp A when the diode conducts. Thus, positive inputs bypass A and are merely buffered by B with a gain of $A_{v+} = 1$. The constraint is simply that $A_{v-} = -1$:

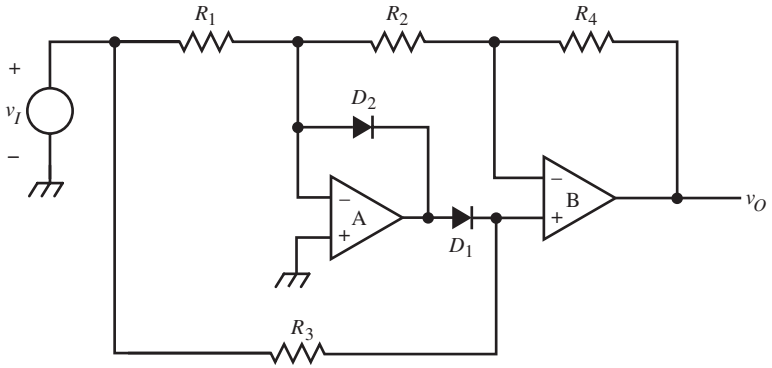
$$R_1 = R_2$$

The positive-gain path involves only one op-amp and has a faster response than the negative-gain path for negative inputs.



A similar kind of circuit with only one op-amp requires matched diodes D_1 and D_2 and current sources as a trade-off for fewer components. In IC form, this is attractive. The positive-gain path is through D_1 and the op-amp, with a gain of one. D_2 also conducts, causing the inverting op-amp input to follow the input. For negative inputs, D_1 is off, D_3 is on to satisfy the current source, and the op-amp inverts with a gain of $-R_2/R_1$. D_2 conducts all the time to balance D_1 or D_3. The constraint is

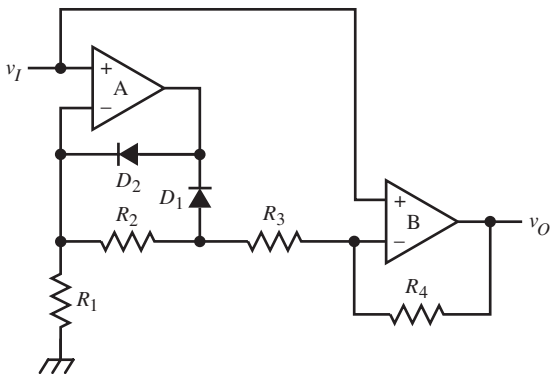
$$R_1 = R_2$$



Another variation on this theme, which allows adjustment of gain at the output by means of R_4, is shown above. The gains are

$$A_{v+} = \frac{R_4}{R_2} + 1,\ A_{v-} = \left(\frac{R_4 + R_2}{R_1} \right)$$

The constraint again applies. R_4 is not constrained and can be used to set the gain. D_2 clamps op-amp A input to virtual ground for positive signals so that R_1 does not affect A_{v+}.



This absolute-value circuit has a high-impedance input but is constrained to a gain of one. For $A_{v+} = 1$ then

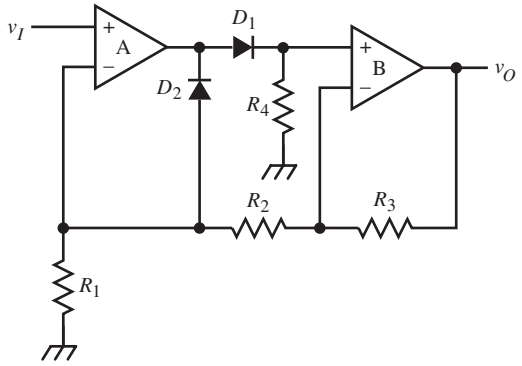
$$A_{v-} = \left(\frac{R_2}{R_1} + 1\right) \cdot \left(-\frac{R_4}{R_3}\right) + \left(\frac{R_4}{R_3} + 1\right) = 1 - \frac{R_2 \cdot R_4}{R_1 \cdot R_3}$$

The constraint is

$$R_4 = 2 \cdot \frac{R_1 \cdot R_3}{R_2}$$

One combination of resistors satisfying the constraint is

$$R_1 = R_2 = 2 \cdot R_3 = R_4$$



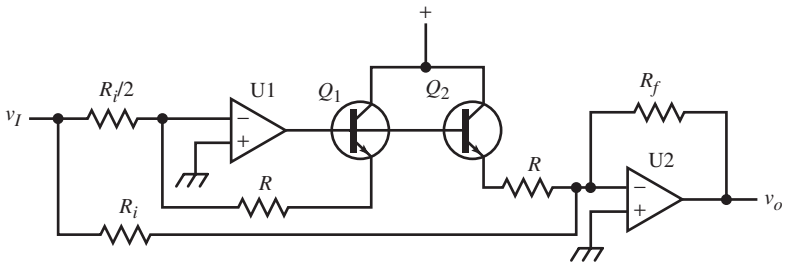
This absolute-value circuit also has a high-impedance input, with gains

$$A_{v+} = \frac{R_2 + R_3}{R_1} + 1, \quad A_{v-} = -\frac{R_3}{R_2}$$

and with the constraint

$$\frac{R_1}{R_2} = \frac{(R_3/R_2 + 1)}{(R_3/R_2 - 1)}$$

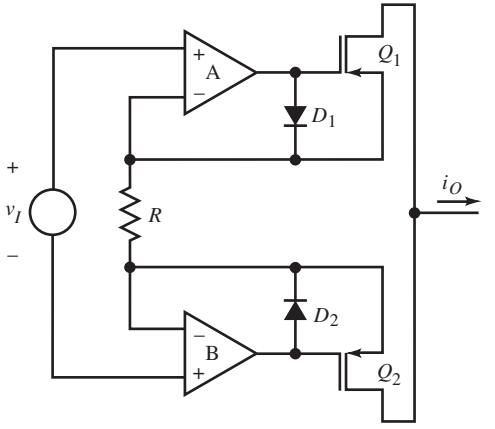
Implicit in the constraint equation is the additional constraint that $R_3 > R_2$.



The absolute-value circuit shown above is essentially that used in the AD534 root mean square (rms) circuit. The transistors and resistors must match for accuracy. For positive v_I, the BJTs are cut off, and input current flows in the lower path through R_i to the input of U2. For negative v_I, both BJTs conduct and present the same voltage at their emitters. Q_1 completes the path back to the U1 input while Q_2 conducts half that current through the lower R_i via $R/2$ to v_I. This causes the output op-amp to conduct the same amount of current as in R_i, and the circuit transfer function is

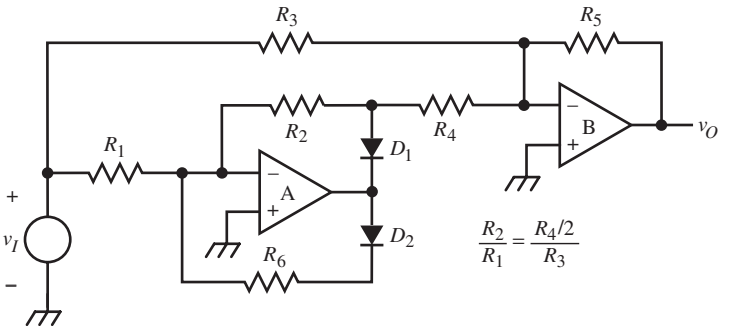
$$\frac{v_O}{|v_I|} = -\frac{R_f}{R_i}$$

As with most other absolute-value circuits, the path for one polarity (positive v_I in this case) is faster than for the other polarity. Also, design of the above circuit should take into account the reverse *b-e* breakdown-voltage limitation, typically around 5 V to 7 V. This limits not only the U1 negative output saturation voltage for positive inputs but also the positive range of v_I. A single-supply op-amp with grounded negative supply terminal allows for maximum positive v_I range.



Not all absolute-value circuits must be designed by matching gains of positive- and negative-gain paths. This circuit resembles an instrumentation amplifier but its output current is unipolar. For $v_I > 0$, D_1 and Q_2 conduct. D_1 reverse-biases Q_1. The negative-gain path is through D_2 and Q_1. The circuit is symmetric and has output current

$$i_O = \frac{|v_I|}{R}$$

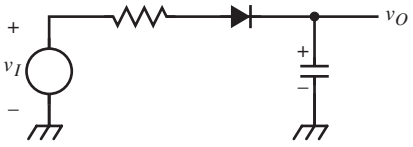


Finally, this circuit has a rectifier section similar to the first absolute-value circuit but with gain set by R_5. The positive-gain path is through A, then B; the negative-gain path is through R_3 and B. For negative inputs, D_2 conducts through R_6, a current-limiting resistor, forcing the inverting input to virtual ground. R_4 and R_2 then shunt the input of B, but this does not affect the gain.

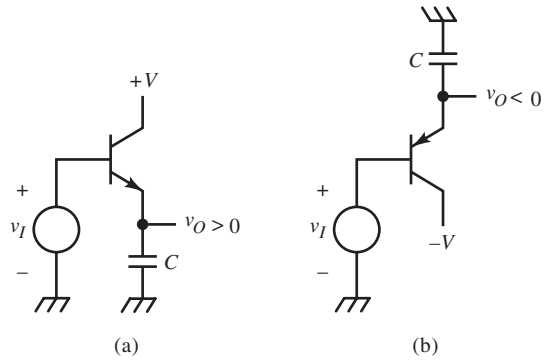
Some op-amps have inherently unipolar outputs and can be fashioned into precision rectifiers. When current mirrors, CMOS inverters, and other elemental circuits are also used, the collection of absolute-value circuits becomes extensive.

PEAK DETECTORS

Peak detectors are a class of circuits that extract the extrema of waveforms. They are essentially rectifier circuits with a memory. A simple example is the rectifier-filter combination of power supplies.

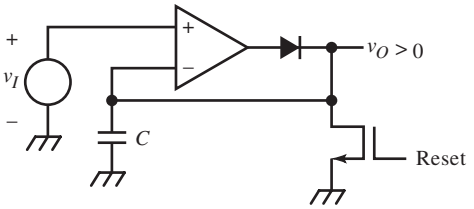


The diode conducts to charge the filter capacitor whenever the source-circuit voltage exceeds the capacitor voltage. The capacitor charges to the peak input voltage.

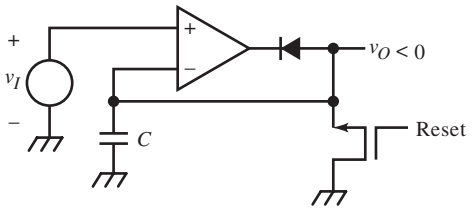


Fast, simple detectors (shown above) of maxima (positive peaks), (a), and minima (negative peaks), (b), use a CC stage that charges C.

Slower, more precise peak detectors, shown below, are reset through FETs that discharge the capacitors. Ideally, after C has been charged to the peak voltage, it retains its charge indefinitely. (In this respect, peak detectors are similar to S/H circuits.) The capacitor must have low leakage (high insulation resistance) to minimize the discharge rate. It must also have low dielectric absorption so that when reset, it retains 0 V until recharged by the input waveform. These requirements suggest a plastic or mica capacitor.

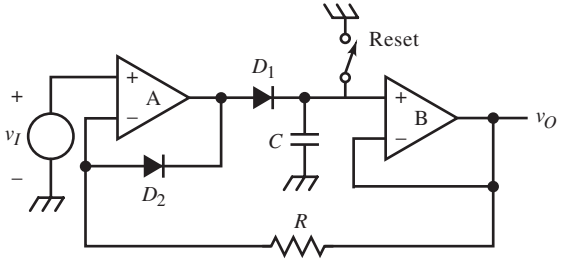


(a)



(b)

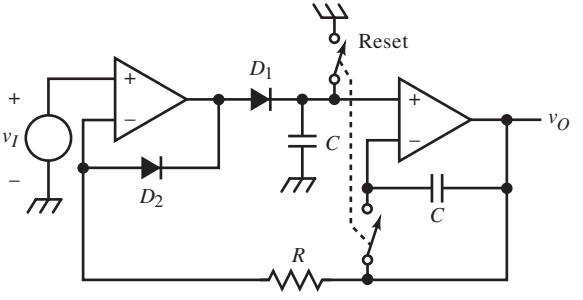
This applies also to its load, including the reset FET, and to the op-amp bias current. For the circuit below, load leakage is minimized by using a high-impedance buffer with low input bias current. When it drives the input op-amp, its offset-voltage error is compensated by being in the loop. The feedback loop also effectively reduces the time constant $r_{out} \cdot C$ by the loop gain, where r_{out} is the resistance in series with the diode. Fast peaks are thus detected more accurately.



To avoid overcharging, the feedback response must not be underdamped or overshoot occurs. Op-amp B must be faster than A to minimize loop delay and avoid overshoot. That is, a single dominant pole in the loop due to A yields a damped response. Because of the output loading of C and the additional pole it causes, op-amp A usually must be frequency compensated.

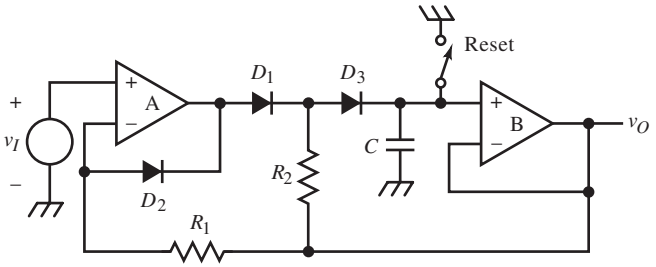
If op-amp A has limited output current, a CC buffer can replace the diode. This increases charging current by the β of the transistor and slew-rate increases by β.

D_2 and R are added to keep op-amp A from being driven into saturation when D_1 is off. R limits D_2 current. This enables the op-amp to respond more quickly as its output now follows the input. This also keeps its output from quickly switching to its saturation limits when D_1 cuts off, thus minimizing transient feed-through to the hold capacitor via D_1 shunt capacitance.



To further minimize bias-current error, a similar capacitor can be placed around op-amp B. Both capacitors are charged by the bias current, resulting in

the same Δv across each. The error voltage on the feedback capacitor subtracts from the hold-capacitor voltage, thus compensating its error. The feedback capacitor, however, must also be reset; two reset switches are required.



Finally, the diode must have a low reverse saturation current to minimize leakage when off. Reverse current varies with reverse voltage and is minimized by minimizing the voltage across the nonconducting diode. This can be done by, again, using the versatile technique of bootstrapping. D_3 and R_2 have been added to the first two-op-amp peak detector. v_O follows the capacitor voltage when holding, and R_2 applies this voltage to the anode of D_3, reducing its voltage to zero. D_1 blocks this node from the varying output of op-amp A, but its leakage is not critical for low values of R_2. When C is charging, R_2 isolates the diode node from the output of op-amp B. With this approach, special low-leakage diodes can be avoided in most applications. An alternative is to replace D_1 of the original peak detector with a junction field-effect transistor (JFET). Its gate-source junction typically has lower leakage than discrete diodes.

The bootstrapping technique of *leakage decoupling* can be applied to the reset switch also. Two switches are placed in series, with a resistor from the output connected to their series connection. Both switches are driven by the Reset signal.

Bootstrapping can also decrease acquisition time if input impedance is not critical. In the first two-op-amp peak detector, add a resistor in series with v_I and a bootstrap diode from output to the op-amp A noninverting input so that it conducts from output to input. Then the first fast peak charges C. Through the

bootstrap diode, the output of op-amp A is now driven to this voltage, which is near the peak. The next fast peak has less voltage over which to slew the output to further charge C.

Most of the charging time occurs when op-amp A output is near the peak voltage because it is not required to supply the large slew-rate-limited currents that a large voltage difference causes. Feedback then increases response time if $r_{out} \cdot C$ corresponds to frequencies at which the loop gain is still high.

The hold capacitor size is chosen as a trade-off between hold time and acquisition time. For fast peaks, a small C is preferred for faster charging. But a smaller capacitor develops hold error at a higher rate than a larger C. Therefore, C is chosen as a compromise between acquisition and hold-time requirements. A two-stage peak detector mitigates the trade-off. The first stage is optimized to be fast, whereas the second stage has a long hold time.

2

Digitizing and Sampling Circuits

Electrical Quantities Both Encode and Represent Information

An electrical quantity in time $x(t)$ is a *signal* when it encodes information. The information is interpreted according to a representational theory, such as logic theory for digital signals or transforms based on analogy for analog signals. The theory of representation is independent of the encoding scheme. In communications theory, encoding is called *modulation*. What the modulating signal represents is independent of its encoding. A thermometer output, for example, can be encoded in analog or digital form but represents temperature, regardless.

Another way to think about information encoding is as two levels of representation. The encoding scheme is a representation at the electrical level, and the encoded information represents a quantity that is independent of electricity. This "more abstract" level of representation has to do with the application. Consequently, electronics is useful in domains that have nothing to do with electronics because both signals and their processing operations have meaningful interpretations for the application.

Information can be encoded as discrete or continuous functions of either an electrical quantity (usually voltage or current) or of time. The information to be encoded and the encoding scheme can be either discrete or continuous. The compatibility of an encoding scheme with the encoded information is a design consideration. For example, discrete functions are often best represented by digital encoding. Engineers sometimes differ over the relative merits of discrete versus continuous encoding and processing of information. The difference between discrete, or *digital*, and continuous, or *analog*, encoding is so important that each constitutes a major subdiscipline within electronics.

A digital signal is discrete in both x and t. Binary encoding is by far the dominant digital encoding of x, where $x \in \{X_L, X_H\}$. These two values or "levels" are named *low* (X_L) and *high* (X_H) and represent binary logic states of *false* and *true*, or, in Boolean algebra, 0 and 1. Whether the low level represents *true* or *false* depends on the polarity of the logic; a low level is false in positive logic and true in negative logic. Digital encoding can have more than two levels. The number of levels equals the modulus or base of the number representation. For example, decimal numbers can be encoded in a 10-level scheme. As the modulus increases, the representation approaches a continuous form.

An example is the output of digital-to-analog converters (DACs). For an 8-bit DAC, 256 discrete levels may adequately approximate a continuous function in some applications. The DAC could, however, be considered an encoder of base-256 numbers. In this sense, continuous waveforms are of infinite modulus, and analog engineers are actually digital engineers who specialize in infinite-base encoding.

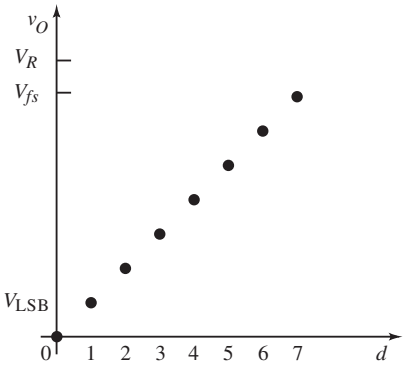
Discrete functions can also be encoded in time. Frequency-shift keying, a kind of binary FM used in modems, is one approach. More common to computer electronics is synchronous and asynchronous serial encoding of alphanumeric characters in ASCII. Many other purely digital encoding schemes make digital encoding and communications a specialty in itself.

Continuous functions can also be encoded purely in time as the width of a (binary) pulse (pulse-width modulation [PWM]) or by its time difference relative to another event in the signal (pulse-position modulation) or simply by the pulse frequency, as is the output of voltage-to-frequency converters (VFCs).

Finally, waveforms that are continuous in x and discrete in t are *sampled waveforms*. These waveforms are of great importance in association with analog-to-digital (A/D) and digital-to-analog (D/A) conversion and with sampled-data systems in general, systems that contain discrete-time waveforms, such as a motor servo controller with a digital position encoder, or any system with sample-and-hold circuits.

Digital-to-Analog Converters

Digital-to-analog converters convert digital input codes to output voltages or currents. The transfer curve for a unipolar three-bit DAC has discrete voltages at the discrete (integer) values of the digital input code d.



The digital code is an ordered set of bits that represent integers. Various number representations are possible, but the most common are shown in the table for three bits.

Integer	Offset Binary	Two's Complement		Sign-Magnitude		
3	111	011		011		+fs
2	110	010		010		
1	101	001		001		
0	100	000	000		100	zs
–1	011	111	↑	101	↑	
–2	010	110	0^+	110	0^-	
–3	001	101		111		
–4	000	100		↑		–fs
	inverted MSB (sign)			MSB 1 ⇒ –		

These are signed (bipolar) representations of integers. The most common number representations are two's complement and offset binary. They differ only in the polarity of their sign bit. Binary-coded decimal (BCD) is also sometimes used, in which the first 10 binary numbers represent a decimal number.

In the above table, the positive full-scale (fs) value is 3, and the negative fs value is −4, one greater in magnitude. This asymmetry results from assigning a state to zero. Sign-magnitude coding is symmetric, but it has two zero states. In general, for n bits, there are 2^n states. The transfer characteristic for a unipolar n-bit DAC is

$$v_O = V_R \cdot \left(\frac{d}{2^n} \right)$$

where V_R is the DAC reference voltage. The fs voltage is less than V_R because the

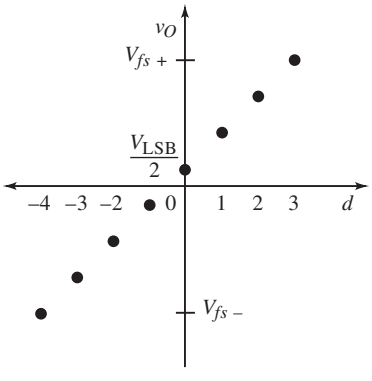
$$\text{maximum } d = 2^n - 1$$

Accordingly,

$$V_{fs} = V_R \cdot \left(\frac{2^n - 1}{2^n} \right) = V_R - \frac{V_R}{2^n} = V_R - V_{\text{LSB}}$$

That is, the fs output is less than V_R by V_{LSB}, the *quantum* voltage: the voltage difference corresponding to a difference of one input state. Because V_{LSB} is the smallest output voltage difference of the DAC, it is also its resolution, its minimum Δv_O.

The DAC input often represents a continuous function, but because it is discrete (or *quantized*), for values between integers the DAC output remains constant. In the plot shown above, v_O is zero over the interval [0, 1). (This is the *least-integer function*). At 1^-, infinitesimally below 1, $v_O = 0$ V, though the correct value is infinitesimally less than V_{LSB}. The output is in error by V_{LSB} at 1^- and has no error at zero. The magnitude of the error can be split so that the error range is $\pm 1/2\ V_{\text{LSB}}$ by offsetting v_O by $1/2\ V_{\text{LSB}}$. Then the DAC output fs magnitudes are also equal.



Quantization error causes *quantization noise,* which is a sawtooth function that cycles between $-V_{\text{LSB}}/2$ and $+V_{\text{LSB}}/2$ between each state ($\Delta d = 1$). The rms value of this noise v_n is

$$\text{rms } v_n = \sqrt{\frac{1}{\Delta d}\int_{-\Delta d/2}^{+\Delta d/2}\left(\frac{V_{\text{LSB}}}{\Delta d}\cdot d\right)^2 \cdot d(d)} = \frac{V_{\text{LSB}}}{\sqrt{12}} \cong 0.3 \cdot V_{\text{LSB}}$$

A signal-to-noise ratio (SNR) definition for n bits is

$$\text{SNR} = \frac{V_R}{\text{rms } v_n} = \frac{2^n \cdot V_{\text{LSB}}}{V_{\text{LSB}}/\sqrt{12}} = \sqrt{12}\cdot 2^n \cong 3.46 \cdot 2^n$$

In decibels, this is

$$\text{SNR (dB)} = 20 \cdot \log\left(\sqrt{12}\cdot 2^n\right) = 20 \cdot \log\sqrt{12} + 20 \cdot n \cdot \log 2 \cong 10.8 + 6.02 \cdot n$$

The dynamic range is about $6 \cdot n$ dB, and quantization noise is about 10.8 dB independent of the number of bits. Each additional bit increases the range by ×2, an octave, and this is 6 dB/octave, the slope of a Bode plot zero.

A different characterization of the SNR is as the ratio of rms signal to rms noise for a sinusoid. The rms value of a sine of amplitude V is $\left(\sqrt{2}/2\right)V$. Then

$$\text{SNR} = \frac{\text{rms sine}}{\text{rms noise}} = \frac{\left(2^n/2\right)\cdot\left(\sqrt{2}/2\right)\cdot V_{\text{LSB}}}{V_{\text{LSB}}/\sqrt{12}} = 2^n \cdot \left(\frac{\sqrt{6}}{2}\right) \cong 1.23 \cdot 2^n$$

In dB scaling, this is

$$\text{SNR (dB)} = 20 \cdot \log\left(\frac{\sqrt{6}}{2}\right) + 20 \cdot \log 2^n \cong 1.76 + 6.02n$$

The dynamic range remains the same under this definition of SNR, but the signal is less relative to the noise. This explains why the constant term of 1.8 dB is less than in the first SNR(dB).

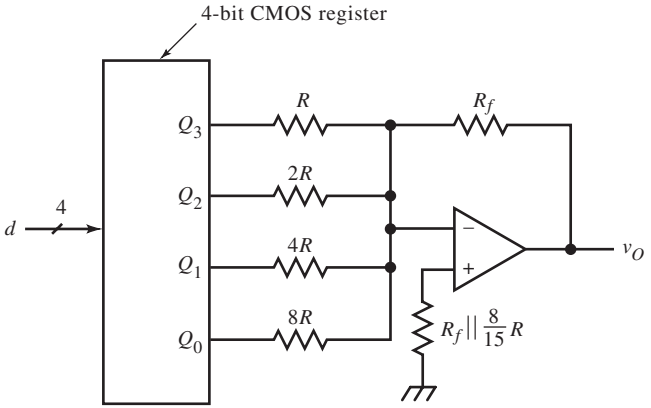
In actual DACs, the step size of quantum V_{LSB} is not constant, which affects the linearity of v_O/d. A measure of this nonlinearity is the *differential linearity error* (DLE), or *differential nonlinearity*, the amount a step differs from V_{LSB}:

$$\text{DLE} = \Delta v_{step} - V_{LSB}$$

If DLE exceeds V_{LSB}, the transfer curve is nonmonotonic, decreasing in output value with increasing d. This behavior can cause limit-cycle oscillation in control system applications. DACs also have offset and scaling (gain) errors, but these are nulled by external adjustment; the DLE cannot be.

DACs often output functions of time, $v_O(t)$; their dynamic response is important. This is characterized by the settling time to within ½ V_{LSB} of error. A dynamic anomaly of DACs is that when a large number of bits change in d, the effects of individual bits on the output are not exactly synchronized. At the output, momentary pulses or "glitches" appear until all the bits settle. This phenomenon is especially evident at midscale, when d changes from 011 to 100 (for a three-bit DAC). The change in most-significant bit (MSB) must be canceled by the combined changes of all the other bits, to within V_{LSB}, the correct Δv_O.

When glitches are unacceptable, as in CRT display systems, the DAC is followed by a deglitcher. These are either hf limiters or samplers with a delay. In delayed samplers, the DAC output is allowed to settle. Then it is sampled, and this value is output. The sampling control signal is delayed from the clock that changes DAC input states.



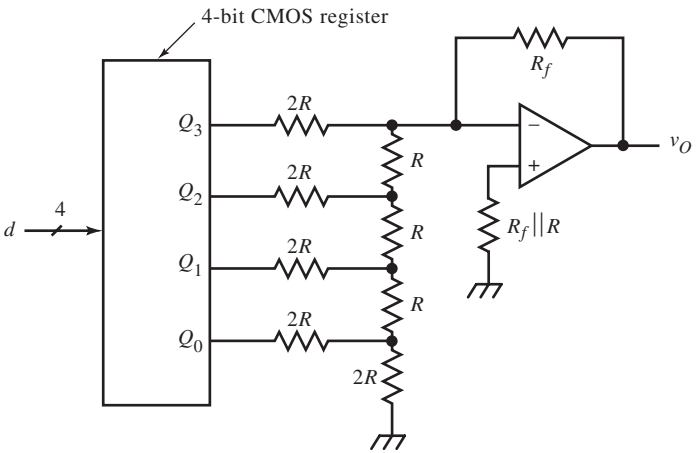
DAC designs are categorized as either bipolar junction transistor (BJT) or complementary metal-oxide semiconductor (CMOS). Both achieve binary weightings of voltage or current for each bit by a resistive network. Unless the number of bits is few (≤4), these are *R-2R* or *resistive ladder networks*. Otherwise, a set of binary-weighted resistors suffice. A four-bit binary-weighted resistor DAC is shown, voltage-driven by a CMOS register. The output is a function of each of the bits b_i of d:

$$v_O = -R_f \cdot \left(\frac{V_R}{R} \cdot b_3 + \frac{V_R}{2R} \cdot b_2 + \frac{V_R}{4R} \cdot b_1 + \frac{V_R}{8R} \cdot b_0 \right) = -\left(\frac{R_f}{R} \right) \cdot V_R \cdot \sum_{i=0}^{3} \left(2^{-i} \cdot b_{3-i} \right)$$

where b_0 is the least-significant bit (LSB) and V_R is the CMOS register supply voltage, the DAC voltage reference. (CMOS digital outputs accurately approach the supply rails.) The resistors must have sufficient precision to minimize DLE. The resistor requiring the most precision is at the MSB, R, since it must be within

$$\frac{\Delta R}{R} = \pm \frac{1}{2} \cdot \frac{R}{R_{\text{LSB}}} = \pm \frac{1}{2} \cdot 2^{-(n-1)} = \pm 2^{-n}$$

A four-bit DAC must have a tolerance on R of 6.3%. A 5% resistor suffices. For eight bits, the tolerance is 0.4%. This is difficult to achieve in monolithic form when the resistor values have such a wide range.

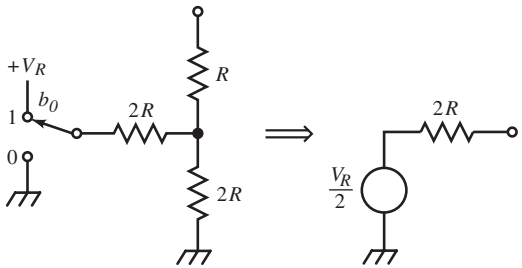


A standard alternative is the *voltage-switching R-2R* network, another four-bit DAC.

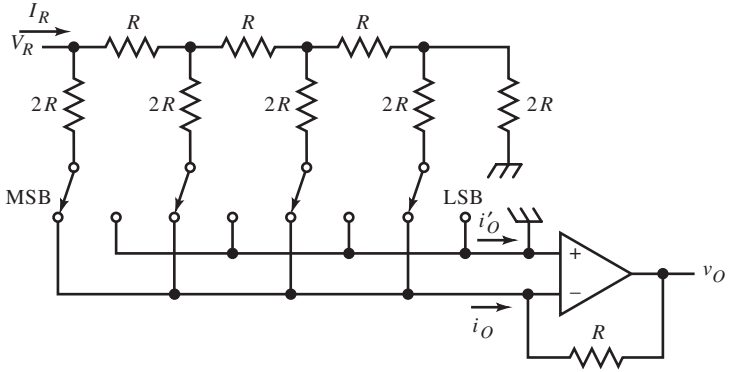
In general, the binary weighting,

$$W(d) = \sum_{i=0}^{n} 2^{-i} \cdot b_{n-i}$$

is the essence of DAC function. This weighting is achieved in the *R-2R* network. Beginning with b_0, if it is 1, Q_0 output is V_R; if it is 0, then the output is 0 V. For $b_0 = 1$, the Thevenin equivalent circuit is shown below.



At each stage of an *R-2R* network, the input resistance is $2R$ and the voltage of the previous stage is halved. From the input end of the network, the b_0 voltage is consequently halved four times. At the output, the op-amp is driven by a source resistance of R (another series R was not added to the network to make it $2R$) and voltage of $W(d)\,V_R$.

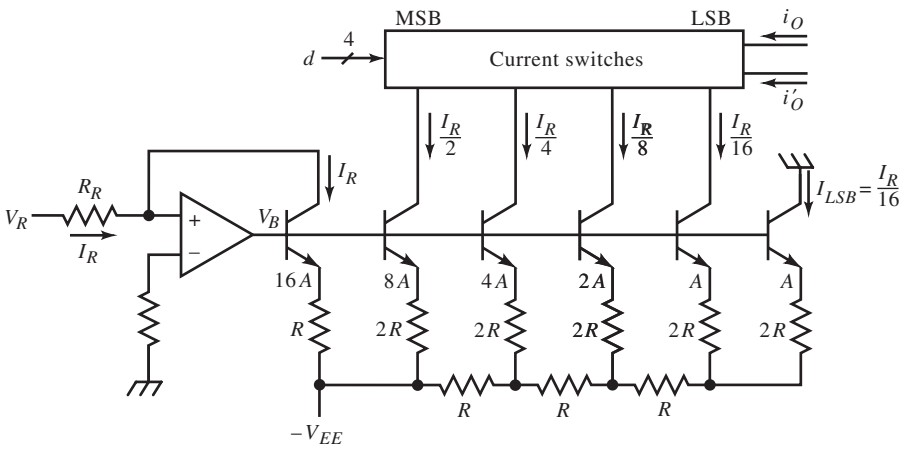


CMOS DACs are typically designed as shown above, with CMOS switches at the output and V_R at the input. This reversal does not change the operation except that the DAC outputs must be kept at ground (or virtual ground, as shown) to avoid errors in output current. This *current-switching R-2R* network is still voltage-driven, and the output is $-WV_R$ with op-amp feedback resistor R. Integrated circuit DACs often include this resistor to ensure its match with those in the network. The outputs are complementary currents that sum to a full-scale current,

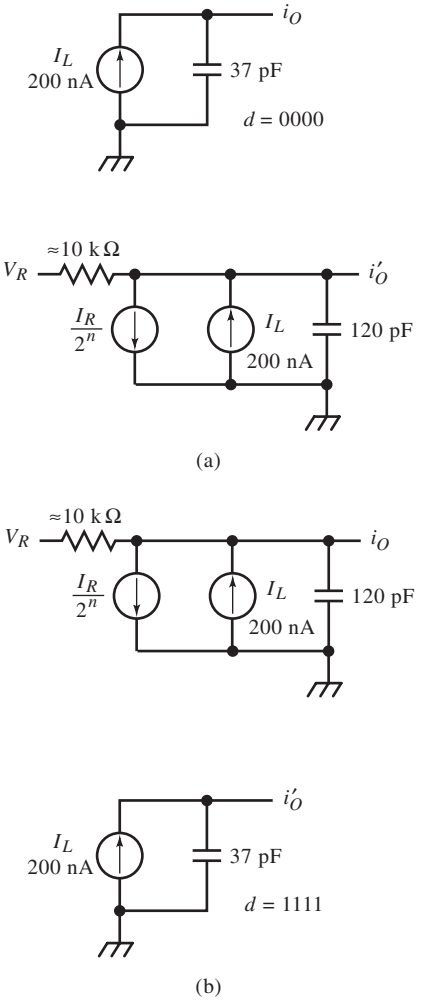
$$i_O + i'_O = I_{fs} = \frac{V_R - V_{\text{LSB}}}{R}$$

Switch resistance must be minimum (or binary-ratioed) for minimum network error. MOSFET switch areas are scaled to achieve equal voltage drops across all switches.

CMOS DAC output impedance changes with d. The two extremes are with all zero and all one bits (next page) for an AD7520, a 10-bit DAC with $R = 10\ \text{k}\Omega$ and leakage current I_L of 200 nA. Response compensation for op-amp input capacitance can be based only on an average or worst-case input state. Because output resistance varies extremely, op-amp bias-current compensation is also suboptimal.



Typical BJT DAC design is based on a current-switching *R*-2*R* network. An input op-amp establishes a reference current I_R in one of several BJTs with emitters connected to an *R*-2*R* network. The emitter areas are ratioed with the current each conducts, to maintain the same V_E for all BJTs. The BJT collectors drive current switches. These can be differential amplifiers (diff-amps) with logic-compatible inputs. Their collectors are connected to either the i_O or i'_O outputs. CMOS DAC outputs have no voltage compliance, but the BJT current outputs from collectors need not be held at a fixed voltage.



The *R*-2*R* network is not switched but is a multiple binary current divider. The voltage in the series-*R* string doubles at each successive stage toward the termination at the LSB. The reference current is established by the op-amp circuit as V_R/R_R. The output is

$$i_O = W(d) \cdot I_R$$

and the complementary output is

$$i'_O = I_{fs} - i_O = \left(\frac{2^n - 1}{2^n} \right) \cdot I_R - i_O$$

The equation for I_{fs} applies here. The relationship between I_R and I_{fs} is the same as for the CMOS DAC:

$$I_{fs} = \left(\frac{2^n - 1}{2^n} \right) \cdot I_R = I_R - I_{\text{LSB}}$$

The complementary output current is also related to $W(d)$ as

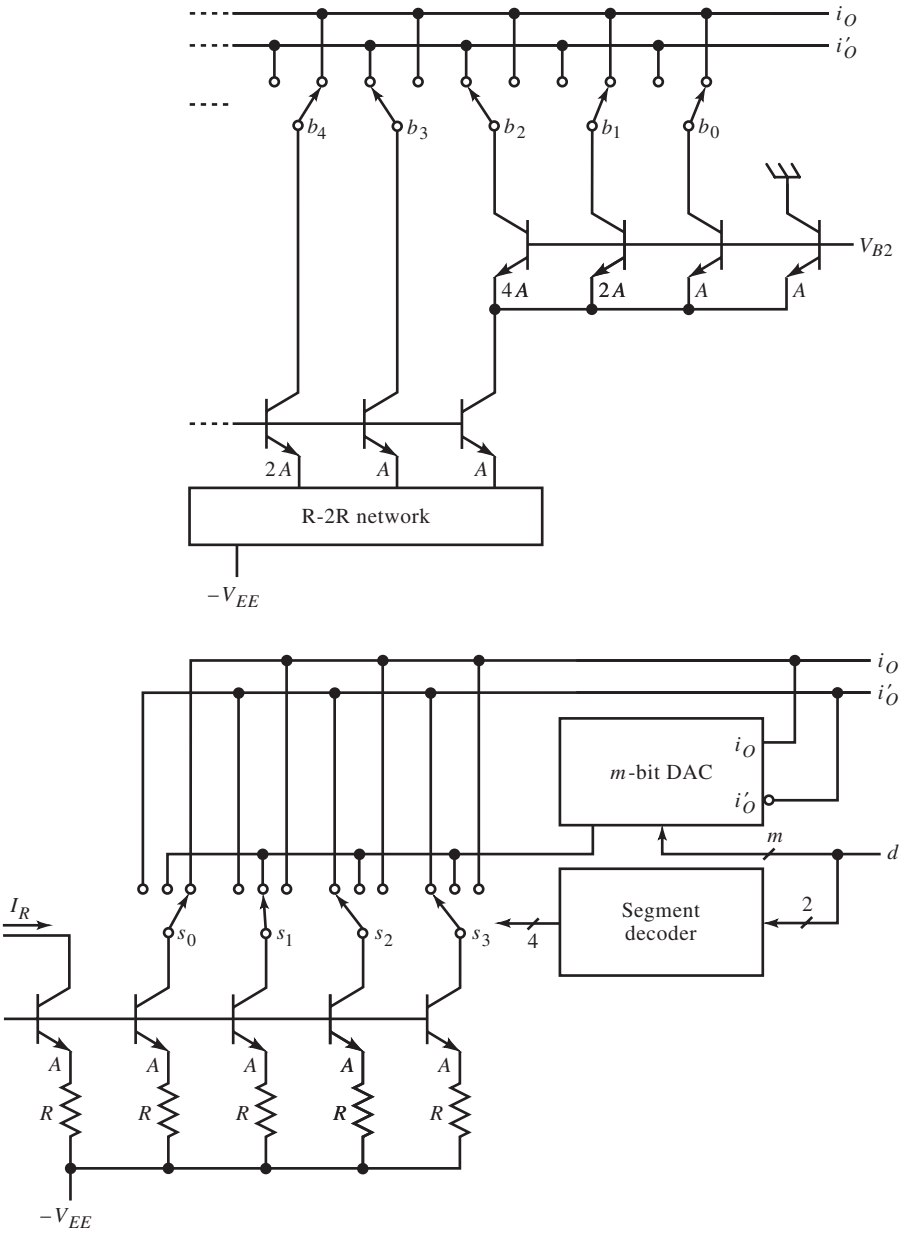
$$i'_O = \overline{W(d)} \cdot I_R$$

where

$$\overline{W(d)} = \sum_{i=0}^{n} 2^{-i} \cdot \overline{b}_{n-i}, \quad \overline{b}_{n-i} = \text{logical complement of } b_{n-i}$$

That is, $\overline{W}$ is the complementary weighting, the result of the bit-wise negation (or one's complement) of d.

Monotonicity among the LSBs, achieved with the *scaled-emitter* technique shown below, is used with the ladder network. The LSB terminating current of the ladder network, instead of being grounded, is fed to a second branch of emitter-scaled BJTs that switch the three LSBs. Monotonicity is ensured by branching.



For DACs with a large number of bits, the branching idea can be realized in a different topology. The *segmented* DAC, shown above for a two-bit segmentation, does current weighting with two networks. The input network is driven by a reference current as before. The two MSBs are decoded by a segment decoder.

They switch four equal currents to either output or to the branching input of an *R*-2*R* network of the remaining *m* LSBs. A segment decoder successively switches more segments with larger MSB codes to i_O as the *m* LSBs divide the current from one segment (s_1 in the figure) between i_O and i_O'. The remaining segment currents go to i_O'. For $m = 2$, when $d \leq 0011$, all segments except s_0 are switched to i_O'. When $d = 0100$, s_0 switches to i_O, and s_1 switches to drive the branch DAC, as shown. The remaining switches stay on i_O'.

Although the branching effect assures monotonicity, the transfer curve linearity can be much worse than the DLE. But in many applications, the DLE is all that matters for linearity. The match of the segment currents does not determine DLE, only the overall linearity. The segmented DAC uses fewer resistors since each segment requires only one resistor, not two as in a ladder network.

A very simple DAC design is the serial-output DAC, easily realized by one filtered output line of a computer. A PWM generator, either in hardware or software, drives a low-pass filter with break frequency far below the PWM frequency. The average output voltage is proportional to the duty-ratio. The disadvantage of this scheme is that it is slow and inherently noisy due to ripple from the filtered pulse. But if the pulse amplitude is accurate and the transitions are fast, a high-resolution output is achievable in direct trade-off with response time.

The ripple amplitude varies with pulse duration, which depends on the duty-ratio *D*. For high or low *D*, ripple is least and is highest at $D = 50\%$. In steady-state, the ripple extends from v_L to v_H around the average, *DV*. The ripple amplitude,

$$\Delta v = v_H - v_L$$

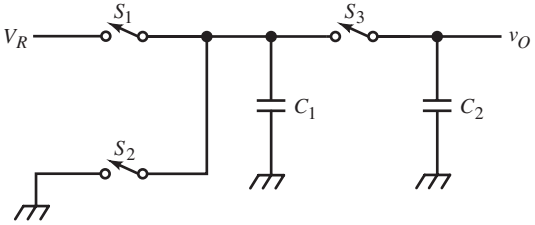
is derived from the decaying exponential, when the pulse is low:

$$\frac{v_H - v_L}{v_H} = \frac{\Delta v}{v_H} = e^{-(1-D)\cdot T/\tau} \cong \frac{\Delta v}{DV}$$

The approximation assumes that ripple is small relative to average output voltage and that $v_H \cong D \cdot V$. Also, τ is the filter time constant. Solving for the ratio of PWM frequency to filter break frequency for 1 LSB of ripple,

$$\frac{f_{\text{PWM}}}{f_{bw}} \cong -\frac{2\pi \cdot (1-D)}{\ln\left(1-2^{-n}/D\right)}, \quad \Delta v = V_{\text{LSB}}$$

For four bits, f_{PWM} must be 23.5 f_{bw} at $D = 0.5$ and 401 f_{bw} for eight bits. At 10% duty-ratio for four bits, the frequency ratio is only 5.8 and is 8.7 for 90%. (At the extremes of D, the approximation fails. At $D = 2^{-n}$ and $D = 1$, the ratio is 0.) Other pulse waveforms from rate-multipliers or statistically biased digital pseudo-random noise have less ripple for the same clock-to-filter frequency ratio but are harder to generate.



A serial-input DAC is shown above, with three switches and two equal capacitors $C_1 = C_2$. A serial digital input begins with the LSB. The DAC operates in two phases for each successive bit. On phase 1, switch S_3 is open, and a serial input bit closes either S_1 (for $b = 1$) or S_2 (for $b = 0$), charging C_1 to

$$q_i = C \cdot b_i \cdot V_R$$

On phase 2, S_1 and S_2 are open, and S_3 is closed. C_2 contains the net charge from previous cycles. On the ith bit on phase 1, this charge is

$$q_{i-1} = C \cdot v_{i-1}$$

On phase 2, S_3 closes, and

$$v_i = \frac{q_i + q_{i-1}}{2 \cdot C} = \left(\frac{V_R}{2}\right) \cdot b_i + \frac{v_{i-1}}{2}$$

For n bits, the voltage is

$$v_n = \left(\frac{V_R}{2}\right)\cdot b_n + \left(\left(\frac{V_R}{4}\right)\cdot b_{n-1} + \left(\left(\left(\frac{V_R}{8}\right)\cdot b_{n-2} + \ldots\right)\right)\right)$$

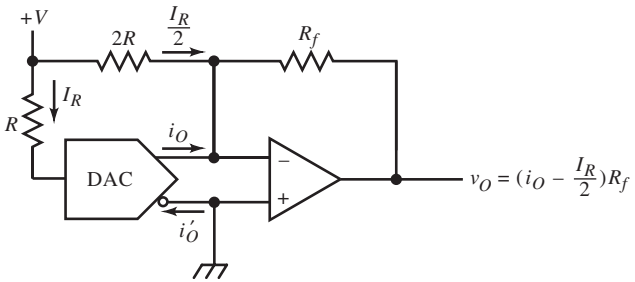
This iterative equation reduces to the closed form of

$$v_n = \left(\sum_{i=1}^{n} 2^{-i}\cdot b_{n-i}\right)\cdot V_R = W(d)\cdot V_R$$

After n bits, v_n is the converted voltage. It must be stored separately for output during the next conversion. The number of bits of monotonic conversion is limited by capacitor matching and switch leakage.

DIGITAL-TO-ANALOG CONVERTER CIRCUITS

In the figure below, a digital-to-analog converter is used as a component in a converter circuit. The DAC schematic symbol is used, with a small circle at i_O' to indicate the complementary output, after the convention of logic symbols. The op-amp output is bipolar and is offset by $I_R/2$ by a resistor of $2R$, where R is the current-reference resistor. Without this offset, v_O is unipolar, ranging from 0 V to $I_{fs}\cdot R_f$. By shifting i_O down by $I_R/2$, v_O at negative fs is one V_{LSB} greater in magnitude than positive fs. $I_R/2$ corresponds to the midscale or zero state of d.



The *offset-binary* output is

$$v_O = \left(i_O - \frac{I_R}{2}\right) \cdot R_f = \left(W - \frac{1}{2}\right) \cdot I_R \cdot R_f$$

Some output values for a four-bit d are tabulated below.

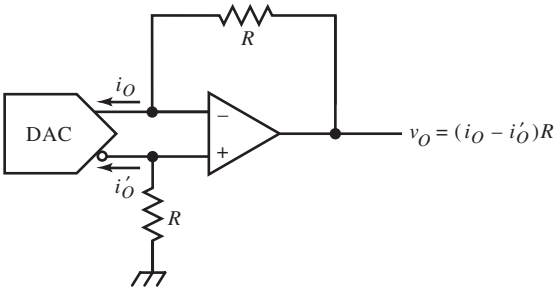
d	$W - ½$
1111	7/16
1000	0
0111	–1/16
0000	–1/2

A two's complement coding of d produces the same results when the MSB is inverted.

In the following DAC circuit, the output range is symmetric about zero. The op-amp is driven differentially by the DAC output so that

$$v_O = i_O \cdot R - i'_O \cdot R = \left(i_O - i'_O\right) \cdot R = \left(W - \bar{W}\right) \cdot I_R \cdot R = \left(2 \cdot i_O - I_{fs}\right) \cdot R$$

The expression for v_O in terms of W follows from the equations for i_O and i'_O.



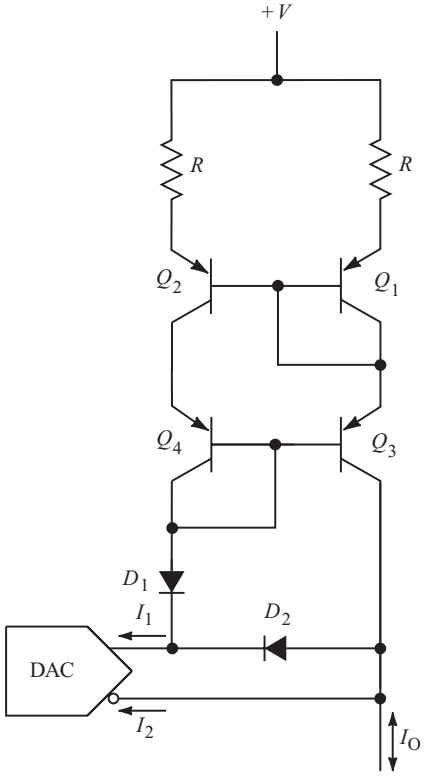
Compared with the offset-binary output, the symmetric-offset output range and step size are twice as large because i'_O is used. The last expression of v_O, compared with the offset-binary v_O, has twice the gain $(2 \cdot i_O)$ and a comparable offset difference of

$$\frac{I_{\text{LSB}}}{2} = \frac{I_R}{2} - \frac{I_{fs}}{2}$$

This leaves the symmetric v_O with a $V_{\text{LSB}}/2$ positive offset relative to the offset-binary output. Some four-bit output values are the following:

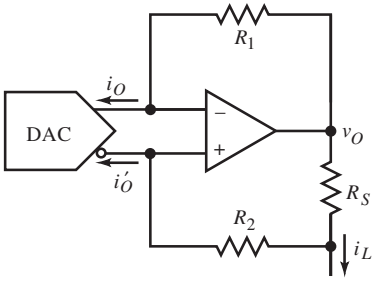
d	$W - \overline{W}$
1111	15/16
1000	1/16
0111	–1/16
0000	–15/16

The extreme states have outputs of equal magnitude while zero is offset by $V_{\text{LSB}}/2$. An inverted output results from exchanging the DAC outputs.



A DAC bipolar current source with a current mirror is shown above, taken from "Current Sources."

Another circuit, one that does not require a current mirror, is shown below. It has similar topology to a Howland current source but is simpler in operation.



The load current is

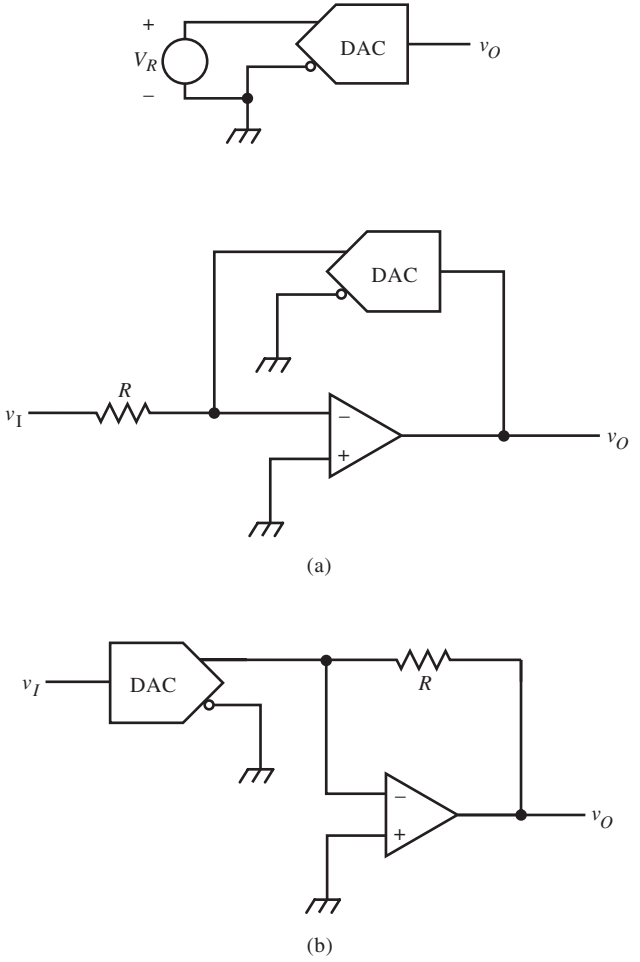
$$i_L = \frac{v_O - v_L}{R_S} - i'_O = \frac{i_O \cdot R_1 - i'_O \cdot R_2}{R_S} - i'_O = \frac{i_O \cdot R_1 - i'_O \cdot (R_2 + R_S)}{R_S}$$

This reduces to

$$i_L = \left(\frac{R_1}{R_S}\right) \cdot (i_O - i'_O), \quad R_1 = R_2 + R_S$$

The differential current output from the DAC is converted into a bipolar single-ended current output.

Because CMOS DACs are passive, switched ladder networks, they can be used "backward" as shown below. The reference voltage is applied across the i_O terminals, and the voltage output is taken from where the reference voltage is normally applied. This scheme is similar to that of the discrete CMOS DAC without the op-amp. The CMOS switches are driven from the supply voltage V_{CC}, and minimum switch resistance (and linearity error) results by keeping V_R well below V_{CC}.



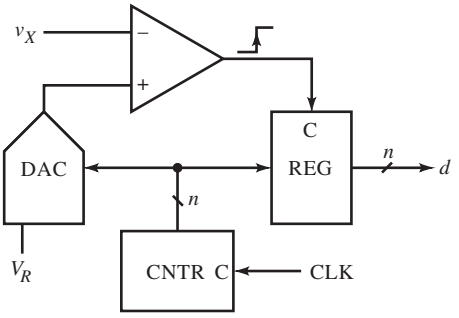
DACs can be combined with op-amps to provide programmable gain. The programmable-gain amplifier (PGA) in (a) above has a wide gain but can have significant voltage offset errors, whereas in (b) the op-amp gain is limited to ×2. In both circuits, R is included in the DAC IC and matches and tracks the ladder resistances. The noninverting configurations are similar in concept. For applications in which the digital input is a dynamic waveform and not merely a scale factor, it is multiplied by v_I; the DAC multiplies a digital by an analog quantity.

PARALLEL-FEEDBACK ADCS

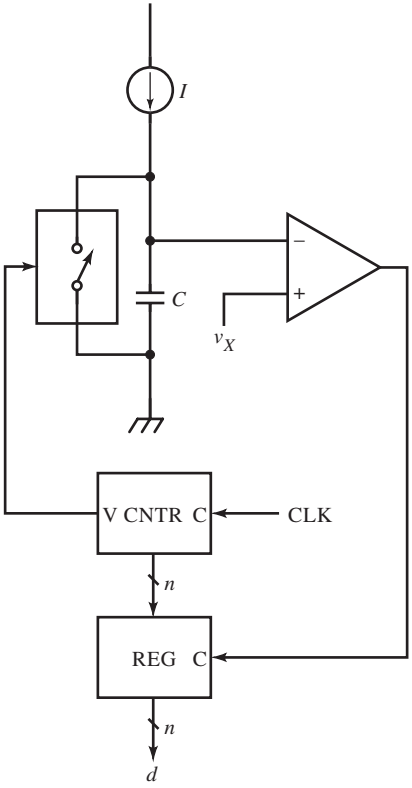
The inverse function of D/A conversion is *analog-to-digital* conversion, performed by A/D converters (ADCs). We consider here four categories of ADCs, which include many variations. Approximate ranges for conversion rate and precision are given below.

ADC Type	Conversion Rate	Precision, Bits
Integrating	0.1 Hz–100 Hz	14–22
Cyclic (serial)	1 kHz–100 kHz	10–16
Parallel-feedback	50 kHz–10 MHz	8–12
Parallel (flash)	10 MHz–1 GHz	4–8

Parallel-feedback converters are based on a concept similar to that of placing a function block in the feedback loop of an op-amp to achieve the inverse function.

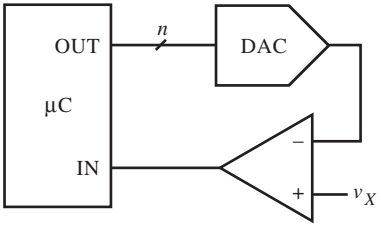


The figure above shows a digital realization of the *ramp* converter. A counter driven by a clock generates a digital sawtooth output. It drives a voltage-output DAC that outputs the ramp in analog form. When it crosses v_X, the comparator output clocks the register and holds the digital count. When the counter overflows, the DAC output resets to its minimum value, and the comparator output goes low, completing the cycle. The comparator output is also an end-of-conversion signal indicating valid register data.



In the analog realization shown above, the same concept uses an analog current-source ramp generator. The counter overflow turns on the reset switch, discharging the capacitor at the end of the conversion cycle. The analog circuit is subject to errors in ramp slope relative to the clock frequency. The digital form, although not having these timing errors, must have an accurate DAC reference voltage.

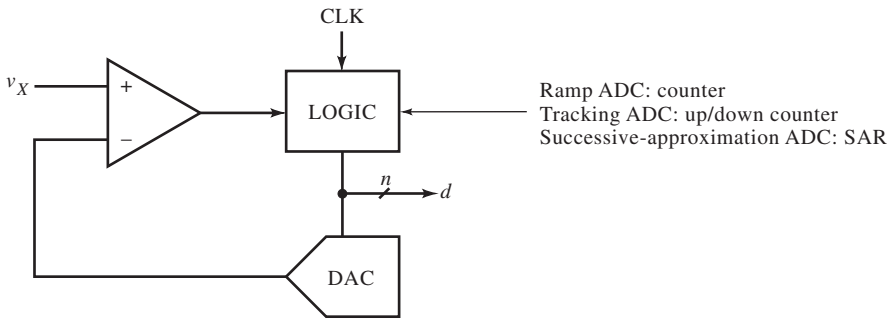
A third realization of the ramp converter makes use of a microcomputer (μC) (or any computer) and minimal additional hardware: an n-bit DAC and a comparator. The μC must have one digital input bit from the comparator and n output bits to drive the DAC. The software algorithm for ramp conversion uses software variable, VX, to hold the digitized value of v_X, and OUT to hold the DAC output value.



The ramp ADC procedure is as follows:

0. **Ramp ADC**
1. Set OUT to zero: OUT ← 0.
2. Input the IN bit.
3. If IN = 0, then VX ← OUT; go to 1.
   Else increment OUT: OUT ← OUT + 1.
4. Output OUT; go to 2.

Conversion time is usually limited by the μC; however, for many applications, it is fast enough, and the few additional components are an advantage. The ramp ADC is a poor technique and is seldom used. The conversion time varies but can take up to 2^n clock periods.



Parallel-feedback converters have a generalized topology. The type of logic block used determines the type of converter. The ramp ADC uses a simple counter. A slightly better ADC is the *tracking* converter. Its logic is a bidirectional

(up/down) counter. As v_X changes, the comparator output causes the counter to count up if the DAC input d is low and down if it is high. The counter servos the DAC to minimize input error at the comparator. Since the counter counts either up or down, the error is always ±1 LSB. For a constant input, a converged counter dithers by one state around the correct value; the comparator output alternates logic levels each clock cycle.

The tracking ADC is an improvement over the ramp ADC because it can be used to follow an input waveform, digitizing it as it occurs (that is, in *real time*). For small input changes, the counter must change only a few states. This is done in a few clock periods, and conversion is fast. For large input changes, such as a square-wave step, the converter shows slew-rate limitations and a longer conversion time. The DAC output *tracking slew-rate*, if limited by the counting rate, is

$$\frac{dW}{dt} = \frac{V_{fs}}{2^n} \cdot f_{\text{CLK}}$$

The clock frequency is limited by the loop delay time: the DAC settling time, comparator delay time, and counter clock-to-output time. For a sinusoidal input,

$$v_X(t) = \frac{V_{fs}}{2} \cdot \sin(\omega \cdot t)$$

its maximum slew-rate is $\omega \cdot V_{fs}/2 = \pi \cdot f \cdot V_{fs}$. Equating to the tracking slew-rate and solving for the maximum fs sine frequency,

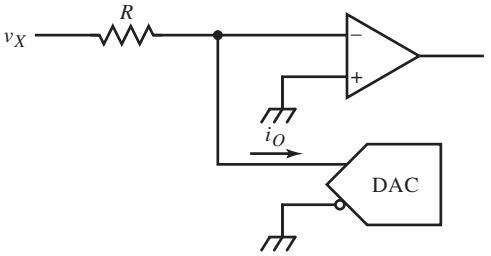
$$\max \text{fs sine } f = \left(\frac{1}{\pi \cdot 2^n} \right) \cdot f_{\text{CLK}}$$

The tracking converter can be implemented with the same hardware as the ramp converter. The general μC-based parallel-feedback ADC also applies generally to parallel-feedback converters. Instead of hardware logic, the software logic distinguishes among parallel-feedback ADC types. A tracking ADC procedure, based on the same software variables as the ramp ADC, is given below:

0. **Tracking ADC**
1. Output OUT.
2. Input IN.
3. If IN = 0, then decrement OUT: OUT ← OUT − 1.
   Else, increment OUT: OUT ← OUT + 1.
4. Set VX to OUT: VX ← OUT.
5. Go to 1.

This procedure is not more complicated than that for the ramp converter but has the performance advantages of the tracking ADC.

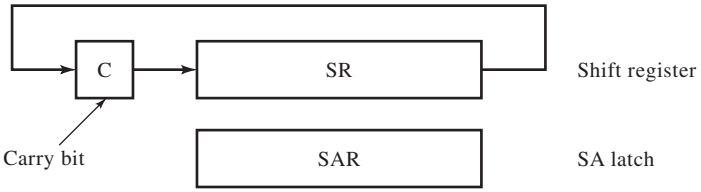
The tracking ADC is useful as a *track-and-hold* (T/H) circuit. The digital output follows the input signal until the clock is gated off or the count clocked into another register. Then the input value, in digital form, is held indefinitely; no analog hold circuit can do this. As for sampling circuits, and also for peak detectors, a capacitor can accurately maintain its charge for only a limited time.



Parallel-feedback ADCs compare DAC voltage to v_X at the comparator input, as shown previously. Current-output DACs require an additional *I*-to-*V* converter stage. In the ADC scheme above, a current-output, bipolar or CMOS, DAC forms a voltage difference with v_X by dropping $i_O \cdot R$ in series with it. The comparator now senses this difference against 0 V. This *current-mode* comparison works with bipolar inputs. The inputs of the comparator must be reversed from voltage-mode comparison, or the complementary current output of the DAC must be used instead.

A third parallel-feedback converter is the *successive-approximation* (SA) converter, a very common conversion technique and the most popular of the parallel-feedback converters. It takes $n + 1$ clock cycles to convert n bits using a bit-wise iterative algorithm. It determines one bit per clock cycle after an initialization cycle.

The parallel-feedback ADC logic block is a *successive-approximation register* (SAR). This register can be realized by an n-bit shift register (SR) and n latches. (A latch is a kind of flop with a level-sensitive clock input. When the clock is asserted [high], its output follows its input. When the clock is unasserted [low], the output remains the value at the falling edge of the clock.) At the start of conversion, the SR bits are cleared, and the MSB is set. The latch feeds this digital midscale value to the DAC. If v_X is larger than midscale, the comparator output is high. When the clock goes low, the MSB is latched. The next clock edge shifts the 1 bit in the SR to the $n - 1$ bit position, and the cycle is repeated. In effect, beginning with the MSB, n decisions are made, each of which narrows the range of possible values for v_X by half. The convergence rate of this procedure is on the order of $\log_2(n)$, and the conversion time is independent of v_X.



The generic μC-based hardware is again used to implement a μC-based SA ADC. The procedure is only slightly more complicated than previous ones but is usually well worth the speed increase. Besides the IN and OUT address locations, the software model is shown above. SAR is a variable that emulates the SAR latch. Variable SR emulates the shift register, which has an additional "carry-bit" stage that is included in the shift loop, as shown. This formulation suggests the efficiency of assembly-language programming because most μCs have a carry bit and a rotate instruction that includes the carry bit (C). Both software variables can be held in μC registers.

In the following procedure, bit-wise logic operations of AND, OR, and NOT (logic negation) are used and are μC instructions. For μCs without a NOT instruction, X is complemented by using the exclusive-OR (EOR or XOR) instruction with binary 1111 . . . (all binary ones, a two's-complement −1, or hexadecimal FFF . . .) and X.

0. **Successive-approximation ADC**
1. Clear SR and SAR: SR ← 0; SAR ← 0.
   Set C to one: C ← 1.
2. Rotate SR right.
3. If $C = 1$, then return.
4. SAR ← SAR OR SR
5. Output SAR to OUT: OUT ← SAR.
6. Input from IN.
7. If IN = 1, then go to 2.

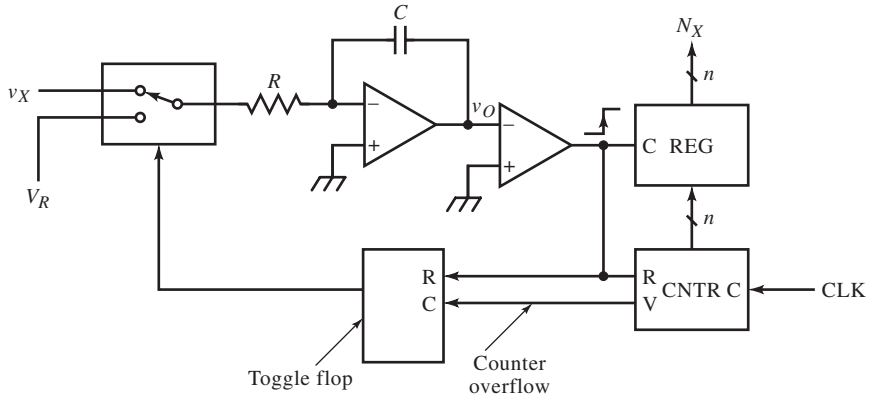
   Else, set SAR to SAR AND $\overline{\text{SR}}$: SAR ← SAR AND (NOT SR).
   (Alternative: SAR ← SAR AND (SR EOR 1111 . . .).
8. Go to 2.

The 1 bit, initially in C, is shifted right, into SR, one bit per iteration. When it gets back to C (step 3 checks this), the procedure is done. Step 4 sets the SR 1 bit in the SAR. If the comparator (IN) is high, v_X is still greater than the SAR value, and this test bit is left set. If IN is low, the set bit made SAR too large, and it is cleared in step 7. Each bit, beginning at the MSB, is tested and then left set or cleared in SAR.

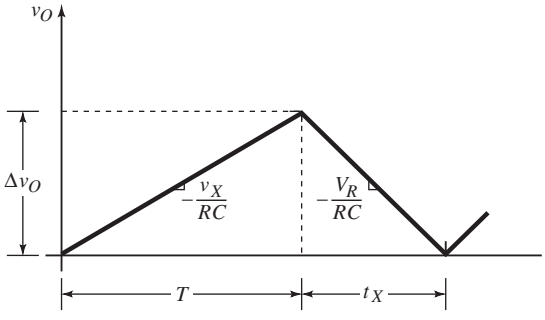
A speed enhancement for SA converters is to increase the clock rate after the first or second bit is determined. These bits have the most range and require the most slew time of the loop hardware. The less-significant bits cause less comparator voltage change and can be determined more quickly, allowing an increased clock frequency at the expense of more digital hardware.

The ramp and SA converters do not function correctly unless v_X is constant during conversion. For dynamic inputs, a sample-and-hold (S/H) circuit must precede the ADC.

INTEGRATING ADCs



A second category of ADC integrates v_X and outputs its average value over the conversion period. The *dual-slope* ADC shown above is an instance. The input to an op-amp integrator is switched between input $v_X < 0$ and positive voltage reference V_R. The integrator output zero-crossing is detected by a comparator, and the count of a free-running counter is clocked as the digitized output. The conversion starts when the counter is reset and v_X is switched into the integrator. The ramp output has a slope of $-v_X/RC$ and ramps up until the counter overflows. For an n-bit counter, this phase lasts 2^n clock cycles or T amount of time. In the second phase, the reference is integrated instead. Because its polarity is opposite that of v_X, the slope changes polarity, as shown in the graph below.



When the integrator output crosses zero, the comparator latches the count. The second phase lasts for t_X time. The converter then begins another cycle.

The change in integrator output voltage, Δv_O, is the same for both phases:

$$\Delta v_O = \frac{v_X}{RC} \cdot T = \frac{V_R}{RC} \cdot t_X \Rightarrow t_X = \left(\frac{v_X}{V_R} \right) \cdot T$$

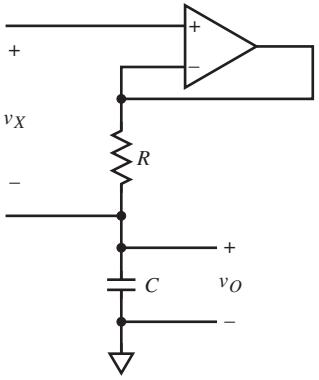
For a constant-frequency clock, the counts relate to the times by

$$N = f_{\text{CLK}} \cdot \Delta t$$

Therefore, the output count is

$$N_X = \left(\frac{v_X}{V_R} \right) \cdot 2^n$$

Dual-slope converter accuracy is not dependent on long-term drift in R, C, or f_{CLK}, only V_R. What this analysis assumes is a perfect op-amp and comparator. Their input offsets and delay times degrade converter accuracy.



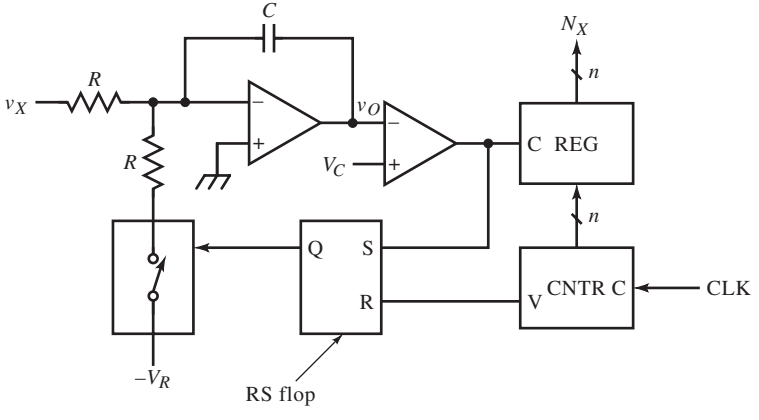
Most actual dual-slope converters correct for offset by introducing a third auto-zero phase before phase 1. In addition, for digital voltmeters (DVMs), a high input impedance is desired, and a buffer amplifier is added before the

integrator. An alternative is the noninverting integrator shown above. If the DVM ground is "floating" (not connected) to the measured source, then the ×1 buffer provides high input impedance as it supplies the charging current for C through R.

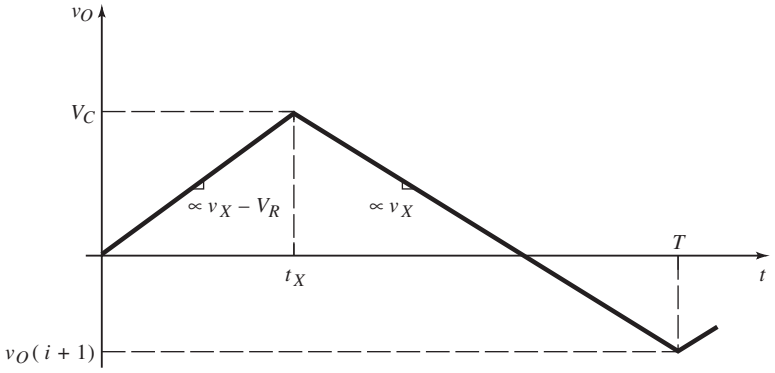
Bipolar inputs require another reference, $-V_R$. Reference selection is determined by the comparator output at the end of phase 1. Another design option for bipolar inputs is to exchange the input terminals by switching. This scheme, however, has difficulty with v_X near zero. Offsets can cause the readings for $+v_X$ and $-v_X$ to have different magnitudes. More significantly, when offsets dominate the input, the converter can integrate with the wrong (shallow) slope. When the reference is integrated, it is of the same polarity, and v_O never crosses zero. To avoid switching in the wrong polarity of reference, hysteresis around zero is sometimes added. But all of this is avoided with two references.

Another input circuit is a V/I converter and a current reference. This eliminates R from the integrator and could also eliminate the op-amp in some designs.

The accuracy of the dual-slope ADC is extended by the *triple-slope* ADC. An additional comparator senses that v_O is approaching 0 V and switches in a smaller reference and another counter. The slope magnitude decreases for this next phase and the time duration is extended. The extra counts contribute additional LSBs.



At somewhat less speed, the simpler *modified dual-slope* converter uses only one switch and integrates the input during both phases. In phase 1, the negative reference $-V_R$ is integrated along with the input. If $V_R > |v_X|$, the integrator output has a positive slope. When it reaches comparator threshold voltage V_C, the reference is switched off, and v_X integrates until the counter overflows at T. The integrator voltage, v_O, at this time depends on v_X. The next conversion cycle thus begins at a different initial v_O.



The conditions for convergence of v_O (and a steady digitized value) are found by solving for $v_O(i)$ where i is the cycle index. For the new cycle,

$$v_O(i+1) = v_O(i) + u \cdot t_X(i) + d \cdot [T - t_X(i)]$$

where the slopes are

$$u = -\frac{v_X - V_R}{RC} = \frac{V_R - v_X}{RC}, \; d = -\frac{v_X}{RC}; \quad v_X, V_R > 0$$

Also, from the *v-t* graph,

$$t_X(i) = \frac{V_C - v_O(i)}{u} = \frac{V_C - v_O(i)}{V_R - v_X} \cdot RC$$

Substituting for t_X in $v_O(i+1)$ gives

$$v_O(i+1) = \left(\frac{d}{u}\right) \cdot v_O(i) + \left[V_C \cdot \left(1 - \frac{d}{u}\right) + dT\right] = a \cdot v_O(i) + b$$

This difference equation is solved by expanding several iterations, beginning with $i = 0$. The resulting recursion equation for $i + 1 = n$ is

$$v_O(n) = a^n \cdot v_O(0) + \left(\frac{1 - a^n}{1 - a}\right) \cdot b, \quad |a| < 1$$

and is attained by using the geometric-series formula

$$\sum_{k=0}^{N-1} z^k = \frac{1 - z^N}{1 - z}, \quad |z| < 1$$

The series converges only when $|z|$ decreases with increasing k. For the converter, the convergence condition is

$$\left|\frac{d}{u}\right| < 1 \Rightarrow -d < u \Rightarrow \frac{v_X}{RC} < \frac{V_R - v_X}{RC} \Rightarrow v_X < \frac{V_R}{2}$$

That is, v_X must not exceed half the reference voltage V_R. Or, in time, $t_X < T/2$. The converged (steady-state) value of v_O can be found by letting n go to infinity in $v_O(n)$ or by setting

$$v_O(i+1) = v_O(i)$$

in $v_O(n+1)$ and solving for v_O:

$$v_O = V_C + \left(\frac{u \cdot d}{u - d}\right) \cdot T$$

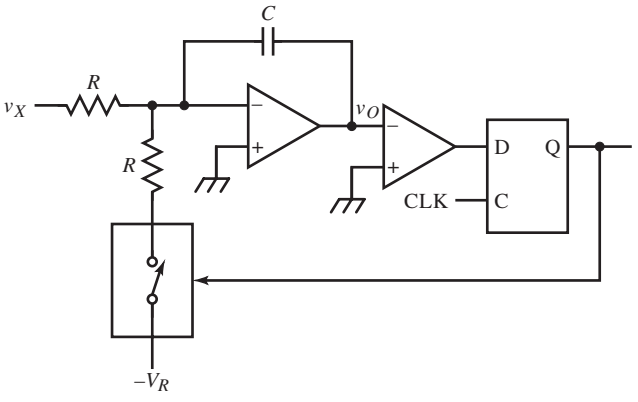
where the second term is always negative, as required for $v_O < V_C$. With v_O, the *steady-state* t_X from $t_X(i)$, is

$$t_X = \left(\frac{-d}{u-d}\right)\cdot T = \left(\frac{v_X}{V_R}\right)\cdot T$$

But this is the same as the previous t_X, and the digital output is expressed by N_X. The modified dual-slope converter has the same transfer characteristic as the dual-slope ADC, though its dynamic response is first-order and takes a few cycles to converge.

In this realization of the modified dual-slope ADC, the Rs must match. V_C need not be accurate – only stable during convergence. Both Rs can be eliminated by driving the integrator with a V/I converter for v_X and replacing R and $-V_R$ with I_R. The switch must then be a current switch. This can be accomplished by letting the flop output divert I_R through a diode. For low leakage, a transistor is used instead. The RS flop consists of two cross-coupled NOR gates. The other two gates in a quad NOR-gate IC implement the clock generator.

Because t_X must be kept less than $T/2$ for stability, the fs t_X is set at $T/4$ by adding two additional bits to the counter (for $n+2$ bits total). This wastes 50% of the available integration time but is easy to implement (by a dual-flop IC) and gives the converter a near-100% overrange capability, an additional half-digit. Besides the register and counter, the total parts count is less than a dozen to implement a three-digit DVM. (A featureless converter such as this is usually called a digital panel meter [DPM].)



Dual-slope ADCs require a large v_O range to achieve precision. An idea that is the digital analog of the virtual ground is realized in the *charge-balancing* (or quantizing or *sigma-delta*, Σ-Δ) ADC, shown above. The circuit topology is very similar to the modified dual-slope ADC, but it works differently. The big circuit difference is that the flop driven by the comparator is clocked, a D-type flop instead of an RS flop.

On a given cycle of the clock, the reference is switched in or out of the integrator to keep v_O near ground. The comparator output gives the sign of the error. In other words, v_O is nulled by discrete-time feedback. The number of clock cycles that the flop was high, N_X, over the total number of conversion counts N, indicates v_X.

The transfer characteristic is calculated by constructing the charge-balance equation for the total charge from v_X and $-V_R$ input to the integrator. For $v_O = 0$, they must be equal, or

$$Q_X = Q_R$$

These charges are the sums of the per-cycle charges:

$$q_X = \left(\frac{v_X}{R}\right) \cdot T_{\text{CLK}}, \quad q_R = \left(\frac{V_R}{R}\right) \cdot T_{\text{CLK}}$$

The total charge of each depends on the number of cycles each is integrated. Then

$$Q_X = q_X \cdot N = \left(\frac{v_X}{R}\right) \cdot N \cdot T_{\text{CLK}}, \quad Q_R = q_R \cdot N_X = \left(\frac{V_R}{R}\right) \cdot N_X \cdot T_{\text{CLK}}$$

Then substituting into Q_X and solving for the output,

$$N_X = \left(\frac{v_X}{V_R}\right) \cdot N = \left(\frac{v_X}{V_R}\right) \cdot 2^n$$

for an n-bit conversion-time counter. This result is, again, the same as for the previous converters.

The charge-balancing circuit is also used as a modulator for serial digital telecommunications (in CODECs) and in audio and speech processing.

An advantage of the integrating ADC is its measurement of the average v_X. By integrating, it has inherent noise rejection and does not need a S/H circuit. The noise rejection capability is quantified by beginning with a constant V_X with sinusoidal noise added:

$$v_X = V_X + V_N \cdot \sin \omega_N t$$

The integrator averages v_X over the conversion period T, so that

$$\text{avg } v_X = \frac{1}{T} \cdot \int_0^T v_X \cdot dt = V_X + \frac{V_N}{\omega_N T} \cdot (1 - \cos \omega_N T)$$

The normal-mode rejection (NMR) of the noise is

$$\text{NMR} = \frac{\text{input noise}}{\text{output noise}} = \frac{V_N}{(V_N / \omega_N T) \cdot (1 - \cos \omega_N T)} = \frac{\omega_N T}{1 - \cos \omega_N T}$$

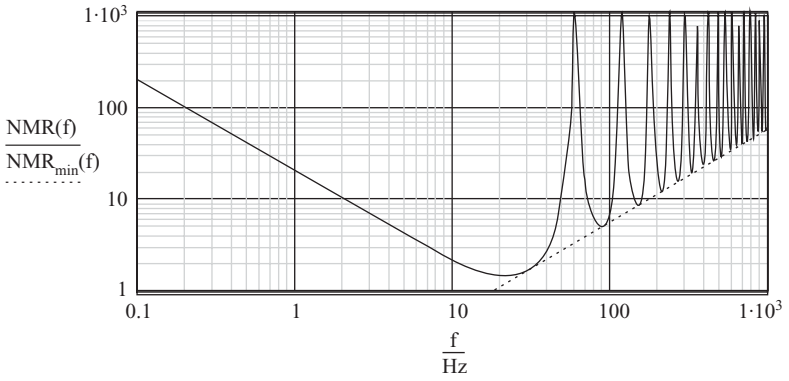
For $\omega_N = 2\pi \cdot f_N$, and $1 - \cos(2x) = 2 \cdot \sin^2 x$, then

$$\text{NMR} = \frac{\pi \cdot f_N \cdot T}{\sin^2 (\pi \cdot f_N \cdot T)}$$

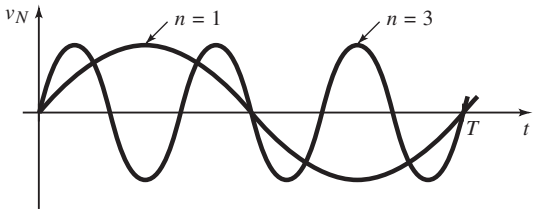
In the decibel scale this is

$$\text{NMR (dB)} = 20 \cdot \log \text{NMR} = 20 \cdot \{\log (\pi \cdot f_N \cdot T) - \log [\sin^2 (\pi \cdot f_N \cdot T)]\}$$

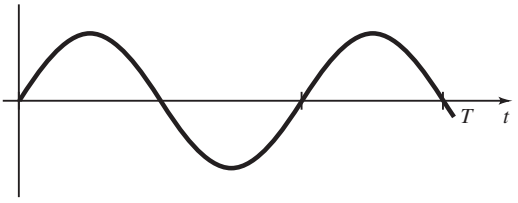
NMR is plotted below on a log-log graph, for $T = 1/60$ Hz, and f_N from 0.1 Hz to 1 kHz.



At $f_N = n/T$, for whole-number n, NMR is infinite. In practice it is typically about 60 dB of rejection. An exact number of noise cycles fit the integration interval T, and the sum of the areas of their positive and negative half-cycles cancel, as in (a) below.



(a)



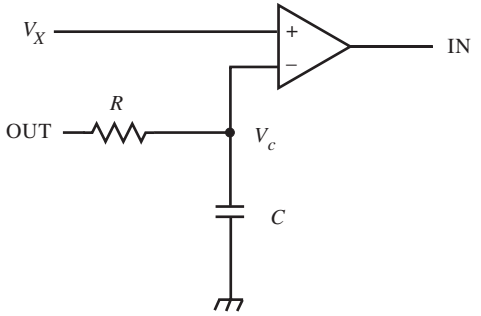
(b)

As f_N varies from n/T, the half-cycles of noise at the ends of the interval are truncated and contribute some fraction of a half-cycle. Figure (b) shows the second worst case after that of $f_N = 0.5/T$, in which an entire extra positive half-cycle is integrated at $f_N = 1.5/T$. For rejection of power-line noise, f_N is often chosen to be a multiple of the power-line frequency.

More significant are the NMR minima of $\pi \cdot f_N \cdot T$. They occur midway between the maxima, at $f_N = 1.5 \cdot n/T$. As f_N increases, according to NMR(dB), the NMR minima increase at 20 dB/dec. At noise frequencies of about n times $1/T$, about n cycles of noise occur during T. The more half-cycles, the less each contributes to the integrated total. Thus, a fraction more of a half-cycle contributes less error the higher f_N is. Note that NMR is the reciprocal of the integrator frequency response, which rolls off at 20 dB/dec with periodic notch filters.

Simple µC-Based Σ-Δ ADCs

Microcontrollers (µCs) often contain a comparator that can implement a precise ADC with the addition of only an external resistor and capacitor. The technique is to implement a charge-balancing or Σ-Δ (or Δ-Σ) ADC. The basic scheme uses a comparator that outputs µC input bit IN and requires one µC output bit, OUT. The circuit is shown below.



In µC software, the ADC routine is best implemented as an interrupt routine, driven by a timer of period t_{INT}, the interrupt period. In the circuit above, the ADC reference voltage is the µC supply ($V_R = V_{CC}$). This assumes that the µC has

CMOS output bits, so that the outputs for negligible current are near the rails:

$$\text{OUT bit CMOS levels: } \begin{cases} 0 \rightarrow 0 \text{ V (ground)} \\ 1 \rightarrow V_{CC} \end{cases}$$

If greater accuracy than V_{CC} is required, instead of driving R directly from OUT, use it to switch accurate analog switches between reference ground (for 0) and an accurate V_R (for 1). If the OUT-bit voltage levels are close enough to the rails, then an accurate V_{CC} can be supplied as the reference.

Σ-Δ *RC* Constraint for *n*-Bit Accuracy

The charge-balance voltage waveform on the capacitor is a constant voltage with a small exponential ripple riding on it, at the frequency of the OUT switching. If this varying voltage becomes too large, the ADC will not be linear enough for n-bit conversion. The larger is the RC time constant, the smaller the ripple. How large must RC be to ensure n bits of linearity? The ripple voltage,

$$\Delta v_C = v_H - v_L \leq V_{LSB} = 2^{-n} \cdot V_R$$

and

$$\frac{\Delta v_C}{v_H} = 1 - \frac{v_L}{v_H} = 1 - e^{-t_{INT}/R \cdot C}$$

where v_H and v_L are the maximum and minimum of v_C. At full scale, $v_H = V_R$ and

$$2^{-n} \geq 1 - e^{-t_{INT}/R \cdot C}$$

or

$$R \cdot C \geq \frac{t_{INT}}{\ln\left(1 - 2^{-n}\right)^{-1}} \cong 2^n \cdot t_{INT}, \quad n >> 1$$

For t_{INT} = 1 ms, and n = 8 bits, then $R \cdot C \geq 256$ ms. For n = 10, $R \cdot C \geq 1.024$ s. The allowable measurement rate is comparable to that of DMMs.

Σ-Δ Algorithm

The ADC algorithm, coded as part of the interrupt routine, sets or clears OUT to keep $v_C = v_X$. In other words, charge balance is maintained on C so that $\Delta q = 0$. This can be expressed using $\Delta q = i \cdot \Delta t$, where $i = v/R$:

$$\frac{V_R - v_X}{R} \cdot N_X = \frac{v_X}{R} \cdot (N - N_X)$$

or

$$\frac{N_X}{N} = \frac{v_X}{V_R}$$

where N is the number of t_{INT} cycles during the measurement. After N intervals, the measurement ends, and the N_X accumulated during this measurement interval is related to v_X by N and V_R:

$$v_X = \left(\frac{V_R}{N}\right) \cdot N_X$$

N is a software parameter and $V_R = V_{CC}$ of the μC. For each interrupt, the following routine is executed:

If IN = 1: OUT ← 1; increment N_X

If IN = 0: OUT ← 0

At the end of the measurement, after N interrupts (or intervals of t_{INT}), then execute:

measured $N_X \leftarrow N_X$

Reset $N_X \leftarrow 0$

Unmatched R_U and R_L

A refinement that can be brought to the minimalist ADC is to account for different resistance values in series with the OUT switches. Let R_U be the series resistance when OUT = 1 (high) and R_L when it is 0 (low). Then

$$\frac{V_R - v_X}{R_U} \cdot N_X = \frac{v_X}{R_L} \cdot (N - N_X)$$

Given the two switch resistance values, the measured voltage, as a fraction of the reference voltage is

$$\frac{v_X}{V_R} = \frac{\left(\frac{R_L}{R_U}\right) \cdot N_X}{N - \left(\frac{R_L}{R_U} - 1\right) \cdot N_X}$$

This equation presents the onerous μC task of division, despite the precalculated constant, R_L/R_U. This refinement is best left for DSPs, which usually facilitate division. As μCs become like DSPs, this improvement becomes feasible to implement.

Auto-Calibration

A more elegant method of producing an accurate measurement without external reference switching can be applied to systems in which multiple channels are multiplexed into the ADC. If two additional MUX inputs are available and the ADC is linear, two-point calibration can be applied. Two reference voltages, which can be 0 V and V_R, are applied to the ADC, resulting in $N_X(0\text{ V}) = N_0$ and $N_X(V_R) = N_R$. A plot of v_X versus N_X will then have two known points on it, corresponding to the known input voltages. The equation for the calibration line is

$$v_X = \left(\frac{V_R - V_{os}}{N_R - N_0}\right) \cdot N_X + V_{os}$$

where the expression in parentheses is the slope of the line. In general, the offset voltage, V_{os}, can be of either polarity, requiring negative N_X. To get around this, two precision resistors forming a divider from V_R can provide instead a known accurate voltage of $\alpha \cdot V_R$, where α is the attenuation ratio of the divider. For this more general case, the equation of the line can be written by equating slope expressions

$$\frac{v_X - \alpha \cdot V_R}{N_X - N_\alpha} = \frac{V_R - \alpha \cdot V_R}{N_R - N_\alpha}$$

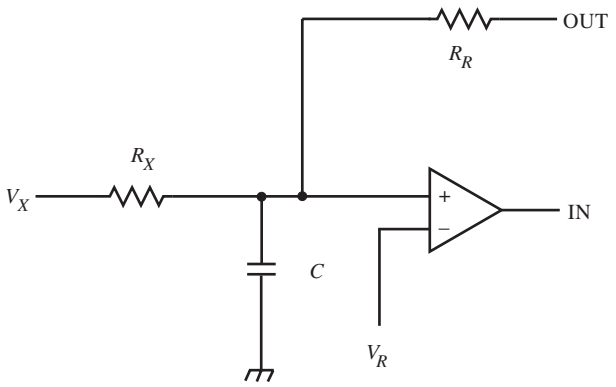
Solving for v_X,

$$v_X = \left[\left(\frac{N_X - N_\alpha}{N_R - N_\alpha} \right) \cdot (1 - \alpha) + \alpha \right] \cdot V_R = \left[m \cdot (1 - \alpha) + \alpha \right] \cdot V_R$$

By making $\alpha = 1/2$, then m must be divided by two, a right-shift instruction. To add ½ to it for rounding, increment m before right-shifting. The resulting number is the fraction of V_R that is v_X.

Inverting Σ-Δ ADC

An inverting Σ-Δ converter uses one additional resistor, as shown below.



The RC time constants must still be much greater than t_{INT}; a low OUT is 0 V, and a high level is $V_{CC} = a \cdot V_R$. Charge balance on the capacitor is maintained by the ADC algorithm, keeping $V_C = V_R$. This results in $\Delta Q = 0$ C:

$$\frac{v_X - V_R}{R_X} \cdot N + \frac{V_{CC} - V_R}{R_R} \cdot N_X = \frac{V_R}{R_R} \cdot (N - N_X)$$

or

$$\frac{v_X}{V_R} = \frac{R_X}{R_R} \cdot \left(1 - a \cdot \frac{N_X}{N}\right) + 1$$

For $R_X = R_R$, and $a = 2$, then

$$\frac{v_X}{V_R} = 2 \cdot \left(1 - \frac{N_X}{N}\right)$$

The following chart summarizes the transfer function.

N_X	v_X
0	$2 \cdot V_R = V_{CC}$
$N/2$	V_R
N	0 V

The interrupt routine for the ADC is given below:

If IN = 1: OUT $\leftarrow$ 0; increment N_X

If IN = 0: OUT $\leftarrow$ 1

At the end of the measurement, after N interrupts (N intervals of t_{INT}), then execute the following routine:

measured $N_X \leftarrow N_X$

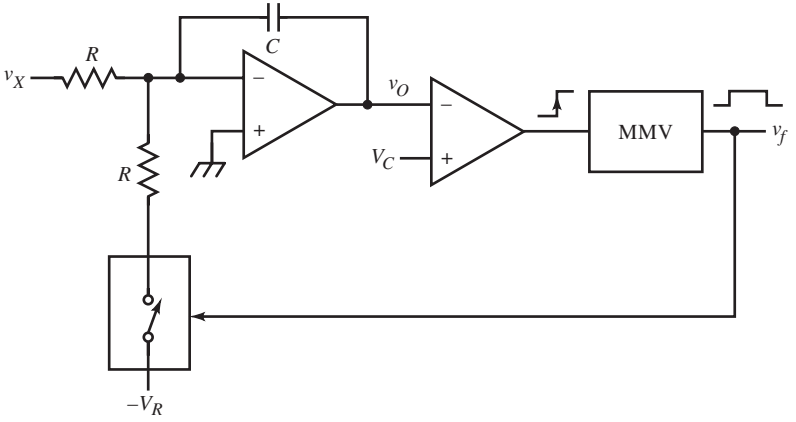
Reset $N_X \leftarrow 0$

These minimal-component ADCs are often adequate for slow, low- to medium-precision, μC-based ADC requirements. Besides few components, other advantages of the Σ-Δ ADC are that it does not need an anti-aliasing filter or S/H circuit preceding its input. Its integrating function reduces noise bandwidth of the measurement. It is an optimal solution for many μC-based applications.

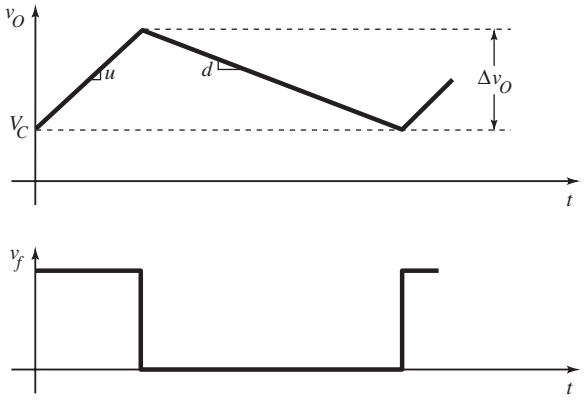
The inverting Σ-Δ ADC input circuit could be extended to have a second-order, cascaded *RC* filter using the same software routine, with a total of four external resistors and two capacitors. This adumbration is left to the imagination of the reader. With sufficiently low t_{INT}, which is achievable on faster μCs and DSPs, high precision can be attained with a medium-performance comparator.

VOLTAGE-TO-FREQUENCY CONVERTERS

A special kind of integrating ADC converts input voltage or current to a pulse frequency. It is a kind of linear voltage-controlled oscillator (VCO) or FM modulator with digital output. The topology of the *voltage-to-frequency* (V/F) converter is similar to previous integrating ADCs.



As with parallel-feedback converters, the topological variations among integrating ADCs is in the logic block driven by the comparator. For the *asynchronous VFC,* a MMV replaces the flop of the charge-balancing converter.



Operation resembles the modified dual-slope ADC. When the integrator output v_O goes below V_C, the comparator output goes high, triggering the MMV and turning on the reference switch. The MMV time-out is t_h, the time that the output pulse v_f is high. During t_h, v_O ramps up with a slope of u. When the MMV times out, the reference is switched out, and $v_X > 0$ causes v_O to ramp down with slope d. Slopes u and d are the same as those for the modified dual-slope ADC. The change in v_O over one cycle is

$$\Delta v_O = u \cdot t_h = \frac{V_R - v_X}{R \cdot C} \cdot t_h$$

From this,

$$t_l = -\frac{\Delta v_O}{d} = \left(\frac{V_R}{v_X} - 1 \right) \cdot t_h$$

The output period is the sum of the half-cycles, or

$$T = t_h + t_l = \left(\frac{V_R}{v_X} - 1\right) \cdot t_h + t_h = \frac{V_R}{v_x} \cdot t_h$$

Finally, the output frequency is

$$f = \frac{1}{T} = \left(\frac{v_X}{V_R}\right) \cdot \frac{1}{t_h}$$

This formula is similar to that of previous integrating ADCs except that it depends on t_h, the MMV time-out duration, instead of a counter overflow period. Because t_h is typically set by an RC circuit, asynchronous VFC accuracy is limited by it. The accuracy also depends on the matching of the R, but the analysis could have been based on an input current i_X and reference current I_R instead. The resistors are implementation-dependent and not fundamental to the operating principle.

The LM331 is an eight-pin VFC IC, shown below in (a). Instead of using an op-amp integrator, it avoids op-amp error by integrating with a shunt RC that is maintained at v_X. The shunt RC voltage v_O must be kept small to avoid nonlinearity. If the exponential waveforms of v_O (b) have a time constant RC that is much larger than t_h, they are approximately linear.

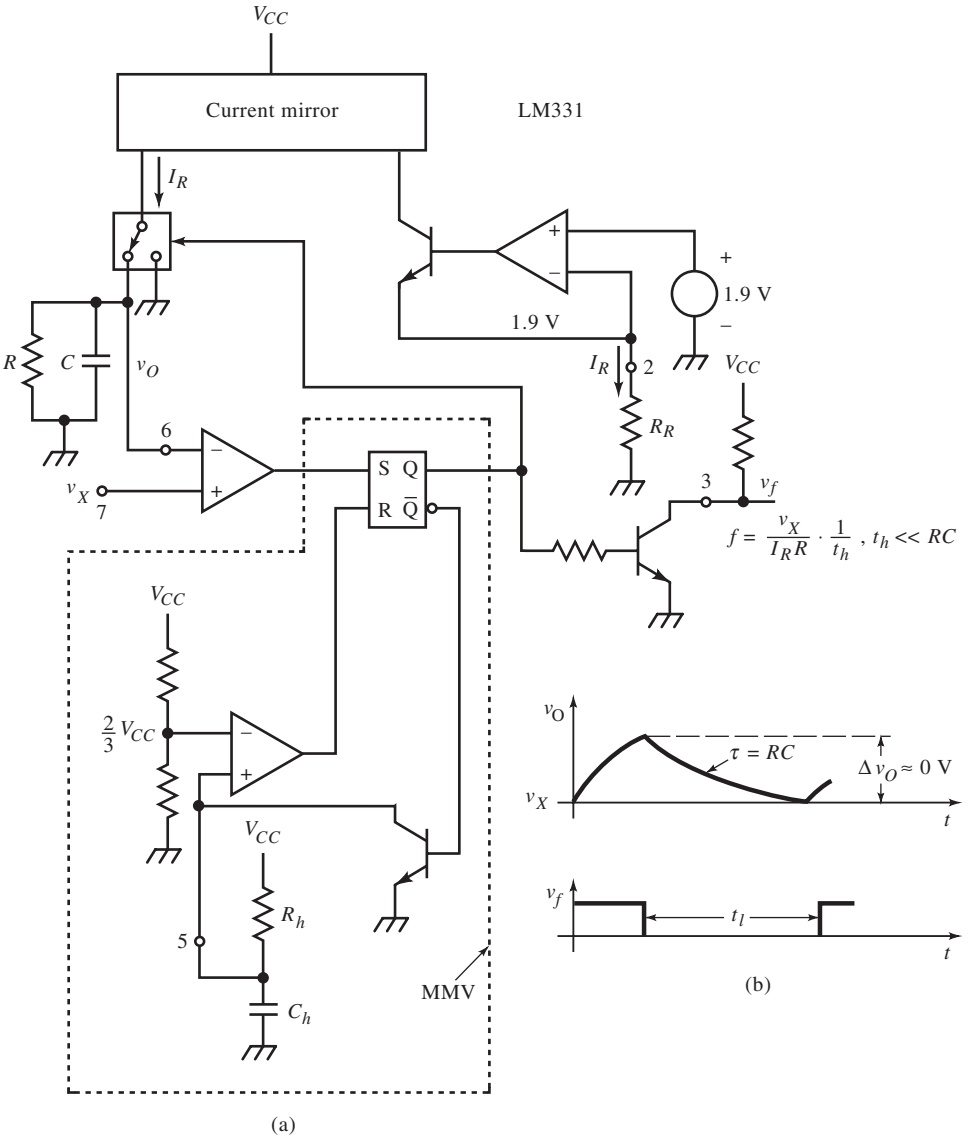
By keeping $v_O \cong v_X$, the LM331 performs charge balancing at v_O. The charge through R over T must be the reference charge during T, or

$$\frac{v_O}{R} \cdot T \cong \frac{v_X}{R} \cdot T = I_R \cdot t_h, \quad t_h << RC$$

Solving for $f = 1/T$,

$$f = \left(\frac{v_X}{R \cdot I_R}\right) \cdot \frac{1}{t_h}, \quad t_h << RC$$

Also, $I_R \cong 1.9\text{V}/R_R$.



This result is also valid for a charge-balancing VFC with a linear integrator but without the constraint. The typical fs frequency is 10 kHz at an output duty-ratio of 50%. Unlike the modified dual-slope ADC, no convergence condition exists, but as t_l approaches zero, the fs frequency asymptotically approaches $1/t_h$

of 20 kHz. When the MMV timing is based on a threshold voltage of $(2/3) \cdot V_{CC}$, then

$$t_h = R_h \cdot C_h \cdot \ln 3 \cong 1.1 \cdot R_h \cdot C_h$$

A more precise analysis, calculated from the exponential v_O, yields a period of

$$T = R \cdot C \cdot \ln\left[\left(\frac{I_R \cdot R}{v_X}\right) \cdot \left(e^{-t_h/R \cdot C} - 1\right) + 1\right]$$

For $t_h << RC$,

$$\left(e^{-t_h/RC} - 1\right) \cong 0$$

Apply the approximations

$$\ln(1 + x) \cong x, \quad x \cong 0; \quad e^x \cong 1 + x, \quad x \cong 0$$

to T:

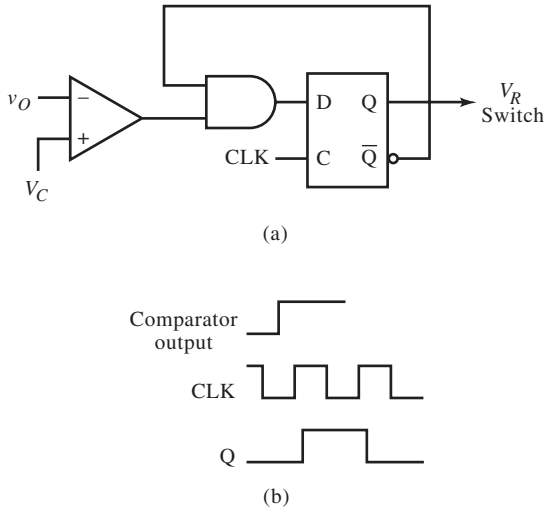
$$T \cong R \cdot C \cdot \left[\left(\frac{I_R \cdot R}{v_X}\right) \cdot \left(\frac{t_h}{R \cdot C}\right)\right] = \left(\frac{I_R \cdot R}{v_x}\right) \cdot t_h, \quad t_h << RC$$

This period is consistent with the asynchronous-VFC *f*. For applications in which a compressed scale for v_X is desired, the nonlinearity of this converter can be advantageous, thereby invoking the adage, "If you can't fix it, feature it."

The VFC is most sensitive to noise at zero scale (zs), when the down-slope *d* is shallowest, causing comparator output jitter among crossings of its threshold and thereby jittering *f*. However, the VFC is an integrating type of ADC because a frequency measurement requires counting v_f over a known period. This counting function is the digital equivalent of integration. The longer the count interval, the more the input is averaged, the greater the precision, and also the slower the conversion rate. For faster conversion at the same precision, the fs frequency

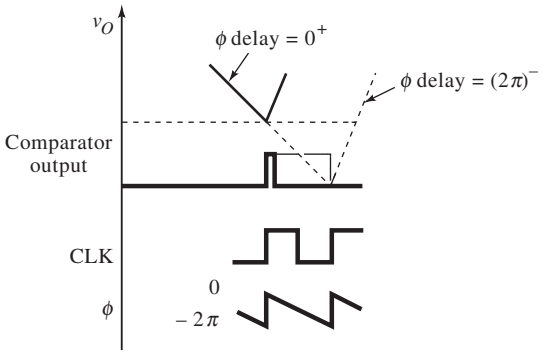
must be increased. By changing count intervals, we can make speed-precision trade-offs without a converter change.

The drift in MMV t_h can be averted by using a digital timer with an accurate clock. Then t_h would, on the average, be accurate. Because the clock is asynchronous with the comparator output, the timer has phase jitter and the time-out varies up to a complete clock cycle. Elaborate schemes have been devised to synchronize a digital counter with an asynchronous trigger to produce an accurate time-out. One simpler scheme combines an analog ramp generator with a counter. The ramp slope is set to V_C/T_{CLK}, where V_C is a comparator threshold. The trigger starts the ramp. It runs up until the active clock edge occurs. The ramp output is held constant until the counter overflows. (More likely, it is a down counter that underflows.) The ramp is restarted. When it crosses V_C, the comparator signals the time-out. The counter counts one less cycle than is required for the time-out because the ramp generator adds a cycle. Its slope error affects the time-out as an error in only one clock period.



Instead of substituting a clocked timer for the MMV, the *synchronous* (or clocked) VFC (figure a) operates similar to the charge-balancing ADC except that the reference is turned on for only one clock cycle at a time. The

comparator output switches the reference only at the active clock edge (b). The D-flop input is gated to enable its output to be high for the cycle, thus generating the output pulse. In commercial synchronous VFCs, the flop output triggers a MMV that sets the output pulse width.



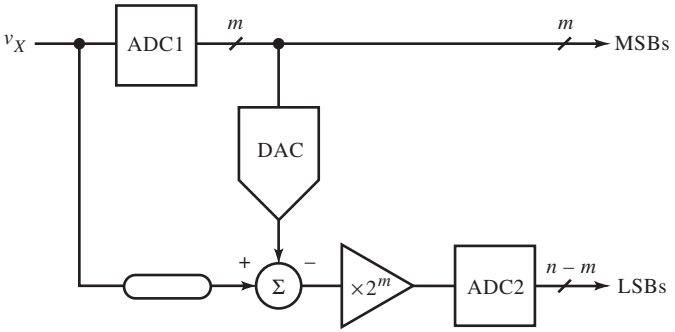
The dual-slope waveform of v_O is synchronous with the clock only at discrete values of v_X. For v_X between these quantum levels, the average level of v_O slowly drifts due to accumulating phase error. The comparator edge drifts relative to the clock, causing the reference on-time to change linearly. This causes the average level of v_O to ramp up or down. When the phase between comparator and clock outputs drifts by a full clock cycle (or 2π radians of phase), the comparator and clock are again in sync; v_O has drifted to a quantum level where the phase error is zero. Comparator and clock edges can coincide, and the output can be indeterminate for some time, causing frequency jitter. A trigger-generator circuit is required for synchronizing edges, for higher performance.

Parallel and Recursive Conversion Techniques

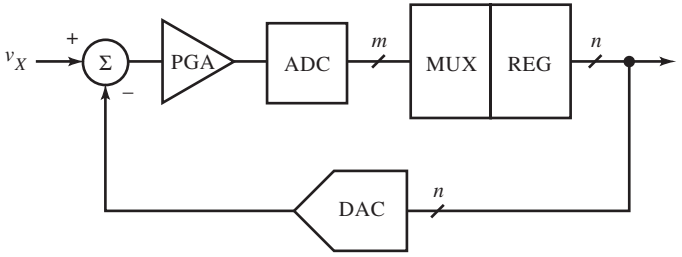
The fastest ADCs are *parallel* or *flash* converters. They have a resistive-divider string of 2^n resistors for an n-bit converter. Each resistor drops V_{LSB} and sets the reference input on one of 2^n latching comparators that drive an encoder. A clock stores the data as 2^n decisions are made simultaneously: $2^n - 1$ for n bits

of conversion plus one for overrange detection. No S/H function is required. Because the circuit complexity grows exponentially with the number of bits, these converters trade off cost, simplicity, and low power for speed. Also with complexity comes a loss of precision because many parts must meet design tolerances. Integration on a single chip helps alleviate the burden of matching parts.

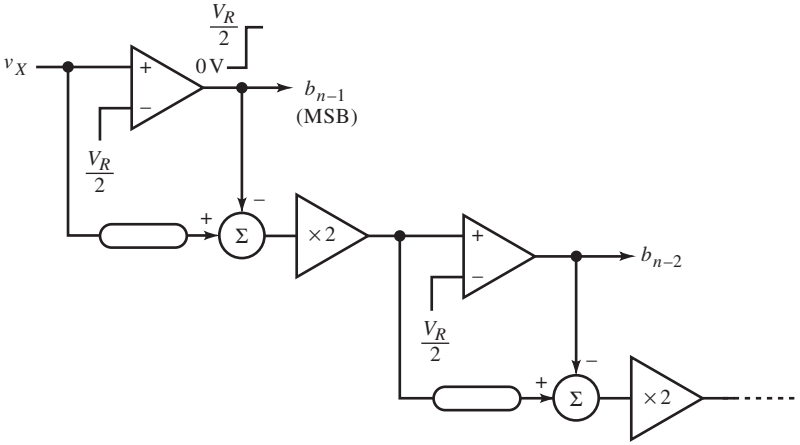
Parallel ADC power is reduced by CMOS implementation. Switched-capacitor comparators designed from CMOS logic inverters reduce power over BJT comparators and can be easily auto-zeroed. But for many applications, the optimum criteria are less complexity and more precision at somewhat reduced speed. This has led to conversion topologies that use m-bit parallel ADCs to digitize $n > m$ bits by iteration.



The *multistage* or *subranging* flash converter has two stages of flash ADCs. The first ADC converts m bits. These MSBs drive a DAC. Its output is subtracted from the input. This remainder or residue is a fraction of one V_{LSB} of the first converter. It is the difference between v_X and the m-bit quantized v_X. The second ADC converts this remainder for the remaining $n - m$ LSBs. If its input range is the same as ADC1, then each V_{LSB} (each step) of ADC1 spans the input range of ADC2, and the remainder must be multiplied by 2^m for correct scaling. Consequently, ADC1 must have n-bit accuracy in the placement of its voltage levels or steps. Also, to avoid misalignment in time, or phase error, subtraction from v_X requires that v_X be delayed by the same amount as the path delay of ADC1 and the DAC.



This idea can be taken further. To save on ADCs and DACs, the *recursive subranging* ADC has a feedback topology instead of the feedforward topology of the multistage flash ADC. In effect, it is a parallel-feedback converter with an m-bit comparator (the ADC) instead of the usual one-bit comparator. It requires n/m iterations or cycles for n-bit conversion. For each iteration, beginning with the m MSBs, the ADC output is stored in m bits of an n-bit output register. The multiplexer (MUX) directs the bits. The PGA gain is increased by 2^m each iteration. This ADC technique requires a hold circuit for v_X.

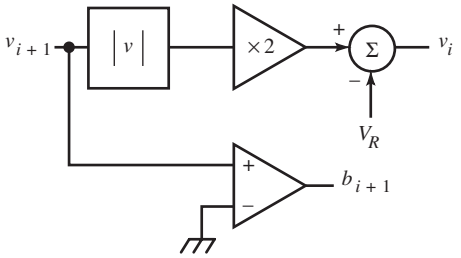


The multistage idea can be taken to its limits by converting one bit per stage. In this *n-stage* flash ADC, each ADC is a comparator designed to have accurate output levels of 0 V and $V_R/2$. Instead of iterating in time, this design iterates

hardware stages. It needs no hold on v_X since v_X ripples through the stages, being processed as it goes, much like a distributed amplifier. Because it is a bit-wise converter, it implements the SA algorithm in space (hardware) instead of in time as the SA ADC does:

$$v_i = 2 \cdot \left(v_{i+1} - b_{i+1} \cdot \left(\frac{V_R}{2} \right) \right)$$

where $v_n = v_X$.



This same idea has been used recursively in John Fluke Co. DVMs, called the *recirculating-remainder* or *cyclic* converter, shown above. It follows a similar recursive equation:

$$v_i = 2 \cdot |v_{i+1}| - V_R$$

where $v_n = v_X$. In the n-stage flash ADC, the remainder passed to the next stage is always positive. Here, the error is bipolar; its sign determines the bit. It is made positive by $|v|$, amplified by two and then subtracts V_R. The block diagram above can be repeated, like the n-stage flash, or a S/H can hold the output for recirculation n times.

The serial bit output is ordered MSB first, but the encoding is in *Gray code*. This code is commonly used in mechanical shaft position encoders because only one bit changes between adjacent states. If the bit outputs are misaligned, an error of only ±1 LSB occurs. Gray-code encoders are used in fast flash

converters for the same reason: Any time-skew among output bits between two successive outputs results in at most 1 LSB of error. Gray code is converted to offset binary by the formula

$$b_i = b_{i+1} \oplus g_i$$

where b_i are output offset-binary bits, g_i are input Gray-code bits, and $\oplus$ is the exclusive-OR logic operation.

Time-Domain Sampling Theory

The explanation of A/D conversion assumes that a voltage at one point in time is converted. For a dynamic input, some means of sampling a voltage at an instant and holding this voltage constant is essential to the conversion process. Even flash converters require that all comparators sense v_X at the same instant. Delays in the latching clock and the inputs among comparators causes this time instant to be instead a time interval t_a, called the *aperture uncertainty* or *aperture jitter*. Besides this, there is delay from the clock edge to when the input is actually sampled, or *aperture delay*.

Aperture jitter limits the maximum sine frequency of v_X that can be digitized. For a frequency f, all comparators must settle within one V_{LSB} or $2^{-n} \cdot V_{fs}$ for n bits. The sine slew-rate is $2\pi \cdot f \cdot V_{fs}$ for the worst case. Then t_a must be less than the time taken to slew 1 LSB, or V_{LSB}; that is,

$$t_a < \frac{V_{\text{LSB}}}{\max(dV/dt)} = \frac{2^{-n} \cdot V_{fs}}{2\pi \cdot f \cdot V_{fs}} = \frac{1}{2^{n+1} \cdot \pi \cdot f}$$

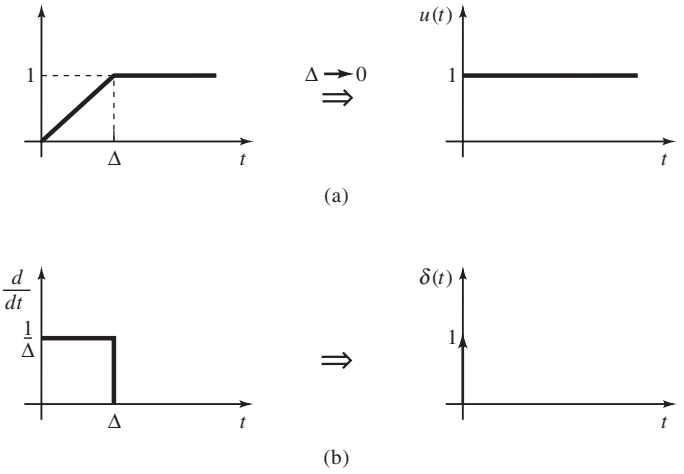
The maximum sine frequency for a given aperture jitter is thus

$$\max \text{ sine } f = \frac{1}{2^{n+1} \cdot \pi \cdot t_a}$$

An eight-bit converter with 100 ps aperture jitter has a maximum digitizing bandwidth of about 6 MHz. By its nature, aperture jitter is a statistical quantity, leading to root mean square (rms) values of the quantities calculated with it.

ADCs that require their input to be held constant over their full conversion period must be preceded by a sampling circuit that then holds the sampled value constant. These are S/H circuits. DACs are inherently digital hold circuits. They hold the sampled output constant and effect a *zero-order hold* (ZOH). A S/H variation is the track-and-hold (T/H) circuit. Its output follows the input in the tracking mode. S/H theory also applies to T/H circuits.

S/H circuits are based on an underlying theory that has general application to discrete-waveform (or sampled-data) systems. Its development in the time domain begins with the step (a) and impulse (b) functions, shown below.



The step function is derived by taking the limit of $v(t)$ in (a) as $\Delta \to 0$. Then for $t < 0$, $u(t) = 0$, but at $t = 0^+$, it is 1. Similarly, the impulse function (in a limiting sense) is derived in (b) as the derivative of $v(t)$. As $\Delta \to 0$, the width of the rectangular pulse goes to zero, but the amplitude goes to infinity. The area remains constant in the limiting process and is the value or "amplitude" of the impulse. In the limit,

$$\lim_{\Delta \to 0} \int_0^{\Delta} \frac{1}{\Delta} \cdot dt = 1$$

The unit impulse has unit value at $t = 0$ and is zero elsewhere so that

$$\int_{-\infty}^{+\infty} \delta(t) \cdot dt = 1$$

Now multiply $\delta(t)$ by a continuous function, $v(t)$, in the integral. Because δ is nonzero only at zero, $v(0)$ effectively weights $\delta(t)$, and

$$\int_{-\infty}^{+\infty} v(t) \cdot \delta(t) \cdot dt = v(0)$$

More generally, if δ is shifted in time by $k \cdot T$, then

$$\int_{-\infty}^{+\infty} v(t) \cdot \delta(t - kT) \cdot dt = v(kT)$$

This can also be expressed as an integral with t as upper bound:

$$\int_{C}^{t} v(\tau) \cdot \delta(\tau - kT) \cdot d\tau = v(kT), \quad t > kT$$

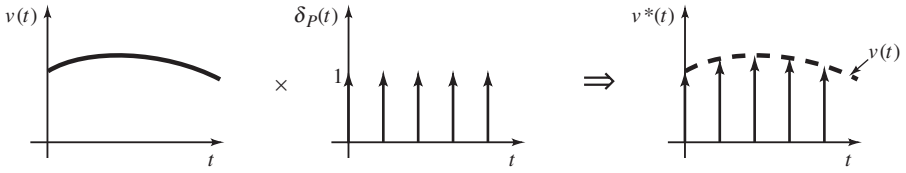
The impulse function is central to sampling theory. A periodic sequence (or "train") of impulses conveniently characterizes the sampling process. A repetitive $\delta(t)$ with period T_s is the sum of an infinite number of time-shifted impulses spaced T_s apart, or

$$\delta_P(t) = \sum_{k=-\infty}^{\infty} \delta(t - kT_s)$$

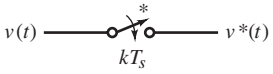
When $v(t)$ is multiplied by $\delta_P(t)$, a sampled form of $v(t)$, or $v^*(t)$, results. ($v(t)$ is real and thus $v^*(t)$ does not designate a complex conjugate.) For $t \geq 0$,

$$v^*(t) = v(t) \cdot \delta_P(t) = \sum_{k=0}^{\infty} v(t) \cdot \delta(t - kT_s) = \sum_{k=0}^{\infty} v(kT_s) \cdot \delta(t - kT_s)$$

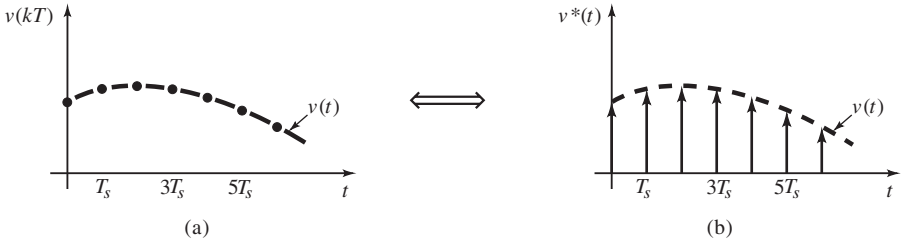
The resulting function is nonzero only where the impulses occur, with values determined by $v(t)$.



The amplitudes of the impulses, though infinite, graphically represent their area values, which are determined by $v(t)$. This is the behavior of the *ideal sampler*, a switch that closes only for an instant.



Two graphic representations of the sampled $v(t)$ are shown below.



From $v(kT)$, the discrete $v(t)$ for $t = kT_s$ in (a) are the integral of each weighted impulse of (b). $v^*(t)$ represents the sampled $v(t)$ as a sum or series, whereas $v(kT_s)$, $k = 0, 1, \ldots$ represents the sampled $v(t)$ as a sequence and $\delta_P(t)$ can also be interpreted as a series from a sequence of unit impulses.

Frequency-Domain Sampling Theory

In the frequency domain, sampling is *impulse modulation*; $v(t)$ amplitude-modulates the impulse train. The Laplace transform and Fourier series reveal another perspective on sampling and lead to important design criteria.

To derive the Laplace transform of $v^*(t)$, begin with $\delta(t)$ and apply

$$\int_{-\infty}^{+\infty} v(t)\cdot\delta(t)\cdot dt = v(0)$$

First,

$$\mathcal{L}\{\delta(t)\} = \int_0^{\infty} \delta(t)\cdot e^{-st}\cdot dt = e^0 = 1$$

Second, the Laplace transform of δ_P is

$$\mathcal{L}\{\delta_P(t)\} = \sum_{k=0}^{\infty} e^{-skT_s} = \Delta_P(s)$$

where T_s is the sampling period. One period of a function $f_1(t)$, such as a single cycle of a sinusoid, can be made repetitive as the series

$$f(t) = f_1(t) + f_1(t-T_s) + f_1(t-2T_s) + \cdots = \sum_{k=0}^{\infty} f_1(t-kT_s)$$

Given $\mathcal{L}\{f_1(t)\} = F_1(s)$, then

$$\mathcal{L}\{f(t)\} = F_1(s) + F_1(s)\cdot e^{-sT_s} + F_1(s)\cdot e^{-s2T_s} + \cdots = \sum_{k=0}^{\infty} F_1(s)\cdot e^{-skT_s}$$

Applying the formula,

$$\sum_{k=0}^{\infty} z^k = \frac{1}{1-z}, \quad |z| < 1$$

results in

$$\mathcal{L}\{f(t)\} = \frac{F_1(s)}{1-e^{-sT_s}}, \quad \left|e^{-sT_s}\right| < 1$$

This is now applied to δ_P:

$$\mathcal{L}\{\delta_P(t)\} = \Delta_P(s) = \frac{1}{1 - e^{-sT_s}}, \quad \left|e^{-sT_s}\right| < 1$$

Finally, the Laplace transform of $v^*(t)$ is

$$\mathcal{L}\{v^*(t)\} = V^*(s) = \sum_{k=0}^{\infty} v(kT_s) \cdot \mathcal{L}\{\delta(t - kT_s)\} = \sum_{k=0}^{\infty} v(kT_s) \cdot e^{-skT_s}$$

This infinite series of exponentials in s makes $V^*(s)$ nonalgebraic and is unwieldy for linear systems analysis. It does, however, resemble the Laplace transform of $v(t)$ for $t = kT_s$. It is simplified by a change of variable,

$$z \equiv e^{sT_s}$$

Solving for s,

$$s = \frac{1}{T_s} \cdot \ln z$$

and substituting for s yields

$$V^*(s)\big|_{s=(1/T_s)\ln z} = \mathcal{Z}\{v(t)\} = \sum_{k=0}^{\infty} v(kT_s) \cdot z^{-k}$$

The operator $\mathcal{Z}$ is the *Z transform*. The Z transform of $v(t)$ is written as $V(z)$, with the understanding that this is not $V(s)$ with z substituted for s. Note that z is a shifting variable; z^{-k} shifts $v(kT)$ by k periods. The Z transform is used in sampled-system analysis the way that the Laplace transform is used with continuous functions. The s-domain offers a continuous view of discrete signals and the z-domain a discrete view of continuous signals.

Now $v^*(t)$ is expressed using the Fourier series. Repetitive $v(t)$ with frequency ω_s can be expressed as the sum of sinusoids at integer multiple frequencies (or *harmonics*) of ω_s:

$$v(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cdot \cos n\omega_s t + \sum_{n=1}^{\infty} b_n \cdot \sin n\omega_s t = \sum_{n=-\infty}^{\infty} c_n \cdot e^{jn\omega_s t}$$

↑ static term ↑ even harmonics ↑ odd harmonics

where

$$a_n = \frac{2}{T_s} \cdot \int_{-T_s/2}^{T_s/2} v(t) \cdot \cos(n\omega_s t) \cdot dt, \quad b_n = \frac{2}{T_s} \cdot \int_{-T_s/2}^{T_s/2} v(t) \cdot \sin(n\omega_s t) \cdot dt$$

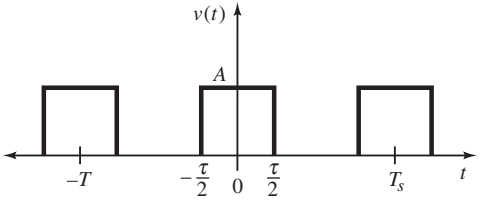
and for the exponential Fourier series,

$$c_n = \frac{1}{T_s} \cdot \int_{-T_s/2}^{T_s/2} v(t) \cdot e^{-jn\omega_s t} \cdot dt$$

The two representations are equivalent and are related by

$$\|c_n\| = \frac{1}{2} \cdot \sqrt{a_n^2 + b_n^2}, \quad \vartheta_n = \tan^{-1}\left\{\frac{b_n}{a_n}\right\}$$

Even functions of time have no sine terms; odd functions have no cosine terms. Some $v(t)$ can be made odd by subtracting an average offset. The odd function is then transformed and the offset is added as a constant term.



In actual samplers, the sampling waveform is an approximation to an impulse train. It has finite amplitude and time duration. The effect this has on sampling can be found by assuming the sampling waveform to be a pulse train with ampli-

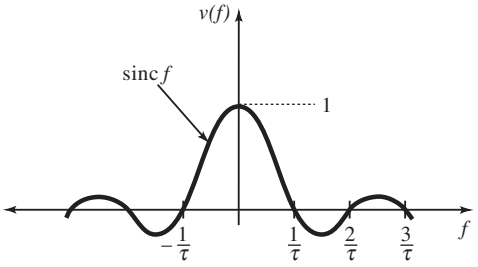
tude A and pulse width τ. Let $\omega_n = n \cdot \omega_s$. As $v(t)$ is centered around $t = 0$, it is an even function, and

$$a_n = \frac{2A}{T_s} \cdot \int_{-\tau/2}^{\tau/2} \cos \omega_n t \cdot dt = \frac{2 \cdot A \cdot \tau}{T_s} \cdot \frac{\sin(\omega_n \tau/2)}{\omega_n \tau/2}, \; n \text{ even}$$
$$a_n = 0, \; n \text{ odd}$$

where the discrete a_n have the continuous *envelope* of the form

$$\operatorname{sinc} x \equiv \frac{\sin x}{x}$$

It is shown below for $x = \omega_n \tau/2 = n \cdot \pi \cdot (\tau/T_s)$. Instead of an impulse, $\operatorname{sinc}(x)$ is the result of finite-width sampling pulses.



As $\tau \to 0$, the pulse train approaches an impulse train. The separation of $a_n(\omega)$ decreases in frequency. If instead we let T_s increase, then the effect is the same; harmonic frequency separation decreases. As $T_s \to \infty$, the a_n merge into a continuous sinc function with a continuous frequency spectrum:

$$\lim_{T_s \to \infty} \Delta\omega_n = \lim_{T_s \to \infty} \left(\frac{2\pi \cdot (n+1)}{T_s} - \frac{2\pi \cdot n}{T_s} \right) = \lim_{T_s \to \infty} \frac{2\pi}{T_s} = 0$$

As $T_s \to \infty$, the function becomes aperiodic, and the Fourier series becomes the *Fourier transform*:

$$\mathcal{F}\{v(t)\} = V(j\omega) \equiv \int_{-\infty}^{+\infty} v(t) \cdot e^{-j\omega t} \cdot dt$$

In the limit, the Fourier series of $v(t)$ undergoes these changes:

$$\sum_{n=-\infty}^{\infty} \rightarrow \int_{-\infty}^{+\infty}, \quad c_n \rightarrow V(j\omega), \quad n\omega_s \rightarrow \omega, \quad \Delta t \rightarrow dt$$

Except for the lower limit of integration, the Fourier transform is a special case of the Laplace transform when $s = j\omega$. The unit step and impulse functions have no Fourier series, but they have Fourier transforms.

As T_s increases (or τ decreases), the sinc response broadens until, in the limit, it is constant over all frequencies. Thus, the frequency response of an impulse is independent of frequency, as is the Laplace transform of $\delta(t)$.

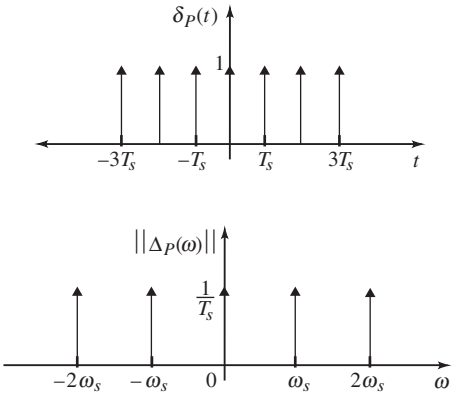
The frequency spectrum for δ_P is

$$c_n = \frac{1}{T_s} \cdot \int_{-T_s/2}^{T_s/2} \left(\sum_{k=-\infty}^{\infty} \delta(t - kT_s) \right) \cdot e^{-jn\omega_s t} \cdot dt = \frac{1}{T_s}$$

This spectrum is also flat for all frequencies, with a constant amplitude of $1/T_s$. It differs from the spectrum for $\delta(t)$ in that it is discrete. The Fourier series of the impulse train is

$$\delta_P(t) = \sum_{k=-\infty}^{\infty} \delta(t - kT_s) = \frac{1}{T_s} \cdot \sum_{n=-\infty}^{\infty} e^{jn\omega_s t}, \quad \omega_s = \frac{2\pi}{T_s}$$

The waveforms of δ_P in both the time and frequency domains are shown below.



This representation of δ_P leads to a different expression for $\mathcal{L}\{v^*(t)\}$ from that derived previously:

$$\mathcal{L}\{v^*(t)\} = \int_{-\infty}^{\infty} v(t) \cdot \left(\frac{1}{T_s} \sum_{n=-\infty}^{\infty} e^{jn\omega_s t} \right) \cdot e^{-st} \cdot dt$$

The index n is independent of t allowing the summation to be removed from the integral:

$$\frac{1}{T_s} \cdot \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} v(t) \cdot e^{jn\omega_s t} \cdot e^{-st} \cdot dt = \frac{1}{T_s} \cdot \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} v(t) \cdot e^{-(s-jn\omega_s)t} \cdot dt$$

The resulting integral is the Laplace transform, $V(s - jn\omega_s)$. Thus,

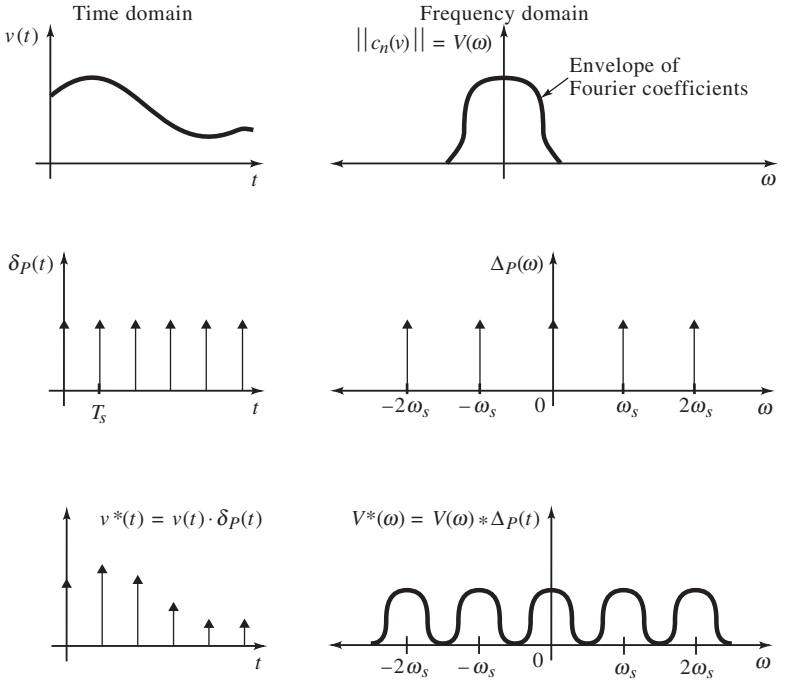
$$V^*(s) = \frac{1}{T_s} \cdot \sum_{n=-\infty}^{\infty} V(s - jn\omega_s)$$

This expression for V^* has a geometric interpretation in the s-domain. The transform of $V(s)$ is periodic in ω_s so that

$$V^*(s) = V^*(s - j\omega_s)$$

$V(s)$ repeats along the $j\omega$-axis at intervals of $j\omega_s$.

Previously, $\delta_P(t)$ was expressed as a series of complex sinusoids with amplitude $1/T_s$ and frequencies of $n \cdot \omega_s$. The frequency spectrum of $v(t)$ is convolved (or *heterodyned*) in the frequency domain with the spectrum of $\delta_P(t)$, as shown below.



Multiplication in one domain corresponds to convolution in the other. The sine and cosine terms in $v(t)$ multiply by the terms of δ_P to produce sum and difference frequencies according to the trigonometric formulas

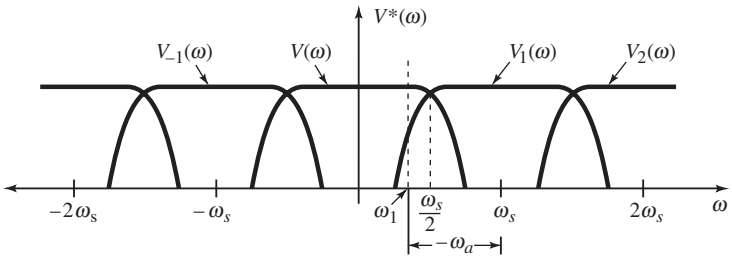
$$\cos\alpha \cdot \cos\beta = \frac{1}{2}\cos(\alpha - \beta) + \frac{1}{2}\cos(\alpha + \beta)$$

$$\cos\alpha \cdot \sin\beta = \frac{1}{2}\sin(\alpha + \beta) - \frac{1}{2}\sin(\alpha - \beta)$$

The frequency-domain plots are the magnitude envelopes of the complex Fourier coefficients, the amplitudes of the harmonics. For $V^*(j\omega)$, the spectrum of $v(t)$ is centered around harmonics of ω_s. Thus, the effect of sampling is to generate frequency-shifted copies (or bands) of $V(j\omega)$ centered around harmonics of ω_s.

THE SAMPLING THEOREM (NYQUIST CRITERION)

The *sampling theorem* gives a criterion for recovery of $v(t)$ from $v^*(t)$. If ω_s is greater than twice the highest frequency in $V(j\omega)$, then the frequency-shifted bands of $V(j\omega)$ do not overlap and can be separated by filtering.



Alias frequency = $\omega_a = \omega_1 - \omega_s$

The *Nyquist criterion* for recoverability of the original continuous signal is

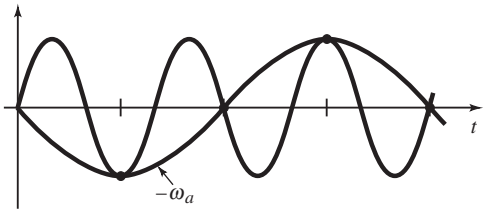
$$\omega_s > 2 \cdot \omega_h$$

where ω_h is the highest frequency component of $V(j\omega)$. The original signal is recoverable from its sampled form when the highest frequency component is less than the *Nyquist frequency*, $\omega_s/2$. In the plot above, the band $V_1(j\omega)$ is a replica of $V(j\omega)$ centered at ω_s. It has frequency components below ω_s that overlap with the positive frequency components of $V(j\omega)$. These are negative frequencies in $V(\omega)$ shifted up in frequency by ω_s.

The significance of negative frequency components in $V(j\omega)$ is that they are inverted (180° phase-shifted) from their corresponding positive counterparts. The magnitude of $V(j\omega)$ is symmetric around $\omega = 0$; it is an even function and $V(-j\omega) = V(j\omega)$. The phase, however, is an odd function and is negative for $\omega < 0$; for negative n, the angle of c_n is $\vartheta = -n \cdot \omega_s \cdot t$. Then $\vartheta(-n) = -\vartheta(n)$.

In the plot, $V(j\omega)$ and $V_1(j\omega)$ are symmetrical around the Nyquist frequency. In effect, V has been folded over at $\omega_s/2$. The larger ω_h is, the further back toward lower frequencies the folding extends. These folded frequency

components from V_1 are *alias* frequencies in $v^*(t)$ and have a frequency of ω_a relative to ω_s.



The significance of an alias frequency in the time domain is that a sequence of samples has more than one frequency interpretation. In the figure above, $V(j\omega)$ has one frequency component at $\omega = (3/4) \cdot \omega_s$. The samples also fit a sinusoid of $\omega = -(1/4) \cdot \omega_s$, an alias frequency within the band of $V(j\omega)$. The alias sinusoid is inverted relative to that of V_1 because its frequency is negative.

More generally, if ω_1 of $V(j\omega)$ is sampled at ω_s, then from the first plot,

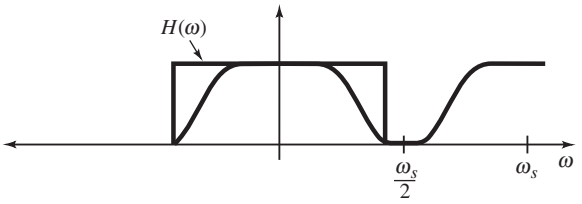
$$\omega_1 = \omega_s - (-\omega_a) = \omega_s + \omega_a$$

and

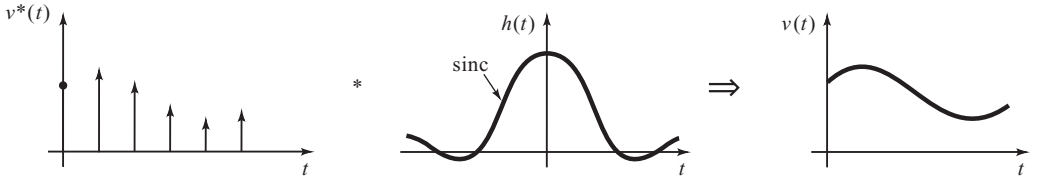
$$\text{alias frequency} = \omega_a = \omega_1 - \omega_s$$

In the plot above, sinusoids of both ω_1 of V and ω_a of V_1 fit the sample points. The discrete samples of $v(t)$ are too few per cycle to eliminate ω_a and $v(t)$ is *undersampled*. The sampling theorem requires more than two samples per cycle for recovery of $v(t)$. Such a $v(t)$ is *oversampled*.

Recovery of $V(j\omega)$ from $V^*(j\omega)$ for oversampled signals is achieved by a low-pass filter (LPF) that passes only $V(j\omega)$. The ideal filter magnitude, $\|H(j\omega)\|$, is shown below.



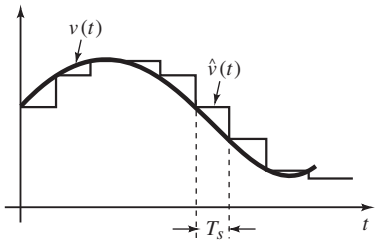
It has an immediate cutoff just above ω_h. The ideal maximum-bandwidth filter has a cutoff at the Nyquist frequency.



In the time domain, this filter function transforms into a sinc function. Nonzero sinc values extend to $t = -\infty$, resulting in a noncausal function that can only be approximated by physical (thus causal) circuits. The pulse shape of the ideal LPF transforms into a sinc function in t just as a pulse in the time domain does in ω. $H(j\omega)$ is multiplied by $V^*(j\omega)$ in ω to recover $V(j\omega)$. In t, $h(t)$ is convolved with $v^*(t)$ to produce $v(t)$. For bandlimited $v(t)$,

$$v(t) = \sum_{k=-\infty}^{\infty} v(kT_s) \cdot \text{sinc}\left(\frac{\omega_s}{2} \cdot (t - kT_s)\right), \quad -\frac{\omega_s}{2} < \omega < \frac{\omega_s}{2}$$

The sinc function operates as an interpolator, filling in the missing values of $v(t)$.



The final derivation of general usefulness is the spectrum of a zero-order hold. This is the frequency response of a S/H circuit. In the time domain, this is a voltage step turned off T_s later:

$$\text{ZOH: } v(t)\cdot u(t) - v(t-T_s)\cdot u(t-T_s)$$

In the *s*-domain, a ZOH can be regarded as an integrator of weighted impulses, producing $\hat{v}(t)$, as shown above. This is the typical output waveform from a S/H or DAC.

The integrated waveform is periodic at the sampling rate. An integrator in s is $1/s$. When it is normalized to be unitless, then it is $1/s\cdot T_s$. A periodic integrator is constructed by integrating for T_s. The Laplace transform of this expression is $H_0(s)$. The normalized ZOH transfer function is thus

$$\text{ZOH: } H_0(s) = \frac{1}{s\cdot T_s} - \frac{1}{s\cdot T_s}\cdot e^{-sT_s} = \frac{1-e^{-sT_s}}{s\cdot T_s}$$

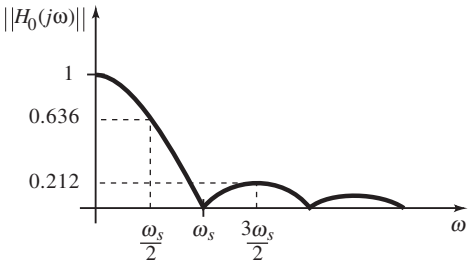
The frequency response of $H_0(s)$ is found by letting $s = j\omega$. Then

$$H_0(j\omega) = \frac{1-e^{-j\omega\cdot T_s}}{j\omega\cdot T_s} = \text{sinc}\left(\frac{\omega\cdot T_s}{2}\right)\cdot e^{-j\omega\cdot T_s/2}$$

The magnitude and phase are

$$\|H_0(j\omega)\| = \left|\text{sinc}\left(\frac{\omega\cdot T_s}{2}\right)\right|, \quad \angle H_0(j\omega) = \frac{-\omega\cdot T_S}{2}$$

Once again, the sinc function appears. The magnitude plot of the frequency response is shown below.



The phase response is linear and only time-shifts the output. The phase delay can be seen in the time plot by noting that a best-fit of $v(t)$ to $\hat{v}(t)$ requires $v(t)$ to be shifted to the right (delayed in time) by half a step, or by $-T_s/2$, in agreement with $\angle H_0(j\omega)$. Ideal recovery of $v(t)$ from $\hat{v}(t)$ requires an inverse sinc filter, or *sinc compensator*. This compensator can be implemented in either digital or analog form. It is digital if it precedes a DAC or follows an ADC and analog if it follows a DAC or precedes an ADC.

SAMPLING CIRCUITS

Sample-and-hold or track-and-hold circuits are switched between the sample or track state and hold state by a digital control line. Ideally, the input voltage at the instant of switching to HOLD is retained as a constant at the output of the S/H. T/Hs are similar to S/Hs; in the non-held state, the output follows the input. In a S/H this is not necessarily so, though most S/Hs are actually T/Hs. The sampling impulse of sampling theory corresponds to the active edge of the HOLD signal.

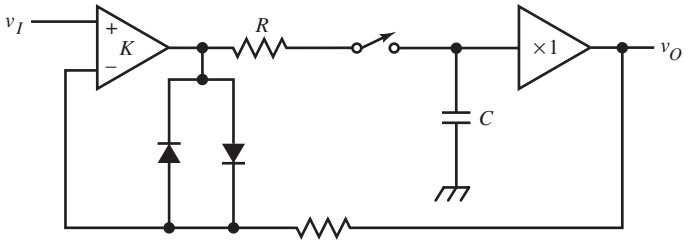
The speed of a S/H is determined by the *acquisition time*, the time from when sampling or tracking of the input begins to when a settled, held output is available. This time has two terms. The first is the time from when tracking begins to the time when the hold capacitor follows the input waveform. A large initial difference between v_I and v_C requires slewing time before tracking is accurate. The second term is the setting time at v_C when the hold state begins. In addition to acquisition time, aperture delay and jitter also apply to S/H circuits.

Several errors are associated with S/H circuits, and their design considerations are closely related to those of peak detectors. Errors occur in the sampling process or in the hold state. The first are dynamic sampling errors. *Digital delay* causes the effective sampling instant to be delayed. For a rising input, this translates into a voltage error of

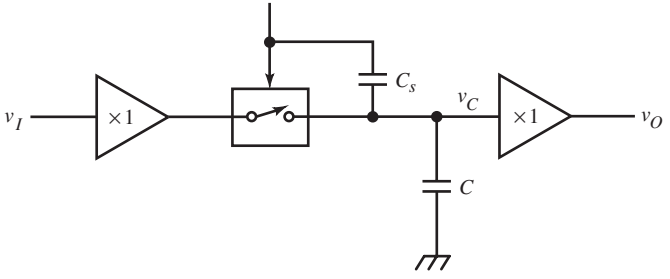
$$v_\varepsilon = \left(\frac{dv_I}{dt}\right) \cdot t_d$$

where the digital delay time t_d is multiplied by the waveform slew-rate. The second cause of error is *analog advance*. If the input is delayed instead, an

effective negative delay occurs in sampling since the waveform lags behind where it should be when sampling occurs. A rising input waveform is below where it should be and a negative error occurs. It is equivalent to sampling the waveform in advance of the actual sampling instant.

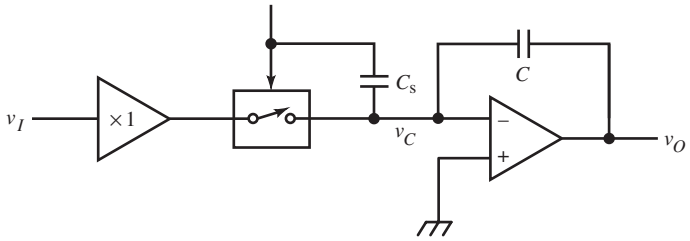


The dominant cause is a voltage lag on the hold capacitor. Its charging always lags somewhat behind the source. This is largely due to charging-source resistance R in the S/H circuit shown above. By closing the loop with an op-amp input, we reduce the charging time constant RC by $K + 1$. The diodes around the op-amp keep its output from saturating when in the hold state. Waveform advance is the major cause of delay error and is compensated by delaying the sampling command.



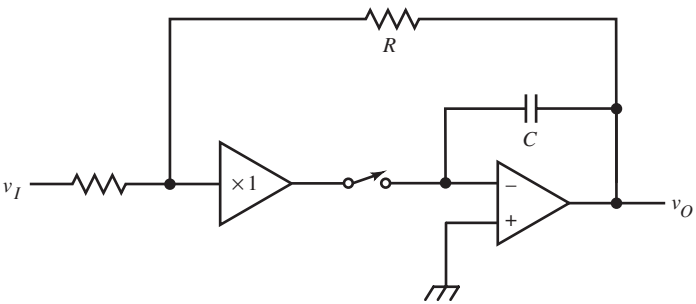
A third dynamic error is due to stray capacitance C_s between the hold capacitor C and the sampling command line. When this line switches, it causes charge to flow through C_s into C. If the capacitor voltage v_C is plotted against a range of constant-voltage inputs, the plot is linear. Its slope represents a hold gain. As

the input voltage v_I increases, the step of extra voltage on C grows in size because the voltage between the hold line v_H and v_C varies linearly with v_I. As the difference between the sample level, V_s, of v_H, and v_C increases, C_s is charged more, and this charge is transferred to C when v_C changes to the hold state. The hold step, or *pedestal*, thus increases with v_I.

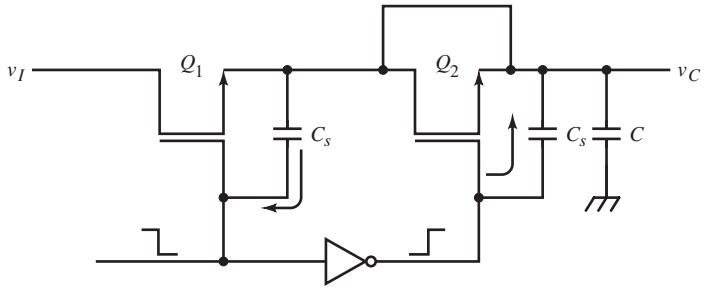


A circuit that avoids this problem is shown above. The hold capacitor is the feedback C of the op-amp. The op-amp isolates v_C from the switch node by holding it at virtual ground. Then the voltage across C_s is independent of v_I and the same amount of charge is transferred to C on switching. The charge on C_s is $C_s \cdot V_s$. The hold gain varies somewhat as C_s varies with voltage as do semiconductor junctions.

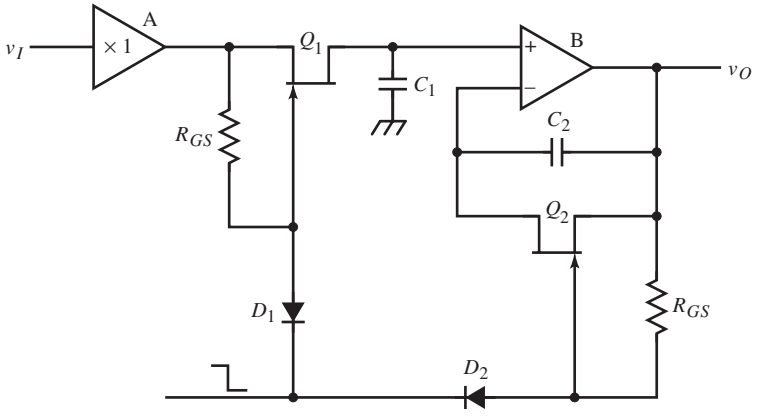
The hold capacitor dielectric absorption must be low to avoid recovery effects during the hold state. Its leakage causes static sampling error during the hold state. Any other leakage paths for capacitor charge contribute to leakage error. The buffer amplifier and sample switch must be low in leakage. A leakage compensator is shown below.



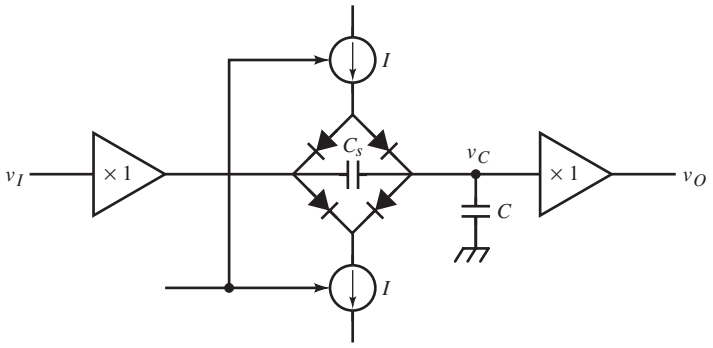
The leakage decoupler R has the same function as in peak detectors. It keeps the voltage across the switch near zero, thus minimizing leakage through it.



Another hold-step compensator places another switch similar to Q_1 in series with it. This additional switch is shorted, but its C_s (C_{GS} for a MOSFET) connects to the same node. It is driven with an opposite polarity edge so that its stray charge cancels that of Q_1.



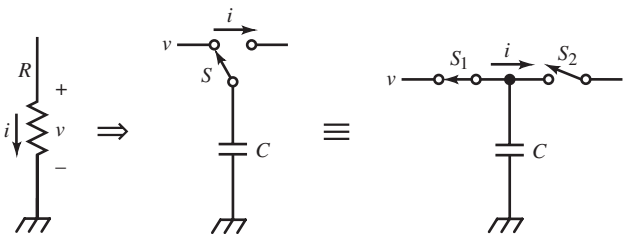
In this S/H, a JFET switch Q_1 passes the waveform through its channel, connected to the buffer A. When the control line goes low, Q_1 cuts off. D_1 conducts a small amount of current through R_{GS} to keep the gate reverse-biased. At the same time, Q_2 is also cut off by a similar circuit. The capacitor C_2, equal to C_1, is a bias-current compensator for the op-amp (as with a previous peak detector). As the hold capacitor C_1 charges with I_B, so does C_2. The differential voltage is canceled at the output. Of course, offset current is not compensated.



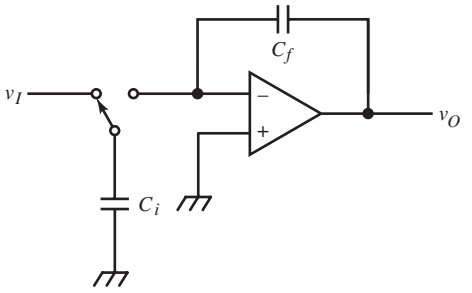
Finally, very fast S/Hs have the additional error of signal leakage through shunt switch capacitance C_s during hold. This circuit uses a diode *sampling bridge*. The bridge has two C_ss in series for each leg of the bridge, or an equivalent of one C_s from input to output. It is current-switched for speed. To reduce additional input-to-output bridge capacitance, the diodes are fed through a metal sheet which functions as a Faraday shield. Sampling bridges of this kind have commonly been used in sampling oscilloscope front ends. The practical limitation in their switching time is often the switching speed of their drivers.

SWITCHED-CAPACITOR CIRCUITS

Switched-capacitor circuits replace resistors with capacitors and switches. In ICs, *diffusion resistors*, made by connecting to the ends of a diffused area, are not optimal since their values are hard to control and they have large areas (their relative values are much better; they match well). Large-value resistors take up so much area that they are often impractical. When accuracy is not important, a kind of resistor made of a thin layer of, say, *n* material between two *p* layers – a *pinch resistor* – can be made large but with a ±20% accuracy. NiCr (nichrome) resistors are very good but costlier to make and trim.



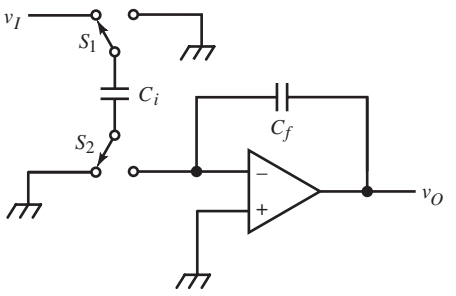
This limitation makes switched-capacitor resistors an attractive alternative. The equivalent resistance is shown above. It is a single-pole, double-throw (SPDT) switch S and a capacitor. The SPDT switch is equivalent to two single-pole, single-throw (SPST) switches, synchronized as shown. When S is switched to the input, it charges to the input voltage v with a charge of $C \cdot v$. When it switches to an output held at ground, it delivers this charge. The output is typically the virtual ground of an op-amp.



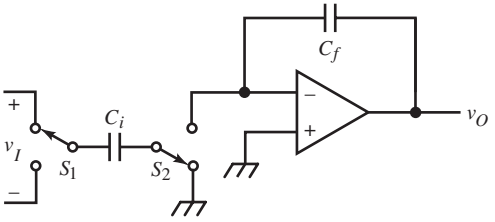
If the switching rate is f_s, then the charge delivered per unit time, or current, is

$$i = C \cdot f_s \cdot v \Rightarrow r = \frac{1}{C \cdot f_s} = \frac{T_s}{C}$$

The equivalent resistance follows directly and is subject to the Nyquist criterion due to switching. The bandwidths of switched-capacitor circuits must be well within the Nyquist frequency for accurate equivalence.



This switching scheme inverts v_I. C_i charges with switches in the position shown. When switched, charge flows out of the op-amp input to ground. In effect, the two-switch circuit is a negative R.



With drive to the grounded side of S_1 from the input instead, the two terminal voltages of v_I subtract upon switching. The differential voltage v_I determines the charges.

CLOSURE

The world of digital electronics merges with analog electronics in digitizing and sampling circuits, but the merged areas – mainly ADCs, DACs, and switched-capacitor and sampling circuits – do not involve logic design. Instead, the underlying theory is an extension of that for continuous functions. The mathematics is similar; difference equations replace differential equations. Sampled circuits also include commutating and switched-capacitor filters and digital signal processing, but the full story, including dithering, FFTs, DSP filters, and windowing, is left for other books.

References

Anderson, T. and Trump, B. (1984, September 6). Clocked V-F Converter Tightens Accuracy and Raises Stability. *Electronic Design*, pp. 235–244.

Angelo, E. J., Jr. (1969). *Electronics: BJTs, FETs, and Microcircuits*. New York: McGraw-Hill, Ch. 15.

Anonymous (1981, April 1). Voltage-Reference Basics. *EDN*, pp. 64–65.

Anonymous (1988, April 14). What's a Delta-Sigma a-d Converter? *Electronic Design*, p. 50 (boxed section).

Anonymous (1989, February 23). What's a Recursive Subranging ADC? *Electronic Design*, p. 122 (boxed section).

Baedtke, R. Triangle Waveform Generator Having a Loop Delay Compensation Network. Tektronix patent.

Bello, V. G. (1972, March). Design of a Diode Function Generator Using the Diode Equation and Iteration. *IEEE Transactions on Circuit Theory*, pp. 213–214.

Bray, D. (1983). The Use of Bipolar Semiconductor Junctions in Linear Circuit Design. Interdesign Monochip Application Note APN-33.

Brokaw, A. Paul (1975). *The Best of Analog Dialogue*, p. 71.

Connors, S. (1974, June 5). Voltage-to-Frequency Converters: A / D's with Advantages. *EDN*, pp. 49–54.

DeVito, L. (1986, October 16). Synchronous V/ F Converter Aids Linearity in Data Acquisition. *EDN*, pp. 183–190. (See 13 Nov. 1986 *EDN* for Part 2.)

Fattaruso, J. W. and Meyer, R. G. (1985, April). Triangle-to-Sine Wave Conversion with MOS Technology. *IEEE Journal of Solid-State Circuits*, Vol. SC-20, No. 2, pp. 623–631.

Fennel, J. W. (1969, July 1). Feedback Improves Sample-and-Hold Performance. *EDN*, pp. 63–64.

Feucht, D. (1975, June 5). Exploring Function-Generator Design Problems and Solutions. *EDN*, pp. 37–43.

Gibbons, J. F. and Horn, H. S. (1964, September). A Circuit with Logarithmic Transfer Response over 9 Decades. *IEEE Trans. Circuit Theory*, Vol. CT-11, No. 3, pp. 378–384.

Gilbert, B. (1977, August 18). Circuits for the Precise Synthesis of the Sine Function. *Electronic Letters*, Vol. 13, No. 7, pp. 506–508.

Globas, G., Zellmer, J. and Cornish, E. (1974, September). A 250-MHz Pulse Generator with Transition Times Variable to Less than 1 ns. *Hewlett-Packard Journal,* Vol. 26, No. 1, pp. 1–7.

Gookin, A. (1977, February). A Faster Integrating Analog-to-Digital Converter. *Hewlett-Packard Journal*, Vol. 28, No. 6, p. 9 (boxed section).

Graeme, J. (1974, September 5). Peak Detector Advances Increase Measurement Accuracy, Bandwidth. *EDN,* pp. 73–78.

Graeme, J. G. (1987, September 3). Low-Cost Quad Op Amps Boost Circuit Performance. *EDN*, pp. 213–222.

Gray, P. E. and Searle, C. L. (1969). *Electronic Principles: Physics, Models, and Circuits.* New York: Wiley, Ch. 4.

Henry, P. (1987, May 14). JFET-Input Amps Are Unrivaled for Speed and Accuracy. *EDN,* pp. 161–169.

Kapoor, A. and Bowers, D. (1986, March 6). Diverse Circuits Exploit Matching in Quad-transistor IC. *EDN*, pp. 223–232.

Kirby, S. (1983, April 21). Deglitcher Circuit Refines d-a-Converter Output. *Electronics*, p. 151.

Kitchin, C. and Counts, L. (1985, August 22). Multifunction IC Provides Diverse Math Functions. *EDN*, pp. 175–180.

Klein, G. and Hagenbeuk, H. (1967, June). An Accurate Triangular-Wave Generator with Large Frequency Sweep. *Electronic Engineering*, pp. 388–390.

Klein, G. and Hagenbeuk, H. (1967, November). Accurate Triangle-Sine Converter. *Electronic Engineering*, pp. 700–704.

Knapp, R. (1988, February 18). Selection Criteria Assist in Choice of Optimum Reference. *EDN*, pp. 183–192.

Kolber, M. and McNerney, F. J. (1981, July 23). Reduce Capacitor Leakage Resistance for Stable, Long Time Delays. *Electronic Design*, p. 173.

Krauss, G. and Eggert, R. (1974, March). A Moderately Priced 20-MHz Pulse Generator with 16-Volt Output. *Hewlett-Packard Journal*, Vol. 25, No. 7, pp. 10–15.

Little, A. and Burnett, B. (1988, February 4). S / H Amp-ADC Matrimony Provides Accurate Sampling. *EDN*, pp. 153–166.

Lobjinski, M. and Bermbach, R. (1982, June 2). Switched-Capacitor Technique Improves a-d Conversion. *Electronics*, pp. 149, 151.

Mego, T. J. (1987, June 25). Resolve 22 Bits Easily with Charge-Balance ADCs. *Electronic Design*, pp. 109–114.

Meyer, R. G., Sansen, W., Lui, S. and Peeters, S. (1976, June). The Differential Pair as a Triangle-Sine Wave Converter. *IEEE Journal of Solid-State Circuits*, pp. 418–420.

Neal, J. and Surber, J. (1984, May 3). Track-and-Holds Take Flash Converters to Their Limits. *Electronic Design*, pp. 381–388.

O'Farrell, J. J. (1989, January 5). 555 Timer Triggers on Millivolt Signal. *EDN*, pp. 208–209.

Peterson, B. (Ed.) (1979, March 5). Inherently Monotonic DAC-IC Design Eliminates Costly Resistor Matching. *EDN*, pp. 46, 49.

Phillips, A. B. (1962). *Transistor Engineering*. New York: McGraw-Hill, p. 131.

Potson, D. and Swing, C. (1988, October 13). Clamping Op Amp Improves ADC. *Electronic Design*, pp. 124–125.

Regan, T. (1982, September 16). CMOS DACs Work Backwards to Provide Voltage Outputs. *Electronic Design*, pp. 117–122.

Rife, D. (1982, April 28). High Accuracy with Standard ICs: An Elegant ADC's Forte. *EDN*, pp. 137–144.

Shaw, W. S. (1972, May 8). Triangular-Wave Generator Spans Eight Decades. *Electronics*, p. 104.

Sheingold, D. H. (Ed.) (1974). *Nonlinear Circuits Handbook*. Norwood, MA: Analog Devices, Inc.

Sprowl, R. (1988, August 18). PWMDAC Simplifies Output Filtering. *EDN*, p. 220.

Stefenel, R. (1982, July 14). Positive Pulse Triggers 555 Integrated-Circuit Timer. *Electronics*, p. 169.

Williams, J. (1982, May 12). Expand Linear Circuit Functions with Nonlinear Design Schemes. *EDN*, pp. 153–158.

Williams, J. (1982, September 1). Current-Source Alternatives Increase Design Flexibility. *EDN*, pp. 169–174.

Williams, J. (1985, May 16). Design Techniques Extend V / F-Converter Performance. *EDN*, pp. 153–164.

Wyatt, M. A. (1989, March 2). MOSFET Switch Compensates Sampler. *EDN*, p. 176.

Zuch, E. L. (1978, December 6). Keep Track of Sample-Hold from Mode to Mode to Locate Error Sources. *Electronic Design*, pp. 80–87.

INDEX

Index Terms	Links			
P				
Padé approximation	116			
parallel or flash converters	182			
parallel-feedback converters	156	157	177	
Paul Brokaw	11			
peak detectors	86	129	158	201
	204			
peaking	41	58		
peak-to-peak auto	86			
PGA	153	184		
phase shifter	54			
phase-lead compensator	116			
phase-locked loop	106			
piecewise-linear waveform	118			
positive-feedback amplifier	58			
power-line	170			
PSR	18	19	21	
pulse-width modulation	118	136	147	
Q				
quantization noise	139			
quasistable state	66			
R				
R-2R network	142	147		
ramp converter	154	155	157	
RC differentiator	67			
recirculating-remainder ADC	185			