30th Anniversary Reader Bonus

Ask The Applications Engineer

A Selection from *Analog Dialogue*

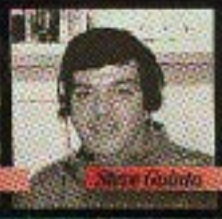
A forum for the exchange of circuits, systems, and software for real-world signal processing

When it comes to trimming

V/F converters

Multi troubles - Trouble from the start

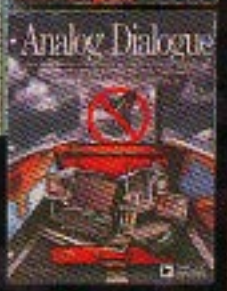
About log compensation resistors



Must a "16-bit" converter be 16-bit monotonic and settle to 16 ppm?

Settling time

Interfacing to serial converters, part I



Interfacing to serial converters, part II

Capacitance and capacitors

Current-feedback amplifiers, part I

Current-feedback amplifiers, part II

ANALOG DEVICES

ANALOG DEVICES TECHNICAL REFERENCE BOOKS

Published by Analog Devices

- **Nonlinear Circuits Handbook**
- **Transducer Interfacing Handbook**
- **Mixed-Signal Design Seminar Notes**
- **Amplifier Applications Guide**
- **System Applications Guide**
- **Linear Design Seminar Notes**
- **Practical Analog Design Techniques**
- **High-Speed Design Seminar Notes**
- **ADSP-21000 Family Applications Handbook**

Published by Prentice Hall (available from Analog Devices)

- **Analog-Digital Conversion Handbook**
- **Digital Signal Processing in VLSI**
- **DSP Applications Using the ADSP-2100 Family (Vols. 1 & 2)**

Ask The Applications Engineer

WORLDWIDE HEADQUARTERS
One Technology Way
P.O. Box 9106, Norwood, MA 02062-9106, U.S.A
Tel: (617) 329-4700, Fax: (617) 326-8703

Printed in the United States of America

TABLE OF CONTENTS

PREFACE AND INTRODUCTION

This special issue of *Analog Dialogue* is offered as a bonus to our faithful readers and an encouragement to new readers. We have reprinted here the popular series, entitled "Ask The Applications Engineer", from its inception in 1988 through Number 23* in 1996.

We are celebrating *Analog Dialogue*'s 30th sequential year in print, just concluded. During those years, *Analog Dialogue* has informed, enlightened, and occasionally educated more than a generation of engineers, scientists, and technicians about Analog Devices's products, practice, and ideas for high-performance signal processing by analog, digital, and mixed means.

This 30th anniversary Selection, *Ask The Applications Engineer*, joins the collection of articles in our 25th anniversary bonus edition, *The Best of Analog Dialogue—1967 to 1991*. Readers of that 224-page collection (still available upon request) will recall its many articles relating to *practice*—as well as articles about landmark products. Our intent in originating the "Ask The Applications Engineer" series has been to supplement that mine of information with a source of what has come to be called FAQ—frequently asked questions (and answers). They are chronicled by a few of our Applications Engineers, but are based on the accumulated wisdom and experience of our worldwide Applications Engineering staff, interacting with tens of thousands of customers worldwide.

Throughout the pages of this volume you will find references to Analog Devices products and publications, just as they appeared in the original edition. If you are interested in receiving copies of publications available from Analog Devices or information about specific products, you may consult our Web site, http://www.analog.com, or phone our Help line, 1-800-ANALOGD (1-800-262-5643) in North America, (617) 937-1428 elsewhere, or one of our local offices. Data sheets are available in North America by 24-hour automated fax by phoning the AnalogFax™ number, 1-800-446-6212. Requests for publications may also be faxed to (617) 326-8703. Naturally, any timely material is on its way to obsolescence the day it appears in print, so it's possible that at the time you read about a particular device, updated or improved versions have become available. For updated information, phone our Help line or one of our local offices.

The articles in this publication appear in chronological order, starting with "Ask The Applications Engineer—1". But in order for the book to be useful as a reference source, we have provided an Index at the back; it will permit the reader to find material relating to a particular interest wherever it appears. We hope you will find this publication helpful and will welcome your suggestions of questions to be answered in the future.

Dan.Sheingold@analog.com

Editor, *Analog Dialogue*

*With one exception: We have not included "Ask The Applications Engineer—4", which was devoted to a description of a 5.25" Component Selection Guide diskette that was novel and useful in its time; however, it has long been obsolete—in this day of the Web site, the CD-ROM, and the selection tree in print. Just about everything else has turned out to include information of timeless value.

Ask The Applications Engineer—1

by James Bryant

MULTI TROUBLES

Q. My multiplexed ADC system is misbehaving . . .

A. Before you go any further, have you grounded all unused multiplexer channels?

Q. No. But how did you know?

A. Because the floating terminal is one of the commonest causes of problems in systems containing CMOS multiplexers. Unused MUX inputs and outputs (whether integrated into a multiplexed ADC or part of a self-contained MUX chip) can pick up signals from stray fields and inject them into the device's substrate, turning on spurious substrate devices. Then, even when the unused channel is turned off, the performance of the *on-channel may be* badly degraded (at *the* unlikely extreme, *the* injection may turn on a spurious four-layer device and destroy some chips).

Whenever a MUX is used, all its inputs and outputs must be connected to a potential between its supply rails. The best way to deal with unused channels is to ground them, but they may be connected to a more-convenient potential within the rails. ▶

TROUBLE FROM THE START

Q. To save power, my ADC is powered up only to make a measurement. The system is very accurate in continuous operation, but unpredictable when power is strobed. Why?

A. When an ADC's power is switched on only to perform a conversion, it may misbehave for three reasons: slow reference turn-on, random initial logic states, and system latch-up.

For various reasons—thermal stabilization, capacitance charging, slow starting of regenerative current mirrors using PNP transistors in band-gap references— it is not uncommon for some voltage references to have relatively large errors for many milliseconds after power-up. Such errors in an ADC's external or internal reference during conversion lead to inaccurate results.

At turn-on, a typical ADC's logic will be in a random state; for a conversion triggered at that time, the ADC may not be able to perform correctly. With one conversion triggered, the logic should return to its correct pre-conversion state—but cases exist where two conversion cycles are necessary before the ADC is certain to perform a valid conversion. Hence, a good general rule is to perform two "dummy" conversions after powerup before relying on the results. (It is also well to recall that some ADCs react badly to having a conversion triggered before the previous conversion is complete; when this happens, one or two "dummy" conversions may be needed to return the logic to a known state.)

If an ADC's external logic is arranged so that the end of the ADC "Busy" signal starts a delay which ends with the start of the next conversion, it is important to realize that if the converter powers up in the Busy state, the Busy signal may remain latched up until a conversion Start pulse has been received. In this case, such a system cannot self-start. If the Busy signal is always present on power-up the problem is almost certain to be recognized—and addressed—during the design of the system; but if the Busy signal is only occasionally present on power-up the system may latch unpredictably. As a rule, control signals to an ADC during start-up should not depend on the logical state of Busy. ▶

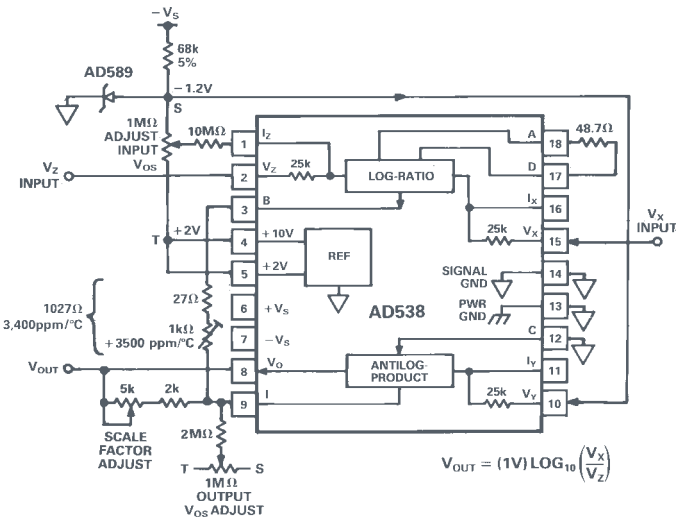
ABOUT LOG COMPENSATION RESISTORS

Q. Designs of logarithmic circuits, including those using the AD538 Y[Z/X]^m unit: (for example, Figure 6 from the AD538 Multifunction Unit data sheet†) call for "kT/q compensation resistors". What are they and where do I get them?*

A. The V_{BE} difference across two opposed silicon junctions, one carrying a current, I, and the other a current, I_{REF}, is $(kT/q)\ln(I/I_{REF})$. Here, k/q is the ratio of Boltzmann's constant to the charge on an electron (about 1/11,605 K/V), and T is the absolute temperature in kelvins.

Although employing similar junctions in isothermal pairs eliminates the effects of temperature-sensitivity of reverse saturation current, the kT/q term is still temperature-dependent. To eliminate this dependency in the application, the logarithmic voltage must be used in a circuit whose gain is inversely proportional to the absolute temperature of the junctions. Over a reasonable range of temperatures near 20°C, this may be arranged by the use of a gain-setting 1-kΩ resistor having a positive temperature coefficient of approximately 3,400ppm/°C—and keeping it at the same temperature as the junctions.

A 3,500 ppm/°C resistor is available from Tel Laboratories, 154G Harvey Road, Londonderry, New Hampshire 03053 (603)-625-8994, Telex: (710)-220-1844, designated Q-81, and from the Precision Resistor Co. Inc., 10601, 75th. St., Largo, Florida 33543 [(813)-541-5771 Telex: 821788], as the PT146. Analog Devices offices in most European countries are aware of local suppliers of these resistors. ▶



*Much useful information about logarithmic and other analog function circuits can be found in the Nonlinear Circuits Handbook, published by Analog Devices ($5.95), P.O. Box 9902, Norwood MA 02062.

†See The Best of *Analog Dialogue* 1967–1991, pp. 164-167.

Ask The Applications Engineer—2

by James Bryant

WHEN IT COMES TO TRIMMING . . .

Q. I need some advice about trimming offsets and gains.

A. Don't!—unless you must. Good alternatives include (a) using headache-free devices, components, and circuits that meet the specs without trimming; (b) taking advantage of digital technology in system applications to make trim corrections in software. Savings provided on occasion by trim potentiometers, in conjunction with loosely spec'd devices, can turn out to be illusory when you consider the effects of circuit design, temperature, vibration, and life on performance and stability—as well as additional paperwork and complexity trimming entails.

Q. Nevertheless, how do I trim the offset and gain errors in analog circuitry?

A. In the correct order and with the correct inputs. If you consider the transfer characteristic of the circuit being trimmed the method to use is generally straightforward.

The simplified ideal transfer characteristic of a linear analog circuit (such as an amplifier, ADC or DAC) is given by the equation:

$$OP = K \times IP \quad (1)$$

where OP is output, IP is input, and K is a scale factor (Note that this simplification hides an enormous number of issues: quantization error in an ADC; dimensionality of K if the input and output are in different forms [e.g. voltage in / current out]; intentional offsets; and many others.)

In a real (non-ideal) circuit, offset and gain errors, OS (referred to the input) and ΔK, respectively, also appear in the equation, which becomes:

$$OP = (K + \Delta K) \times (IP + OS) \quad (2)$$

$$OP = (K \times IP) + [(K \times OS) + (\Delta K \times IP) + (\Delta K \times OS)] \quad (3)$$

Equations (2) and *(3)* are incomplete in that they assume only one offset—at the input—but this is the most-common case. Systems with separate input and output offsets will be considered later.

From *(3)* we see that it not possible to trim gain directly when an unknown offset is present. Offset must be trimmed first. With IP set at 0, the offset trim is adjusted until OP *is* also 0. Gain may then be trimmed: with an input near to full scale (FS), the gain trim is adjusted to make the output obey equation (1).

Q. But what about bipolar ADCs and DACs?

A. Many ADCs and DACs may be switched between unipolar and bipolar operation; such devices, wherever possible, should have their offset and gain trimmed in the unipolar mode. Where it is not possible, or where the converter is to operate only in the bipolar mode, other considerations apply.

A bipolar converter may be considered as a unipolar converter with a large offset (to be precise, an offset of 1 MSB—one-half of full-scale range). Depending on the architecture used, this bipolar offset *(BOS)* may or may not be affected by the gain trim. If it is so affected, equation (1) becomes:

$$OP = K \times (IP - BOS) \quad (4)$$

In this case offset is trimmed at analog zero, after which gain is trimmed near FS—positive or negative, but usually positive. This is normally the method used for DACs where the bipolar offset is within the DAC.

If the bipolar offset is not affected by the gain trim:

$$OP = K \times IP - BOS \quad (5)$$

Here offset is trimmed at FS negative and gain is trimmed at (or very near to—see below) FS positive. This method is used for most ADCs and for DACs where bipolar offset is obtained by the use of op amps and resistors external to the DAC.

Naturally, the method suggested on the data sheet should always be followed, but where a data sheet is unobtainable, in general, offset should be trimmed at analog zero for DACs and FS negative for ADCs—and near FS positive for both.

Q. Why do you keep saying "near" to full scale?

A. Amplifiers and DACs may be trimmed at zero and full scale. In the case of a DAC, all-1's—the largest digital input possible—should produce an output 1 LSB below "full scale," where "full scale" is considered as some constant times the reference; this follows since the output of a DAC is the normalized product of the reference and the digital input.

ADCs are not trimmed at zero and FS. The output of an ideal ADC is quantized, and the first output transition (from 00 . . . 00 to 00 . . . 01) takes place 1/2 LSB above the nominal value of all 0's. Thereafter transitions take place every 1-LSB increase in analog input until the final transition takes place 1 1/2 LSB below FS. A non-ideal ADC is trimmed by setting its input to the nominal value of a desired transition and then adjusting until the ADC output flickers between the two values equally.

The offset of an ADC is therefore trimmed with an input corresponding to the first transition (i.e., 1/2 LSB above zero or above FS negative—which is "near" zero or "near" FS negative); and the gain is then trimmed at the last transition (i.e. 1 1/2 LSB below FS positive—which is "near" FS positive). This procedure results in an interaction between the gain and offset errors during offset trim but it should be too slight to be significant.

Q. Are there any other anomalies resulting in a need to trim "near", rather than at full scale?

A. Synchronous voltage-to-frequency converters (SVFCs) are liable to injection locking phenomena when their output frequency is harmonically related to their clock frequency, i.e., when their output is very close to 1/2, 1/3 or 1/4 of clock frequency. FS for an SVFC is 1/2 clock frequency. The presence of a trim tool can exacerbate the problem. It is therefore advisable to trim the gain of an SVFC at around 95% of FS.

Q. What about circuits requiring both "input" and "output" offset trim?

A. Circuits such as instrumentation and isolation amplifiers often have two stages of dc gain, and the gain of the input stages can be variable. Thus a two stage amplifier, with an input offset, *IOS*, an output offset, *OOS*, a first stage gain of *K*, and a unity-gain output stage, has (for zero input) an output, *OP*, of:

$$OP = OOS + K \times IOS \qquad (6)$$

From (6) it is evident that if the gain is constant we need only adjust either *IOS* or *OOS* to null the total offset (although if the input uses a long-tailed pair of bipolar transistors we will get a better offset temperature coefficient if we trim both—for FET long-tailed pairs this is not necessarily the case). If the first stage gain is to be varied, both offsets must be trimmed.

This is done by an iterative procedure. With zero input, and gain set to maximum, the input offset is adjusted until the output is also zero. The gain is then reduced to its minimum value and the output offset adjusted until the output is zero again. The two steps are repeated until no further adjustment is necessary. Gain trimming should not be done until both *IOS* and *OOS* are pulled; the actual values of the high and low gains used in offset trim are unimportant.

Q. What circuitry should I use for gain and offset trims?

A. Many amplifiers (and a few converters) have terminals for trimming gain and offset. Many more do not.

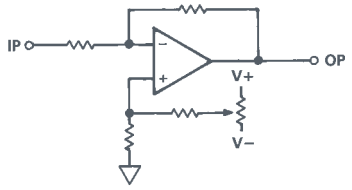
Offset trim is normally performed with a potentiometer connected between two assigned terminals, and its wiper is connected (sometimes via a resistor) to one of the supplies. The correct connections and component values will be given on the device data sheet. One of the commonest differences between op-amps is the value of offset correction potentiometer and which supply it should be connected to.

Where separate terminals are not provided for offset trim, an offset-adjusting constant can usually be added to the input signal. Two basic possibilities are shown in Figures 1a and 1b. Where the correction is being made to a system where a differential input op amp is used as an inverter (the commonest case) the method of la is best to correct for device offsets—but not system offsets. In the single-ended connection, method 1b will work for system offsets but should be avoided where possible for small device offsets, because it often requires a very large value of summing resistance, compared to the signal-input resistances, in order to (i) avoid loading the summing point excessively, (ii) scale the correction voltage properly and produce enough attenuation to minimize the effects of differential supply-voltage drifts. It is often helpful to use resistances between the supplies and the potentiometer to increase trim resolution and reduce dissipation.

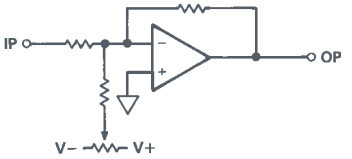
Where gain trim is provided for in a circuit, it will generally consist of a variable resistor. Details of its value and connection will appear on the data sheet of the device. Where gain trim is not required, this resistor may be replaced by a fixed resistor having half the resistance of the maximum value of the recommended trim potentiometer.

Where gain trim is not provided it is not always achievable externally without an additional variable-gain stage. For example, consider a DAC using a ladder network. If the ladder network is used in the current mode (Figure 2a), the input impedance at the reference terminal does not vary with digital code, and the gain of the DAC may be trimmed with a small variable resistor in series with either the reference input or the feedback resistor. However, if the DAC is used in the voltage mode (Fig 2b), then the reference input impedance is code dependent, and gain may only be trimmed by varying the reference voltage—which is not always possible—or the gain of the buffer amplifier.

The possibility of trimming gain in circuits not furnished with gain-trim circuitry, therefore, will depend on individual cases; each must be assessed on its own merits. ▶

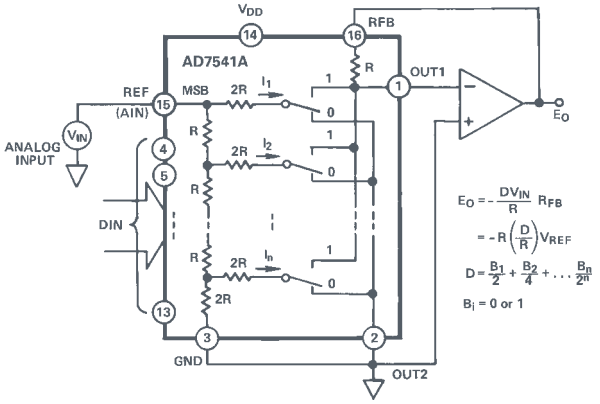


a. Voltage applied to non-inverting input.

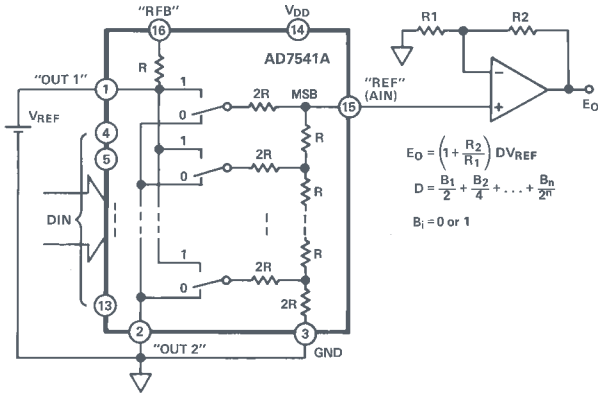


b. Current summed at inverting input.

Figure 1. Two connections for offset adjustment.



a. CMOS DAC connected for current steering. Input impedance is constant.



b. The same DAC connected for voltage output.

Figure 2. Comparing basic DAC circuits.

Ask The Applications Engineer—3

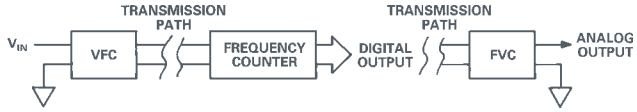
by James Bryant

V/F CONVERTERS

Q. How do I send an analog signal a long distance without losing accuracy?

A. An excellent solution to this common problem is to ship the signal as frequency using a *voltage-to-frequency converter* (VFC), a circuit whose output is a frequency proportional to its input. It is relatively easy to send a frequency signal over a long transmission path without interference via optical isolators, optical fibre links, twisted-pair or co-axial lines, or radio links.

If the data must be digital, the receiver will consist of a frequency counter, easily implemented in a single-chip microcomputer. Frequency is reconverted to analog voltage by a "frequency-to-voltage converter" (FVC)—generally a VFC configured to perform its inverse function, often using a phase-locked loop.

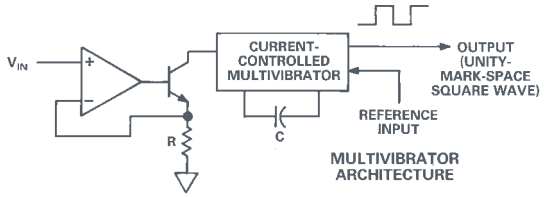


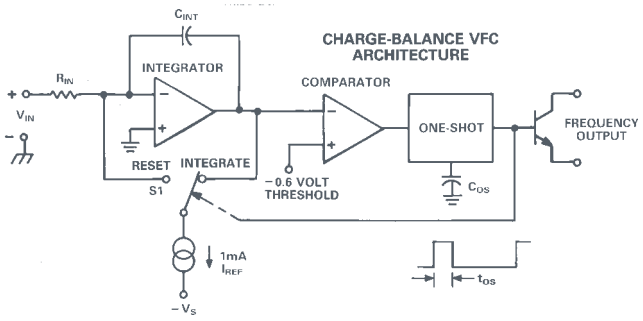
Q. How does a VFC work?

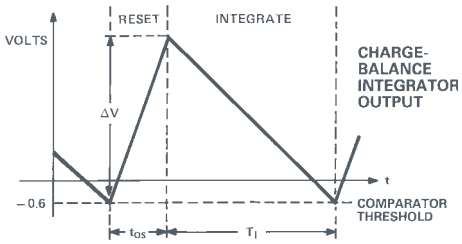
A. There are two common types: multivibrator-(AD537) and charge-balance (AD650) VFCs.*

In the *multivibrator* type, the input voltage is converted to a current which charges and discharges a capacitor. The switching thresholds are set by a stable reference, and the output, which has unity mark-space ratio, is a frequency proportional to the input.

The *charge-balance* VFC uses an integrator, a comparator and a precision charge source. The input is applied to the integrator, which charges. When the integrator output reaches the comparator threshold, the charge source is triggered and a fixed charge is removed from the integrator. The rate at which charge is removed must balance the rate at which it is being supplied, so the frequency at which the charge source is triggered will be proportional to the input to the integrator.







Q. What are the advantages and disadvantages of the two types?

A. The multivibrator is simple and cheap, demands little power, and has unity mark-space (M-S) output—very convenient with some transmission media. But it is less accurate than the charge-balance type and cannot integrate negative input transients.

The charge-balance type is more accurate, and negative input transients are integrated to contribute to the output. It has more-demanding supply requirements and a lower input impedance, and its output is a pulse train, not a unity M-S square wave.

Q. What are the important types of error in a VFC?

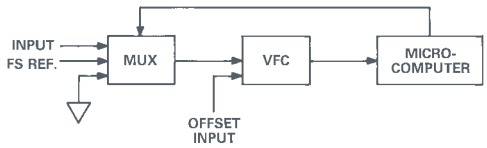
A. The same three as in most precision circuitry: offset errors, gain errors and linearity errors—and their variation with temperature. As with most precision circuitry, offset and gain can be trimmed by the user, but linearity cannot. However, the linearity of VFCs is normally very good (if the capacitors are properly chosen—see below).

Q. How do you trim gain and offset in a VFC?

A. The procedure suggested by theory is to trim offset first at zero frequency and then gain at full scale (FS). But this can give rise to problems in recognizing "zero frequency," which is the state when the VFC is just not oscillating. It is therefore better to trim offset with a small input (say 0-1% FS) and adjust for a nominal frequency, then trim gain at FS, and then repeat the procedure once or twice.

For example, suppose a VFC is being used with FS of 100 kHz at 10-volt input. Ideally, 10 V should give 100-kHz output and 10-mV input should give 100 Hz. Offset is, therefore, trimmed for 100 Hz with 10 mV applied; gain is then trimmed to give 100 kHz at 10 V. But gain error affects the 10-mV offset trim slightly, so the procedure may have to be repeated to reduce the residual error.

If a VFC is used with software calibration a deliberate offset is often introduced so that the VFC has a definite frequency for zero input voltage. The microcomputer measures the VFC outputs at 0 V and FS inputs and computes the offset and scale factor. It may also be necessary to reduce the gain so that the VFC cannot try to exceed its maximum rated frequency.



*Data sheets are available for any of the Analog Devices products mentioned here. An Application Note: "Operation and Applications of the AD654 V-to-F Converter," is also available without charge.

Q. What circuit precautions are necessary when using a VFC?

A. Apart from the usual precautions necessary with any precision analog circuitry (grounding, decoupling, current routing, isolation of noise, etc., a subject for a book, not a paragraph) the main precautions necessary when using a VFC are the choice of capacitor and separation of the input and output.

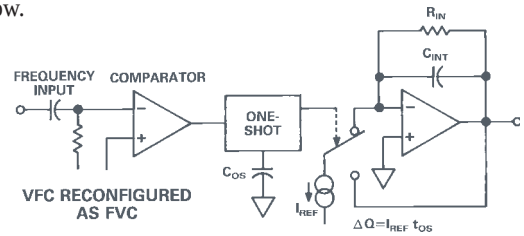
The critical capacitors in a precision VFC (the multivibrator's timing capacitor, and the monostable timing capacitor in a charge-balance type) must be stable with temperature variation. Furthermore, if they suffer from *dielectric absorption,* the VFC will be nonlinear and may have poor settling time.

If a capacitor is charged, discharged and then open-circuited it may recover some charge. This effect, known as dielectric absorption (DA), can reduce the precision of VFCs or sample-hold amplifiers using such capacitors. VFCs and SHAs should therefore use Teflon or polypropylene, or zero- temperature-coefficient (NPO, COG) ceramic capacitors with low DA.

Coupling between output and input of a VFC can also affect its linearity. To prevent problems, decoupling practices and the usual layout precautions should be observed. This is critically important with opto couplers, which require high current drive (10-30 mA).

Q. How do you make a frequency-to-voltage converter?

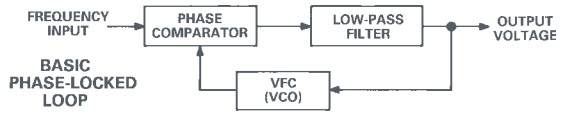
A. There are two popular methods: the input frequency triggers the monostable of a charge-balance VFC that has a resistor in parallel with its integration capacitor; or the input frequency can be applied to the phase/frequency comparator of a phase-locked loop (PLL), which uses a VFC (of either type) as its oscillator. The basic principle of the first type is illustrated below.



For each cycle of the input frequency, a charge, ΔQ, *is* delivered to the leaky integrator formed by R *and* C. *At equilibrium, an equal charge must leak away during each period,* T $(= 1/f)$, of the input, at an average rate, $I = V/R$. Thus, $V = \Delta Q \cdot f \cdot R$.

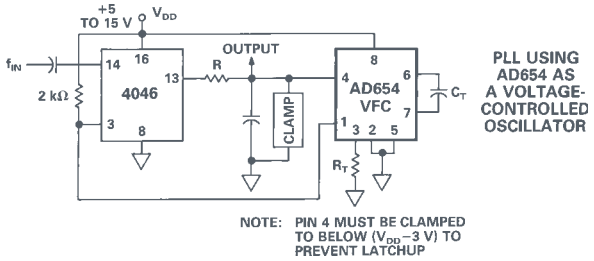
Though the mean voltage is independent of C, the output ripple is inversely dependent on C. The peak-to-peak ripple voltage, ΔV, *is* given by the equation, $\Delta V = \Delta Q/C$. This indicates that ripple is independent of frequency (assuming that the charge, Q, *is* delivered in a short time relative to the period of the input). The settling time of this type of FVC is determined by the exponential time constant, RC, from which the time to settle within a particular error band may be calculated.

From these equations, we see that the characteristics of this type of FVC are interdependent, and it is not possible to optimize ripple and settling time separately. To do this we must use a PLL.



The phase-locked-loop FVC illustrated differs from any other PLL in only one respect: the voltage-controlled oscillator of the normal PLL, which must be monotonic but not necessarily linear, has been replaced by a VFC with a linear control law. In the servo system, negative feedback keeps the VFC's output frequency equal to the input frequency. The output voltage, the VFC's input, is accurately proportional to the input frequency.

Designing PLL systems is beyond the scope of this discussion,[1] but if a 4000-series CMOS PLL, the 4046, is used *just as a phase detector* (its VCO's transfer characteristic is not sufficiently linear), we can build the FVC shown here, with an AD654 VFC.



Q. What is a synchronous VFC?

A. A charge-balance VFC with improved linearity and stability, where the monostable is replaced by a bistable, driven by an external clock. The fixed time during which the precision current discharges the integrator is one clock period of the external clock.

A further advantage of the SVFC is that the discharge does not start when the integrator passes the comparator threshold (at a non-critical rate), but on the next clock cycle. The SVFC output is synchronous with a clock, so it is easier to interface with counters, µPs, etc.; it is especially useful in multichannel systems: it eliminates problems of interference from multiple asynchronous frequency sources.

There are two disadvantages. Since the output pulses are synchronized to a clock they are not equally spaced but have substantial jitter. This need not affect the user of a SVFC for a/d conversion, but it does prevent its use as a precision oscillator. Also, capacitive coupling of the clock into the comparator causes injection-lock effects when the SVFC is at 2/3 or 1/2 FS, causing a small (4-6 bit at 18-bit resolution at 1-MHz clock) dead zone in its response. Poor layout or device design can worsen this effect.

Despite these difficulties the improvement in performance produced by the abolition of the timing monostable makes the SVFC ideal for the majority of high-resolution VFC applications.

Q. Can you have a synchronized FVC?

A. Yes, and with very good performance; it is best done with an FVC- connected SVFC and a clock that is common to both ends of the transmission path. If the input signal to a synchronized FVC is *not* phase related to the clock, severe timing problems can arise, which can only be solved by the use of additional logic (two D flip-flops) to establish the correct phase relationship. ▶

[1]See Gardner, F. M, *Phase-lock Techniques,* 2nd ed., New York: Wiley, 1979, for more detail; also Analog Devices' *Analog-Digital Conversion Handbook.*

Ask The Applications Engineer—5

High-speed comparators provide many useful circuit functions when used correctly.

by John Sylvan

Question: Why can't I just use a standard op amp in a high-gain or open-loop configuration as a voltage comparator?

You can—if you are willing to accept response times in the tens of microseconds. Indeed, if in addition you require low bias-currents, high-precision and low offset voltages, then an op amp might be a better choice than most standard voltage comparators. But since most op amps have internal phase/frequency compensation for stability with feedback, it's difficult to get them to respond in nanoseconds. On the other hand, a low-cost popular comparator, the LM311, has a response time of 200 ns.

Also the output of an operational amplifier is not readily matched to standard logic levels. Without external clamping or level-shifting, an op amp operating as a comparator will swing to within a few volts of the positive and negative supplies, which is incompatible with standard TTL or CMOS logic levels.

My comparator oscillates uncontrollably. Why does this happen?

Examine the power-supply bypassing. Even a few inches of PC trace on the supply lines can add unacceptable dc resistance and inductance. As a result, transient currents while the output is switching may cause supply-voltage fluctuations, which are fed back to the input through the ground and supply lines. Install low-loss capacitors (0.1μF ceramic capacitors) as close as possible to the supply pins of the comparator to serve as a low-impedance reservoir of energy during high-speed switching.[1]

I've installed bypass capacitors, but I still can't keep my high-speed comparator from oscillating. Now what's the problem?

It could be the comparator's ground connection. Make sure that the ground lead is as short as possible and connected to a low-impedance ground point to minimize coupling through lead inductance. Use a ground plane if possible and avoid sockets.

Another cause of the oscillation may be a high source impedance and stray capacitance to the input. Even a few thousand ohms of source impedance and picofarads of stray capacitance can cause unruly oscillations. Keep leads short, *including the ground clip of your scope probe*. For best measurement results use the shortest possible ground lead to minimize its inductance (< 1").

With a slowly moving input signal, my comparator seems to "chatter" as it passes through the transition voltage. Why can't I obtain a single clean transition from the device?

A comparator's high gain and wide bandwidth are typically the source of this problem. Any noise is amplified, and as the signal passes through the transition region, the noise can cause a fast-responding amplifier's output to bounce back and forth. Also, since the device's sensitivity (i.e., *gain)* is higher during a transition, the tendency to oscillate due to feedback increases. If possible, filter the signal to minimize noise accompanying it. Then try using hysteresis which, like backlash in gear trains, requires the input to change by a certain amount before a reversal occurs.

[1]A useful discussion of comparator foibles can be found in *Troubleshooting Techniques Quash Spurious Oscillations*, by Bob Pease, *EDN*, September 14, 1989, pp. 152-6

For example, after a high-to-low transition on the AD790, its built-in hysteresis requires the input voltage (positive input) to increase by 500 μV to produce a low-to-high transition.

If my comparator does not have internal hysteresis, can I add it externally?

Yes, with external positive feedback. This is done by feeding a small fraction of the output of the comparator back to the positive input. This simple technique is shown in Figure 1. The hysteresis voltage from the lower transition point to the upper transition point will depend on the value of the feedback resistor, R_F, the source resistance, R_S, low output level, V_{low} and high output level, V_{high}. The low and high transition points are set by:

$$V_{low} \times \frac{R_S}{R_S + R_F} \text{ and } V_{high} \times \frac{R_S}{R_S + R_F}$$

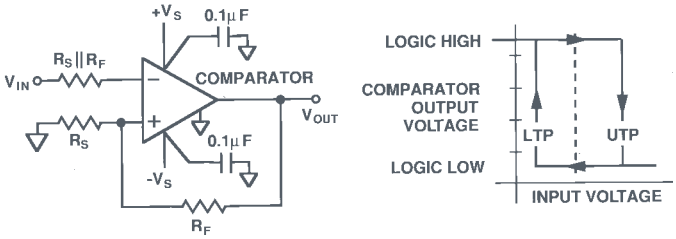
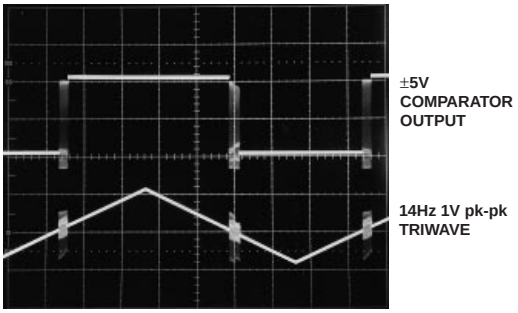
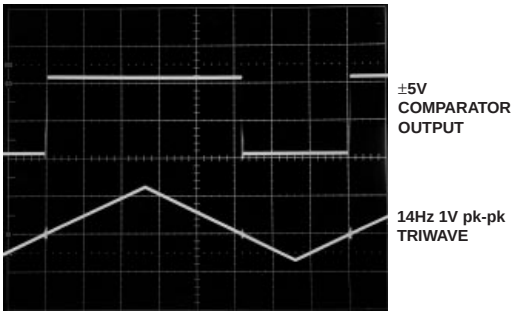


Figure 1. Applying external hysteresis to a comparator.

Figure 2 shows how adding external hysteresis can "clean up" a comparator's response. Figure 2a shows the response of a comparator with bipolar output swing without hysteresis. As the triangular-wave input (trace A) passes through the transition point (ground), the device oscillates vigorously (and couples a portion of the oscillation into ground and the signal-source). Figure 2b depicts the response of the same comparator with 5 mV of external hysteresis applied; it shows a much cleaner transition.



No external hysteresis



5 mV external hysteresis

Figure 2. Hysteresis helps clean up comparator response.

A problem encountered with external hysteresis is that output voltage depends on supply voltage and loading. This means the hysteresis voltage can vary from application to application; though this affects resolution, it need not be a serious problem, since the hysteresis is usually a very small fraction of the range and can tolerate a safety margin of two or three (or more) times what one might calculate. Swapping in a few comparators can help confidence in the safety margin. Don't use wirewound resistors for feedback; their inductance can make matters worse.

What's the difference between propagation delay and prop-delay dispersion? Which of the two specifications is of most concern?

Propagation delay is the time from when the input signal crosses the transition point to when the output of the comparator actually switches. Propagation-delay dispersion is the variation in prop delay as a function of overdrive level. If you're using a comparator in pin-drive electronics in an automatic test system, then prop-delay dispersion will determine the maximum edge resolution. In contrast, propagation delay can be considered as a fixed time offset and therefore compensated for by other techniques.

I have a +5-volt system and don't want to add an additional supply voltage; can I use my comparator with a single supply?

Yes, but to establish the threshold use an adequately bypassed stable reference source well within the common-mode range of the device. The signal level is also to be referenced to this source.

I sometimes see unexpected behavior in my comparator. What could be the cause of this problem?

Examine the common-mode range of the input signal. Unlike operational amplifiers, that usually operate with the input voltages at the same level, comparators typically see a large differential voltage swings at their inputs. If the inputs exceed the device's specified common-mode range (even though within the specified *signal* range), the comparator may respond erroneously. For proper operation, ensure that *both* input signals do not exceed the common-mode range of the specific comparator. For example, the AD790 has a $+V_S$ differential input range, but its common-mode range is from $-V_S$ to 2 volts below $+V_S$.

Can you suggest a circuit that performs autozeroing when the comparator is off-line to minimize drift?

Try the circuit shown in Figures 3 and 4. In the Calibrate mode, the input is disconnected and the positive input of the comparator is switched to ground. The comparator is connected in a loop with a pair of low-voltage sources of opposing polarity charging a buffered capacitor in response to the comparator's output state.

If the comparator's minus input terminal is above ground, then the comparator output will be low, the 1-µF capacitor will be connected to the negative voltage (–365 mV) and the voltage from the buffer amplifier will ramp down until it is below the plus input (ground)—plus hysteresis and any offsets—at which point the comparator switches. If it is below ground, the comparator's output will be high, the capacitor will be connected to the positive voltage (+365 mV), the output from the buffer amplifier ramps up. In the final state, each time the comparator switches (when the ramped change exceeds the hysteresis voltage), the polarity of the current is reversed; thus the capacitor voltage averages out the offsets of the buffer and comparator.

At the end of the Calibrate cycle, the JFET switch is opened, with the capacitor charged to a voltage equal to the offsets of the comparator and buffer ± the hysteresis voltage. At the same time, the Calibrate signal goes low, disabling the feedback to the polarity switch and connecting the input signal to the comparator.

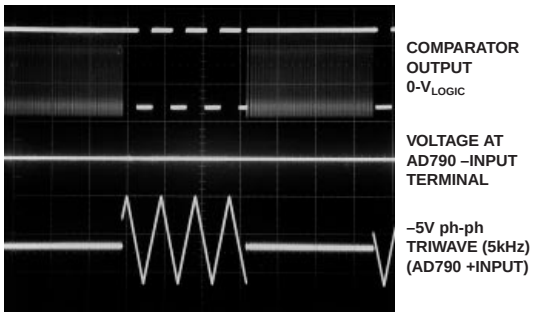


Figure 4. Comparator output, buffer output, and comparator input.

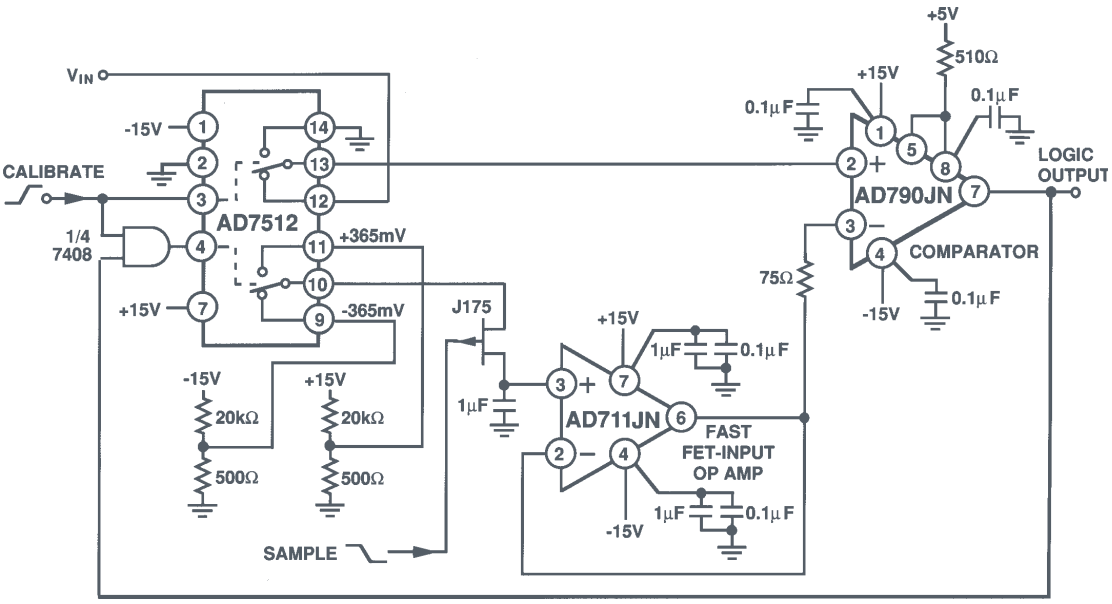


Figure 3. Autozeroed comparator integrates out offsets during calibration cycle.

Ask The Applications Engineer—6

by James Bryant

OP-AMP ISSUES

Q. Why are there so many different types of operational amplifier?

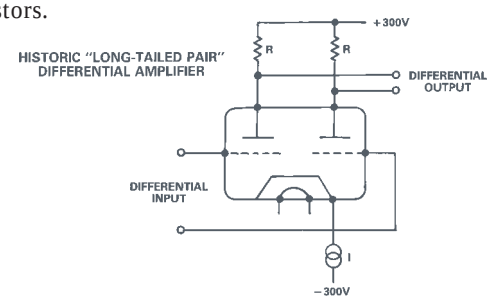
A. Because there are so many parameters that are important in different applications, and because it is impossible to optimize all of them at once. Op amps may be selected for speed, for noise (voltage, current or both), for input offset voltage and drift, for bias current and its drift, and for common-mode range. Other factors might include power: output, dissipation, or supply, ambient temperature ranges, and packaging. Different circuit architectures and manufacturing processes optimize different performance parameters.

Q. Is there any common factor in the design of op-amps?

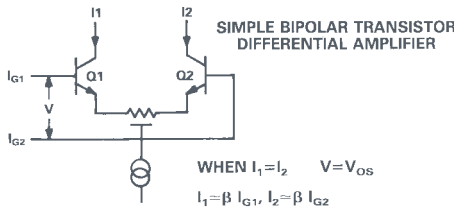
A. Yes—most classical (voltage input) op-amps are three-stage devices, consisting of an input stage with differential input and differential output—with good common-mode rejection—followed by a differential-input, single-ended output stage having high voltage-gain and (generally) a single-pole frequency response; and, finally, an output stage, which usually has unity voltage gain.

Q. So where are the differences?

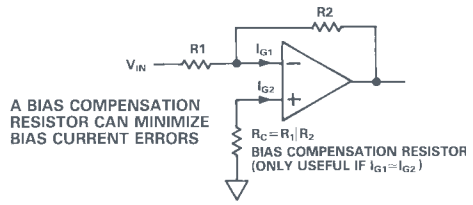
A. There are many possible variations on this basic design. One of the most fundamental is the structure of the input stage. This stage is almost always a long-tailed pair—that is to say, a pair of amplifying devices connected as in the figure—but the choice of devices has a profound effect on the input parameters of the op amp. The figure was drawn with thermionic tubes to avoid any suggestion of partiality in favour of any particular semiconductor device. Since thermionic devices at present are not generally available in IC chip form, a monolithic op-amp will have an input stage built with bipolar or field-effect transistors.



A long-tailed pair built with bipolar transistors is shown in the next figure. Its strong features are its low noise and, with suitable trimming, low voltage offset. Furthermore, if such a stage is trimmed for minimum offset voltage it will inherently have minimum offset drift. Its main disadvantage stems from the proportionality of the emitter and base currents of the transistors; if the emitter current is large enough for the stage to have a reasonable bandwidth, the base current—and hence the bias current—will be relatively large (50 to 1,000 nA in general-purpose op-amps, as much as 10 μA in high-speed ones).

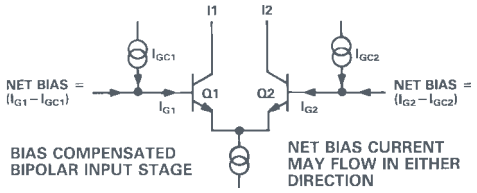


The bias currents in the inverting and non-inverting inputs are unipolar and well matched (their difference is called *offset current)*, and they decrease in a minor way with increasing temperature. In many applications, the accurate matching may be used to compensate for their high absolute value. This figure shows a bias compensation circuit where the bias current in the non-inverting input flows in R_c (known as the bias compensation resistor); this compensates for the voltage drop as the bias current in the inverting input flows through R2. R_c is made nominally equal to the parallel combination of R1 and R2.—it can be trimmed to minimize error due to non-zero offset current).



Such bias compensation is only useful when the bias currents are well-matched. If they are not well-matched, a bias compensation resistor may actually *introduce* error.

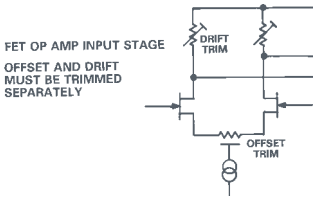
If a bipolar input stage is required without the drawback of such a high bias current, a different form of bias compensation may be used by the chip designer (next figure). The same long-tailed pair is used, but the major portion of the current required by each base is supplied by a current generator on the chip. This can reduce the external bias current to 10 nA or less without affecting the offset, temperature drift, bandwidth or voltage noise. Bias current variation with temperature is quite low.



There are two disadvantages to such an architecture: the current noise is increased and the external bias currents are not well matched (indeed, they may actually flow in opposite directions, or change polarity as chip temperature changes). For many applications these features are no drawback; indeed, one of the most popular low-offset op-amp architectures, the OP-07, uses just such an architecture, as do the OP-27, OP-37 and the AD707, which has a guaranteed offset voltage of only 15 μV. Bias-compensated amplifiers of this type are often recognizable when their data sheets explicitly specify *bipolar* bias current, for example, ±4.0 nA.

Where bias currents of even a few nanoamps are intolerable, bipolar transistors are usually replaced by field-effect devices. In the past, MOSFETs have been somewhat noisy for op-amp

input stages, although modern processing techniques are overcoming this drawback. Since MOSFETS also tend to have relatively high offset voltages, *junction* FETs (JFETs) are used for high-performance low-bias-current op amps. A typical JFET op-amp input stage is shown in this figure.



The bias current of a JFET bears no relationship to the current flowing in the device, so even a wideband JFET amplifier may have a very low bias current—values of a few tens of picoamperes are commonplace, and the AD549 has a guaranteed bias current of less than 60 fA (one electron per three microseconds!) at room temperature.

The qualification "at room temperature" is critical—the bias current of a JFET is the reverse leakage current of its gate diode, and the reverse leakage current of silicon diodes approximately doubles with every 10°C temperature rise. The bias current of a JFET op-amp is thus not stable with temperature. Indeed, between 25°C and 125°C, the bias current of a JFET op-amp increases by a factor of over 1,000. (The same law applies to MOSFET amplifiers, because the bias current of most MOSFET amplifiers is the leakage current of their gate-protection diodes.)

The offset voltage of a JFET amplifier may be trimmed during manufacture, but minimum offset does not necessarily correspond to minimum temperature drift. It has therefore been necessary to trim offset and drift separately in JFET op-amps, which results in somewhat larger values of voltage offset and drift than are available from the best bipolar amplifiers (values of 250 µV and 5 µV/°C are typical of the best JFET op-amps). Recent studies at Analog Devices, however, have resulted in a patented trimming method which is expected to yield much better values in the next generation of JFET op-amps.

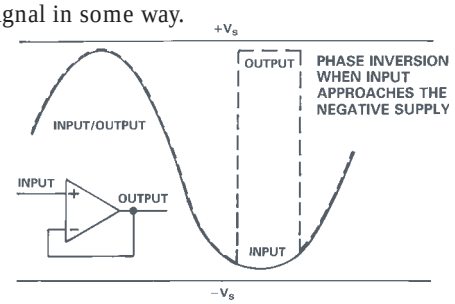
We thus see that there are trade-offs between offset voltage, offset drift, bias current, bias current temperature variation, and noise in operational amplifiers—and that different architectures optimize different features. The table compares the features of the three commonest op-amp architectures. We should note one more category, typified by the new AD705, using bipolar *superbeta* input transistors; it *combines low offset voltage and drift with low bias current and drift.*

CHARACTERISTICS OF OP-AMP INPUT STAGES

	SIMPLE BIPOLAR	BIAS-COMPENSATED BIPOLAR	FET
OFFSET VOLTAGE	LOW	LOW	MEDIUM
OFFSETDRIFT	LOW	LOW	MEDIUM
BIAS CURRENT	HIGH	MEDIUM	LOW-VERY LOW
BIAS MATCH	EXCELLENT	POOR (CURRENT CAN BE IN OPPOSITE DIRECTIONS	FAIR
BIAS/TEMP VARIATION	LOW	LOW	BIAS DOUBLES FOR EVERY 10°C RISE
NOISE	LOW	LOW	FAIR

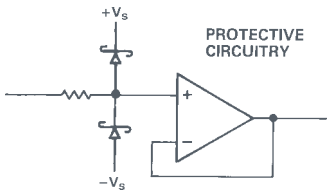
Q. What other features of op amps should the user know about?

A. A common problem encountered with JFET op-amps is phase inversion. If the input common-mode voltage of a JFET op-amp approaches the negative supply too closely, the inverting and non-inverting input terminals reverse functions. Negative feedback becomes positive feedback and the circuit may latch up. This latchup is unlikely to be destructive, but power may have to be switched off to correct it. This figure shows the effect of such phase inversion in a circuit where latch-up does not occur. The problem may be avoided by using bipolar amplifiers, or by restricting the common-mode range of the signal in some way.



A more serious form of latchup can occur in both bipolar and JFET op-amps if the input signal becomes more positive or negative than the respective op-amp power supplies. If the input terminals go more positive than +Vs + 0.7 V or more negative than –Vs – 0.7 V, current may flow in diodes which are normally biased *off.* This in turn may turn on thyristors (SCRs) formed by some of the diffusions in the op- amp, short-circuiting the power supplies and destroying the device.

To avoid such destructive latch-up it is important to prevent the input terminals of op-amps from ever exceeding the power supplies. This can have important implications during device turn-on: if a signal is applied to an op-amp before it is powered it may be destroyed at once when power is applied. Whenever there is a risk, either of signals exceeding the voltages on the supplies, or of signals being present prior to power-up of the op-amp, the terminals at risk should be clamped with diodes (preferably fast low-forward-voltage Schottky diodes) to prevent latchup from occurring. Current-limiting resistors may also be needed to prevent the diode current from becoming excessive (see the figure).



This protection circuitry can cause problems of its own. Leakage current in the diode(s) may affect the error budget of the circuit (and if glass-encapsulated diodes are used, their leakage current may be modulated at 100 or 120 Hz due to photoelectric effects if exposed to fluorescent ambient lighting, thus contributing *hum* as well as dc leakage current); Johnson noise in the current-limiting resistor may worsen the circuit's noise performance; and bias current flowing in the resistor may produce an apparent increase in offset voltage. All these effects must be considered when designing such protection.

The important subjects of noise, interference, bypassing, and grounding demand discussion—but we're out of space! We'll come back to them again in future chats; meanwhile you may want to take a look at some of the references in the footnotes on pages 193-4 of The Best of Analog Dialogue, *1967-1991.*

Ask The Applications Engineer—7

by James Bryant and Lew Counts

OP-AMP ISSUES—NOISE

Q. What should I know about op-amp noise?

A. First, we must note the distinction between noise generated in the op amp and its circuit components and *interference*, or unwanted signals and noise arriving as voltage or current at any of the amplifier's terminals or induced in its associated circuitry.

Interference can appear as spikes, steps, sine waves, or random noise, and it can come from *anywhere:* machinery, nearby power lines, r-f transmitters and receivers, computers, or even circuitry within the same equipment (for example, digital circuits or switching-type power supplies). Understanding it, preventing its appearance in your circuit's neighborhood, finding how it got in, and rooting it out, or finding a way to live with it is a big subject. It's been treated in these pages in the past; those, and a few additional references, are mentioned in the Bibliography.

If all interference could be eliminated, there would still be random noise associated with the operational amplifier and its resistive circuits. It constitutes the ultimate limitation on the amplifier's resolution. That's the topic we'll begin to discuss here.

Q. O.K. Tell me about random noise in op amps. Where does it come from?

A. Noise appearing at the amplifier's output is usually measured as a voltage. But it is generated by both voltage- and current sources. All internal sources are generally *referred to the input,* i.e., treated as uncorrelated—or *independent*—random noise generators (see next question) in series or parallel with the inputs of an ideal noisefree amplifier: We consider 3 primary contributors to noise:

- a noise *voltage* generator (like offset voltage, usually shown in series with the noninverting input)
- two *noise-current* generators pumping currents out through the two differential-input terminals (like bias current).
- If there are any resistors in the op-amp circuit, they too generate noise; it can be considered as coming from either current sources or voltage sources (whichever is more convenient to deal with in a given circuit).

Op-amp voltage noise may be lower than 1 nV/$\sqrt{\text{Hz}}$ for the best types. Voltage noise is the noise specification that is more usually emphasized, but, if impedance levels are high, current noise is often the limiting factor in system noise performance. That is analogous to offsets, where offset voltage often bears the blame for output offset, but bias current is the actual guilty party. Bipolar op-amps have traditionally had less *voltage* noise than FET ones, but have paid for this advantage with substantially greater *current* noise—today, FET op-amps, while retaining their low current noise, can approach bipolar voltage-noise performance.

Q. Hold it! 1 nV/$\sqrt{Hz}$? Where does $\sqrt{Hz}$ come from? What does it mean?

A. Let's talk about random noise. Many noise sources are, for practical purposes (i.e., within the bandwidths with which the designer is concerned), both white and Gaussian. White noise is noise whose power within a given bandwidth is independent of frequency. Gaussian noise is noise where the probability of a particular amplitude, X, follows a Gaussian distribution. Gaussian noise has the property that when the rms values of noise from two or more such sources are added, provided that the noise sources are uncorrelated (i.e., one noise signal cannot be transformed into the other), the resulting noise is not their arithmetic sum but the root of the sum-of-their-squares (RSS).* The RSS sum of three noise sources, V_1, V_2, and V_3, is

$$V_O = \sqrt{V_1^2 + V_2^2 + V_3^2}$$

Since the different frequency components of a noise signal are uncorrelated, a consequence of RSS summation is that if the white noise in a brick-wall bandwidth of Δf is V, then the noise in a bandwidth of 2 Δf is $\sqrt{V^2 + V^2} = \sqrt{2}\,V$. More generally, if we multiply the bandwidth by a factor K, then we multiply the noise by a factor $\sqrt{K}$. The function defining the rms value of noise in a Δf = 1 Hz bandwidth anywhere in the frequency range is called the (voltage or current) *spectral density function*, specified in nV/$\sqrt{\text{Hz}}$ or pA/$\sqrt{\text{Hz}}$. For white noise, the spectral density is constant; it is multiplied by the square root of the bandwidth to obtain the total rms noise.

A useful consequence of RSS summation is that if two noise sources are contributing to the noise of a system, and one is more than 3 or 4 times the other, the smaller is often ignored, since

$$\sqrt{4^2} = \sqrt{16} = 4, \text{ while } \sqrt{4^2 + 1^2} = \sqrt{17} = 4.12$$

[difference less than 3%, or 0.26 dB]

$$\sqrt{3^2} = \sqrt{9} = 3, \text{ while } \sqrt{3^2 + 1^2} = \sqrt{10} = 3.16$$

[difference less than 6%, or 0.5 dB]

The source of the higher noise has become the *dominant* source.

Q. O.K. How about current noise?

A. The current noise of simple (i.e. not bias-current-compensated) bipolar and JFET op-amps is usually within 1 or 2 dB of the Schottky noise (sometimes called the "shot noise") of the bias current; it is not always specified on data sheets. Schottky noise is current noise due to random distribution of charge carriers in the current flow through a junction. The Schottky noise current, I_n, in a bandwidth, B, when a current, I, is flowing is obtained from the formula

$$I_n = \sqrt{2I\,qB}$$

Where q is the electron charge (1.6×10^{-19} C). Note that $\sqrt{2I\,q}$ is the spectral density, and that the noise is white.

This tells us that the current noise spectral density of simple bipolar transistor op-amps will be of the order of 250 fA/$\sqrt{\text{Hz}}$, for I_b = 200 nA, and does not vary much with temperature—and that the current noise of JFET input op-amps, while lower (4 fA/$\sqrt{\text{Hz}}$ at I_b, = 50 pA), will double for every 20°C chip temperature increase, since JFET op-amps' bias currents double for every 10°C increase.

Bias-compensated op-amps have much higher current noise than one can predict from their input currents. The reason is that their net bias current is the *difference* between the base current of the input transistor and the compensating current source, while the noise current is derived from the *RSS sum* of the noise currents.

Traditional voltage-feedback op-amps with balanced inputs almost always have equal (though uncorrelated) current noise on both

[*Note the implication that noise *power* adds linearly (sum of squares).]

their inverting and non-inverting inputs. Current-feedback, or transimpedance, op-amps, which have different input structures at these two inputs, do not. Their data sheets must be consulted for details of the noise on the two inputs.

The noise of op-amps is Gaussian with constant spectral density, or "white", over a wide range of frequencies, but as frequency decreases the spectral density starts to rise at about 3 dB/octave. This low-frequency noise characteristic is known as "1/*f* noise" since the noise *power* spectral density goes inversely with frequency (actually $1/f^{\gamma}$). It has a –1 slope on a log plot (the noise *voltage* (or *current)* $1/\sqrt{f}$ spectral density slopes at –1/2). The frequency at which an extrapolated –3 dB/octave spectral density line intersects the midfrequency constant spectral density value is known as the "l/*f* corner frequency" and is a figure of merit for the amplifier. Early monolithic IC op-amps had 1/*f* corners at over 500 Hz, but today values of 20–50 Hz are usual, and the best amplifiers (such as the AD-OP27 and the AD-OP37) have corner frequencies as low as 2.7 Hz. 1/*f* noise has equal increments for frequency intervals having equal ratios, i.e., per *octave* or per *decade*.

Q. Why don't you publish a noise figure?

A. The noise figure (NF) of an amplifier (expressed in dB) is a measure of the ratio of the amplifier noise to the thermal noise of the source resistance.

$V_n = 20 \log \{ [V_n(\text{amp}) + V_n(\text{source})]/V_n(\text{source})\}$

It is a useful concept for r-f amplifiers, which are almost always used with the same source resistance driving them (usually 50 Ω or 75 Ω), but it would be misleading when applied to op amps, since they are used in many different applications with widely varying source impedances (which may or may not be resistive).

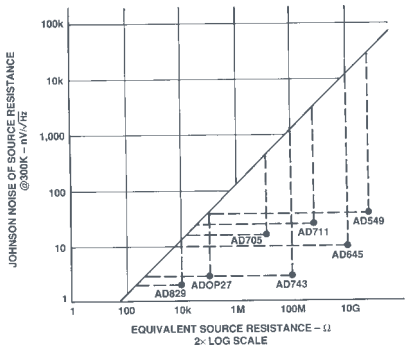
Q. What difference does the source impedance make?

A. At temperatures above absolute zero all resistances are noise sources; their noise increases with resistance, temperature, and bandwidth (we'll discuss basic resistance noise, or *Johnson noise*, in a moment). *Reactances* don't generate noise, but noise currents through them will develop noise voltages.

If we drive an op-amp from a source resistance, the equivalent noise input will be the RSS sum of the amplifier's noise voltage, the voltage generated by the source resistance, and the voltage caused by the amplifier's I_n flowing through the source impedance. For very low source resistance, the noise generated by the source resistance and amplifier current noise would contribute insignificantly to the total. In this case, the noise at the input will effectively be just the voltage noise of the op-amp.

If the source resistance is higher, the Johnson noise of the source resistance may dominate both the op-amp voltage noise and the voltage due to the current noise; but it's worth noting that, since the Johnson noise only increases with the square root of the resistance, while the noise voltage due to the current noise is directly proportional to the input impedance, the amplifier's current noise will always dominate for a high enough value of input impedance. When an amplifier's voltage and current noise are high enough, there may be no value of input resistance for which Johnson noise dominates.

This is demonstrated by the figure nearby, which compares voltage and current noise noise for several Analog Devices op amp types, for a range of source-resistance values. The diagonal line plots vertically the Johnson noise associated with resistances on the horizontal scale. Let's read the chart for the ADOP27: The horizontal line indicates the ADOP27's voltage noise level of about 3 nV/$\sqrt{\text{Hz}}$ is equivalent to a source resistance of less than about 500 Ω. Noise will not be reduced by (say) a 100-Ω source impedance, but it will be increased by a 2-kΩ source impedance. The vertical line for the ADOP27 indicates that, for source resistances above about 100 kΩ, the noise voltage produced by amplifier's current noise will exceed that contributed by the source resistance; it has become the dominant source.



Remember that any resistance in the non-inverting input will have Johnson noise and will also convert current noise to a noise voltage; and Johnson noise in feedback resistors can be significant in high-resistance circuits. All potential noise sources must be considered when evaluating op amp performance.

Q. You were going to tell me about Johnson noise.

A. At temperatures above absolute zero, all resistances have noise due to thermal movement of charge carriers. This is called Johnson noise. The phenomenon is sometimes used to measure cryogenic temperatures. The voltage and current noise in a resistance of *R* ohms, for a bandwidth of *B* Hz, at a temperature of T kelvins, are given by:

$$V_n = \sqrt{4kTRB} \text{ and } I_n = \sqrt{4kTB/R}$$

Where *k* is Boltzmann's Constant (1.38×10^{-23} J/K). A handy rule of thumb is that a 1-kΩ resistor has noise of 4 nV/$\sqrt{\text{Hz}}$ at room temperature.

All resistors in a circuit generate noise, and its effect must always be considered. In practice, only resistors in the input(s) and, perhaps, feedback, of high-gain, front-end circuitry are likely to have an appreciable effect on total circuit noise.

Noise can be reduced by reducing resistance or bandwidth, but temperature reduction is generally not very helpful unless a resistor can be made very cold—since noise power is proportional to the *absolute* temperature, T = °C + 273°.

(to be continued)

REFERENCES

Barrow, J., and A. Paul Brokaw, "Grounding for Low- and High-Frequency Circuits," *Analog Dialogue* 23-3, 1989.

Bennett, W. R., *Electrical Noise.* New York: McGraw-Hill, 1960.

Freeman, J. J., *Principles of Noise.* New York: John Wiley & Sons, Inc., 1958.

Gupta, Madhu S., ed., *Electrical Noise: Fundamentals & Sources.* New York: IEEE Press, 1977. Collection of classical reprints.

Motchenbacher, C. D., and F. C. Fitchen, *Low-Noise Electronic Design.* New York: John Wiley & Sons, Inc., 1973.

Rice, S.O., "Math Analysis for Random Noise" *Bell System Technical Journal* 23 July, 1944 (pp 282–332).

Rich, Alan, "Understanding Interference-Type Noise," *Analog Dialogue* 16-3, 1982.

- - - , "Shielding and Guarding," *Analog Dialogue* 17-1, 1983.

Ryan, Al, and Tim Scranton, "DC Amplifier Noise Revisited," *Analog Dialogue* 18-1, 1984.

van der Ziel, A. *Noise.* Englewood Cliffs, NJ: Prentice-Hall, Inc., 1954.

Ask The Applications Engineer—8

by James Bryant

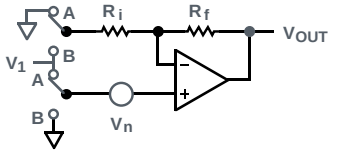
OP-AMP ISSUES

(Noise, continued from the last issue, 24-2)

Q. What is "noise gain"?

A. So far we have considered noise sources but not the gain of the circuits where they occur. It is tempting to imagine that if the noise voltage at the input of an amplifier is V_n and the circuit's signal gain is G, the noise voltage at the output will be GV_n; but this is not always the case.

Consider the basic op-amp gain circuit in the diagram. If it is being used as an inverting amplifier (B), the non-inverting input will be grounded, the signal will be applied to the free end of R_i and the gain will be $-R_f/R_i$. On the other hand, in a non-inverting amplifier (A) the signal is applied to the non-inverting input and the free end of R_i is grounded; the gain is $(1 + R_f/R_i)$



SIGNAL GAIN	AMPLIFIER NOISE GAIN
A: $\left(1+\frac{R_f}{R_i}\right)$	$\left(1+\frac{R_f}{R_i}\right)$
B: $-\frac{R_f}{R_i}$	$\left(1+\frac{R_f}{R_i}\right)$

The amplifier's own voltage noise is always amplified in the non-inverting mode; thus when an op-amp is used as an inverting amplifier at a gain of G, its voltage noise will be amplified by the noise gain of (G+ 1). For the precision attenuation cases, where G<1, this may present problems. (A common example of this is an active filter circuit where stopband gain may be very small but stop-band noise gain is at least unity.)

Only the amplifier voltage noise—and any noise developed by the noninverting-input current noise flowing in any impedance present in that input (for example, a bias-current compensation resistor)—is amplified by the noise gain. Noise in R_i, either Johnson noise or arising from inverting input noise current, is amplified by G in the same way as the input signal, and Johnson noise voltage in the feedback resistor is not amplified but is buffered to the output at unity gain.

Q. What's "popcorn" noise?

A. Twenty years ago this column would have spent a great deal of space discussing popcorn noise, which is a type of low frequency noise manifesting itself as low level (but random amplitude) step changes in offset voltage occurring at random intervals. When played through a loudspeaker it sounds like cooking popcorn—hence the name.

While no integrated circuit process is entirely free from the problem, high levels of popcorn noise result from inadequate processing techniques. Today its causes are sufficiently well understood that no reputable op-amp manufacturer is likely to produce op-amps where popcorn noise is a major concern to the user. {Oat-bran noise is more likely to be an issue in situations where cereal data is concerned[:-)]}

Q. Pk-pk noise voltage is the most convenient way to know whether noise will ever be a problem for me. Why are amplifier manufacturers reluctant to specify noise in this way?

A. Because noise is generally Gaussian, as we pointed out in the last issue. For a Gaussian distribution it is meaningless to speak of a maximum value of noise: if you wait long enough any value will, in theory, be exceeded. Instead it is more practical to speak of the rms noise, which is more or less invariant—and by applying the Gaussian curve to this we may predict the probability of the noise exceeding any particular value. Given a noise source of V rms, since the probability of any particular value of noise voltage follows a Gaussian distribution, the noise voltage will exceed a pk-pk value of 2 V for 32% of the time, 3 V for 13% of the time, and so on:

Pk-pk value	**% of time pk-pk value is exceeded**
2 × rms	32%
4 × rms	4.6%
6 × rms	0.27%
6.6 × rms	0.10%
8 × rms	60 ppm
10 × rms	0.6 ppm
12 × rms	2×10^{-9} ppm
14 × rms	2.6×10^{-12} ppm

So if we define a peak value in terms of the probability of its occurrence, we may use a peak specification—but it is more desirable to use the rms value, which is generally easier to measure. When a peak noise voltage is specified, it is frequently 6.6 × rms, which occurs no more than 0.1% of the time.

Q. How do you measure the rms value of low-frequency noise in the usually specified band, 0.1 to 10 Hz? It must take a long time to integrate. Isn't this expensive in production?

A. Yes, it is expensive, but—Although it's necessary to make many careful measurements during characterization, and at intervals thereafter, we cannot afford the time it would take in production to make an rms measurement. Instead, at very low frequencies in the 1/*f* region (as low as 0.1 to 10 Hz), the peak value is measured during from one to three 30-second intervals and must be less than some specified value. Theoretically this is unsatisfactory, since some good devices will be rejected and some noisy ones escape detection, but in practice it is the best test possible within a practicable test time and is acceptable if a suitable threshold limit is chosen. With conservative weightings applied, this is a reliable test of noise. Devices that do not meet the arbitrary criteria for the highest grades can still be sold in grades for which they meet the spec.

Q. What other op-amp noise effects do you encounter?

A. There is a common effect, which often appears to be caused by a noisy op amp, resulting in missing codes. This potentially serious problem is caused by ADC input-impedance modulation. Here's how it happens:

Many successive-approximation ADCs have an input impedance which is modulated by the device's conversion clock. If such an ADC is driven by a precision op amp whose bandwidth is much lower than the clock frequency, the op amp cannot develop sufficient feedback to provide a stiff voltage source to the ADC input port, and missing codes are likely to occur. Typically, this effect appears when amplifiers like the OP-07 are used to drive AD574s.

It may be cured by using an op amp with sufficient bandwidth to have a low output impedance at the ADC's clock frequency, or by choosing an ADC containing an input buffer or one whose input impedance is not modulated by its internal clock (many sampling ADCs are free of this problem). In cases where the op amp can drive a capacitive load without instability, and the reduction of system bandwidth is unimportant, a shunt capacitor decoupling the ADC input may be sufficient to effect a cure.

Q. Are there any other interesting noise phenomena in high-precision analog circuits?

A. The tendency of high-precision circuitry to drift with time is a noise-like phenomenon (in fact, it might be argued that, at a minimum, it is identical to the lower end of *1/f* noise). When we specify long-term stability, we normally do so in terms of μV/1,000 hr or ppm/1,000 hr. Many users assume that, since there are, on the average, 8,766 hours in a year, an instability of *x*/1,000 hr is equal to 8.8 *x*/yr.

This is not the case. Long-term instability (assuming no long-term steady deterioration of some damaged component within the device), is a "drunkard's walk" function; what a device did during its last 1,000 hours is no guide to its behavior during the next thousand. The long-term error mounts as the square-root of the elapsed time, which implies that, for a figure of *x*/l,000 hr, the drift will actually be multiplied by $\sqrt{8.766}$, or about 3× per year, or 9× per 10 years. Perhaps the spec should be in μV/1,000 $\sqrt{hr}$.

In fact, for many devices, things are a bit better even than this. The "drunkard's walk" model, as noted above, assumes that the properties of the device don't change. In fact, as the device gets older, the stresses of manufacture tend to diminish and the device becomes more stable (except for incipient failure sources). While this is hard to quantify, it is safe to say that—provided that a device is operated in a low-stress environment—its rate of long-term drift will tend to reduce during its lifetime. The limiting value is probably the 1/*f* noise, which builds up as the *square-root of the natural logarithm* of the ratio, i.e., $\sqrt{1n\ 8.8}$ for time ratios of 8.8, or 1.47 × for 1 year, 2.94 × for 8.8 years, 4.4 × for 77 years, etc.

A READER'S CHALLENGE:

Q. A reader sent us a letter that is just a wee bit too long to quote directly, so we'll summarize it here. He was responding to the mention in these columns *(Analog Dialogue* 24-2, pp. 20-21) of the shot effect, or Schottky noise (Schottky was the first to note and correctly interpret shot effect—originally in vacuum tubes[1]). Our reader particularly objected to the designation of shot noise as solely a junction phenomenon, and commented that we have joined the rest of the semiconductor and op-amp engineering fraternity in disseminating misinformation.

In particular, he pointed out that the shot noise formula—

$$I_n = \sqrt{2q\,IB}\ \text{amperes},$$

where I_n is the rms shot-noise current, I is the current flowing through a region, q *is* the charge of an electron, and B is the bandwidth—does not seem to contain any terms that depend on the physical properties of the region. Hence (he goes on) shot noise is a *universal* phenomenon associated with the fact that any current, *I*, is a flow of electrons or holes, which carry discrete charges, and the noise given in the formula is just an expression of the graininess of the flow.

He concludes that the omission of this noise component in any circuit carrying current, including purely resistive circuits, can lead to serious design problems. And he illustrates its significance by pointing out that this noise current, calculated from the flow of dc through any ideal resistor, becomes equal to the thermal Johnson noise current at room temperature when only 52 mV is applied to the resistor—and it would become the dominant current noise source for applied voltages higher than about 200 mV.

A. Since designers of low-noise op amps have blithely ignored this putative phenomenon, what's wrong? *The assumption that the above shot noise equation is valid for conductors.*

Actually, the shot noise equation is developed under the assumption that the carriers are independent of one another. While this is indeed the case for currents made up of discrete charges crossing a barrier, as in a junction diode (or a vacuum tube), it is not true for metallic conductors. Currents in conductors are made up of very much larger numbers of carriers (individually flowing much more slowly), and the noise associated with the flow of current is accordingly very much smaller—and generally lost in the circuit's Johnson noise.

Here's what Horowitz and Hill[2] have to say on the subject:

"An electric current is the flow of discrete electric charges, not a smooth fluidlike flow. The finiteness of the charge quantum results in statistical fluctuations of the current. *If the charges act independently of each other,** the fluctuating current is . . .

$$I\ noise\ (\text{rms}) = I_{nR} = (2\ qI_{dc}\ B)^{1/2}$$

where q is the electron charge (1.60×10^{-19} C) and B is the measurement bandwidth. For example, a "steady" current of 1 A actually has an rms fluctuation of 57 nA, measured in a 10-kHz bandwidth; i.e., it fluctuates by about 0.000006%. The relative fluctuations are larger for smaller currents: A "steady" current of 1 μA actually has an rms current-noise fluctuation, over 10 kHz, of 0.006%, i.e., –85 dB. At 1 pA dc, the rms current fluctuation (same bandwidth) is 56 fA, i.e., a 5.6% variation! Shot noise is 'rain on a tin roof.' This noise, like resistor Johnson noise, is Gaussian and white.

"The shot noise formula given earlier assumes that the charge carriers making up the current act independently. That is indeed the case for charges crossing a barrier, as for example the current in a junction diode, where the charges move by diffusion; *but it is not true for the important case of metallic conductors, where there are long-range correlations between charge carriers. Thus the current in a simple resistive circuit has far less noise than is predicted by the shot noise formula.** Another important exception to the shot-noise formula is provided by our standard transistor current-source circuit, in which negative feedback acts to quiet the shot noise."

*Italics ours

[1] Goldman, Stanford, *Frequency Analysis, Modulation, and Noise*. New York: McGraw-Hill Book Company, 1948, p. 352.

[2] Horowitz, Paul and Winfield Hill, *The Art of Electronics*, 2nd edition. Cambridge (UK): Cambridge University Press, 1989, pp. 431-2.

Ask The Applications Engineer—9

SEMINARS AND SUPPORT

by Chris Hyde

Q. Are performance, quality, reliability, price, and availability the only important considerations in selecting products for use in the critical portions of my designs?

A. There is one more—*support*. A manufacturer's support can be an important factor in shortening the design cycle and approaching optimal part selection—but only if you take advantage of it. Using it can make the difference between getting your product to market on time or losing the edge and market window to your competitors.

Q. What do you mean by support?

A. At Analog Devices, it basically means *help* for the designer. Its constituents include:

- (mostly) free *literature* and *software* [for example, accurate and comprehensive data sheets, data books, selection guides, tutorial and technical reference books, application notes and guides, SPICE models and other useful disk-based material, and serial publications such as *Analog Dialogue* and *DSPatch*]
- *advice* and *information* from our applications engineers, on the phone and in the field, to discuss the technical pros, cons, advantages and pitfalls in using our products to solve your design problems and selecting the right product from among the many choices available
- *samples* and *evaluation boards* from our sales and applications engineers, to try out new products—especially those at the edges of the technology—and
- *seminars*, practical tutorials in various aspects of analog-, digital-, and mixed-signal processing.

Q. That sounds like a rather full plate. What's in it for you?

A. We're really quite pragmatic. The products that we manufacture are—more often than not—state of the art and often pace the knowledge of the engineers who will benefit by applying them. It is in Analog Devices' best interest to assist these engineers in learning how and why to apply these products.

Today, designers are at a crossroads and in need of new forms of guidance. Analog Devices' unique combination of abilities in component design, processes, and functional integration, our long-cultivated capability of combining analog and digital functions on a single chip, our 25 years of experience in helping designers deal with the unique problems of transitioning between the analog and digital worlds—and now our unique contributions in digital signal processing—combine to put us in the forefront of a revolution in system design.

The integration of these capabilities shows up in both the products and the ability to provide support for customers using them to deal with the signal-conditioning chain in its entirety. The chain starts and ends with the analog signal—to condition it, convert it, process it in the digital domain, and convert the result back to analog. The physical and electrical environment is often hostile to signals, and there are many (often quite subtle) things for the designer to consider. We are in a unique position to help engineers from both analog and "mixed" signal-processing technology.

Q. I'd love to go to a real nuts-and-bolts seminar on this topic. Are you planning one?

A. You read our mind. Every year, Analog Devices sponsors a full-day technical course entitled, "[. . . subject . . .] Seminar." Presented by Analog Devices applications engineers, it is designed to assist both analog and digital designers with many of the trickier aspects of both analog and digital signal processing. The seminar combines fundamental concepts, advanced theory, and practical application. Readers who have attended Analog Devices seminars—know that it will not be a "product pitch" (but naturally our discussions of practical application will unabashedly take into account the characteristics of the devices we know best).

Q. When? Where? How much?

A. The widely advertised seminars are given in a variety of locations throughout the United States and elsewhere in the World, and portions thereafter to other groups as the opportunity arises. The minimal cost includes lunch and all materials—including a fresh and typically 500-page book of Notes. These books form a library that is available for sale. To register or obtain more information, consult the analog.com web site. It's an excellent opportunity to get a taste of Analog Devices support.

Q. What seminar books are available?

A. I thought you'd never ask. Here are the *most recent thru 1996*:

- High-speed design techniques (1996)
- Practical analog design techniques (1995)
- Linear design seminar (1994)
- System applications guide (1993)
- Amplifier applications guide (1992)
- Mixed-signal design seminar (1991)

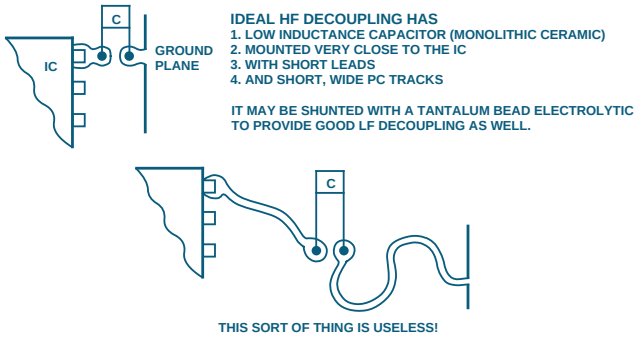
VARIOUS TOPICS

by James Bryant

Q. Tell me something about supply decoupling.

A. All precision analog integrated circuits, even low-frequency ones, contain transistors having cutoff frequencies of hundreds of MHz; their supplies must therefore be decoupled to the ground return at high frequency—as close to the IC as feasible to prevent possible instability at very high frequencies. The capacitors used for such decoupling must have low self-inductance, and their leads should be as short as possible (surface-mounted chip ceramic capacitors of 10- to 100 nF are ideal, but leaded chip ceramics are generally quite effective if the lead length is kept to less than 2 mm (see nearby figure).

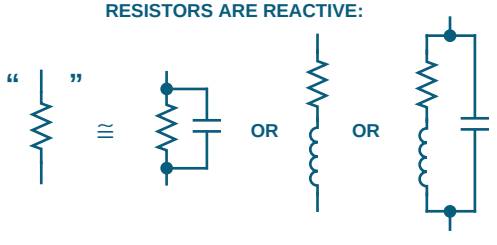
Low-frequency decoupling is also important, since the PSR (power-supply rejection) is normally specified at dc and will deteriorate appreciably with increasing power-supply ripple frequencies. In some high-gain applications, feedback through the common power-supply impedance can lead to low frequency instability ("motorboating"). However, low-frequency decoupling at each IC is not often necessary.



Supply decoupling does more than prevent instability. An op-amp is a *four-terminal* device (at least), since there must be a return path for both input signals and the output circuit. It is customary to consider the common terminal of both op-amp supplies (for op-amps using + supplies) as the output signal return path, but in fact, one of the supplies will be the de facto return path at higher frequencies, and the decoupling of the amplifier's supply terminal for this supply must take into consideration both the necessity of normal high-frequency decoupling and the routing of the output ground.*

Q. In "Ask the Application Engineer," you're always describing non-ideal behavior of integrated circuits. It must be a relief to use a simple component like a resistor and know that you have a near-ideal component.

A. I only wish that a resistor was an ideal component, and that that little cylinder with wire ends behaved just like a pure resistance. Real resistors also contain imaginary resistance components—in other words they're reactive. Most resistors have a small capacitance, typically 1-3 pF, in parallel with their resistance, although some types of film resistors, which have a spiral groove cut in their resistive film, may be inductive, with inductances of a few tens or hundreds of nH.



Of course, wirewound resistors are generally inductive rather than capacitive (at least, at the lower frequencies). After all, they consist of a coil of wire. It is commonplace for wirewound resistors to have inductances of several microhenrys or tens of microhenrys, and even so-called "non-inductive" wirewound resistors, which consist of N/2 turns wound clockwise and N/2 turns wound anticlockwise, so that the inductances of the two half windings cancel out, have a residual inductance of a microhenry or even more. (For higher-resistance-value types, above 10 kΩ or so, the residual reactance may be capacitive rather than inductive, and the capacitance will be higher—by up to 10 pF—than a standard film or composition resistor.)

These reactances must be considered carefully when designing high frequency circuits which contain resistors.

Q. But many of the circuits you describe are for making precision measurements at DC or very low frequencies. Stray inductance and capacitance don't matter in such applications, do they?

A. They actually do. Since transistors (either discrete or within ICs) have very wide bandwidths, if such circuits are terminated with reactive loads, they may sometimes oscillate at frequencies of hundreds or thousands of MHz; bias shifts and rectification associated with the oscillations can have devastating effects on low-frequency precision and stability.

Even worse, this oscillation may not appear on an oscilloscope, either because the oscilloscope bandwidth is too low for such a high frequency to be displayed, or because the scope probe's capacitance is sufficient to stop the oscillation. It is always wise to use a wideband (LF to 1.5 GHz or more) spectrum analyzer to verify the absence of parasitic oscillations in a system. Such checks should be made while the input is varied throughout its whole dynamic range, since parasitic oscillations may sometimes occur over a narrow range of inputs.

Q. Are there any problems with the resistance of resistors?

A. The resistance of a resistor is not fixed but varies with temperature. The temperature coefficient (TC) varies from a few parts per million per degree Celsius (ppm/°C) to thousands of ppm/°C. The resistors with the best stability are wirewound or metal film types, and the worst are carbon composition.

Large temperature coefficients are sometimes useful (an earlier "Ask the Applications Engineer"† mentioned how a +3,500-ppm/°C resistor can be used to compensate for the kT/q term in the equation for the behavior of a junction diode). But in general, the variation of resistance with temperature is likely to be a source of error in precision circuits.

If the accuracy of a circuit depends on the matching of two resistors having different TCs, then, no matter how well-matched at one temperature, they will not match at another; and even if the TCs of two resistors match, there is no guarantee that they will remain at the same temperature. Self-heating by internal dissipation, or external heating from a warm part of the system, will result in a mismatch of temperature, hence resistance. Even with high quality wirewound or metal-film resistors these effects can result in matching errors of several hundred (or even thousand) ppm. The obvious solution is to use resistors which are fabricated in close proximity on the same substrate whenever good matching is necessary for system accuracy. The substrate may be the silicon of a precision analog IC or a glass or metal thin-film substrate. In either case, the resistors will be well-matched during manufacture, will have well-matched TCs, and will be at nearly the same temperature because of their proximity.

(This discussion will be continued in a future issue.)

†*Analog Dialogue* 22-2, 1988. p.29.

*This issue is developed in detail in the free application note, "An IC amplifier user's guide to decoupling, grounding, and making things go right for a change," by Paul Brokaw. [AN-202]

Ask The Applications Engineer—10

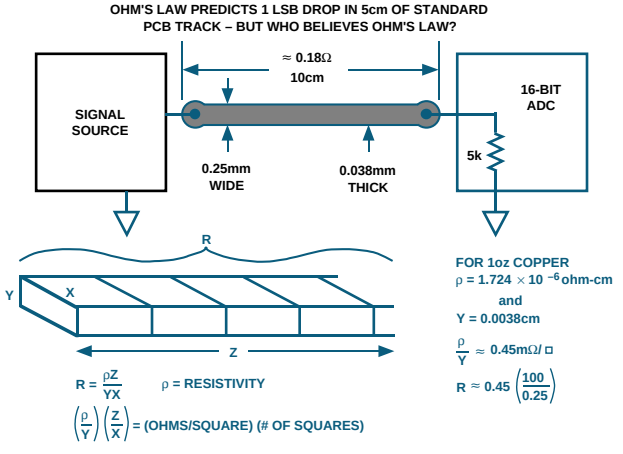
by James Bryant

Q. In the last issue of Analog Dialogue you told us about some of the problems of a simple resistor. [More will appear in a future issue.] *Surely there must be some component that behaves exactly as I expected it to. How about a piece of wire?*

A. Not even that. You presumably expect your piece of wire or length of PC track to act as a conductor. But room-temperature superconductors have not yet been invented, so any piece of metal will act as a low-valued resistor (with capacitance and inductance, too) and its effect on your circuit must be considered.

Q. Surely the resistance of a short length of copper in small-signal circuits is unimportant?

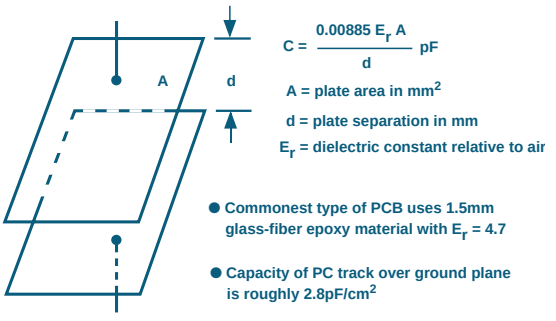
A. Consider a 16-bit a/d converter with 5-kΩ input impedance. Suppose that the signal conductor to its input consists of 10 cm of typical PC track—0.25 mm (0.010") wide and 0.038 mm (0.0015") thick. This will have a resistance of approximately 0.18 Ω at room temperature, which is slightly less than 2×2^{-16} of 5 kΩ; this introduces a gain error of 2 LSB of full scale.



One might argue that the problem would be reduced if PC tracks were made wider—and indeed, in analog circuitry it's almost always better to use wide tracks; but many layout drafters (and PC Design programs) prefer minimum-width tracks for signal conductor. In any case it's especially important to *calculate* the track resistance and its effect in every location where it might cause a problem.

Q. Doesn't the capacitance of the extra width of track to metal on the board's underside cause a problem?

A. Rarely. Although the capacitance of PC tracks is important (even in circuits designed for low frequencies, since LF circuits can oscillate parasitically at HF) and should always be evaluated, the extra capacitance of a wider track is unlikely to cause a problem if none existed previously. If it is a problem, small areas of ground plane can be removed to reduce ground capacitance.



Q. Hold it! What's a ground plane?

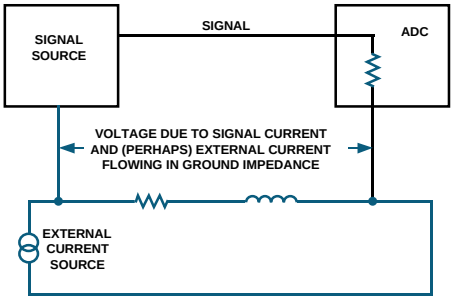
A. If one entire side of a PCB (or one entire layer, in the case of a multi-layer PCB) consists of continuous copper which is used as ground this is known as a "ground plane." It will have the least possible resistance and inductance of any ground configuration. If a system uses a ground plane, it is less likely to suffer ground noise problems.

Q. I have heard that ground planes are hard to manufacture.

A. Twenty years ago there was some truth in this. Today improvements in PC adhesives, solder resists and wave-soldering techniques make the manufacture of ground-plane PCB's a routine operation.

Q. You say that a system using a ground plane is "less likely" to suffer ground noise problems. What remaining ground noise problems does it not cure?

A. The basic circuit of a system having ground noise is shown in the diagram. Even with a ground plane the resistance and inductance will not be zero—and if the external current source is strong enough it will corrupt the precision signal.

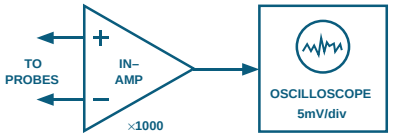


The problem is minimized by arranging the PCB so that high currents do not flow in regions where ground voltages can corrupt precision signals. Sometimes a break or slot in a ground plane can divert a large ground current from a sensitive area—but breaks in a ground plane can also reroute signals into sensitive areas, so the technique must be used with care.

Q. How do I know what voltage drops are present in a ground plane?

A. They should generally be measured; however, it is sometimes possible to calculate them from the resistance of the

ground plane material (standard 1 oz copper has resistance of 0.45 mΩ/square) and the length through which currents flow, but the calculation can be complicated. At DC and low frequencies (dc-50 kHz), voltage drops can be measured with an instrumentation amplifier such as the AMP-02 or the AD620.



The amplifier is set to a gain of 1,000 and connected to an oscilloscope with a gain of 5 mV/div. The amplifier may be powered from the same supply as the circuit being tested—or from its own supply—but the grounds of the amplifier, its supply if separate, and the oscilloscope must be connected to the power ground of the circuit under test at the power supply.

The voltage between any two points on the ground plane may then be measured by applying the probes to those points. The combination of the amplifier gain and oscilloscope sensitivity give a measurement sensitivity of 5 μV/div. Amplifier noise will swell the oscilloscope trace to a band about 3 μV wide but it is still possible to make measurements with about 1-μV resolution—sufficient to identify most low-frequency ground noise problems; and identification is 80% of a cure.

Q. Are there any cautions about performing this test?

A. Any alternating magnetic fields which thread the probe leads will induce voltages in them. This can be tested by short-circuiting the probes together (and resistively to ground to provide a bias current path) and observing the oscilloscope trace; ac waveforms observed that result from inductive pickup may be minimized by repositioning the leads or taking steps to eliminate the magnetic field. It is also essential to ensure that the ground of the amplifier is connected to the system ground; without this connection the amplifier, with no return path for bias current cannot work; grounding also ensures that this connection does not disturb the current distribution that is being measured.

Q. What about measuring HF ground noise?

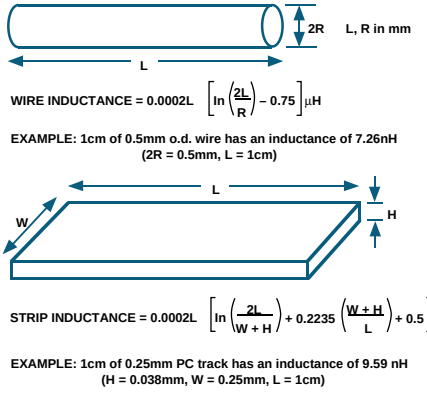
A. It is hard to make a suitable instrumentation amplifier with wide bandwidth, so at HF and VHF a passive probe is more suitable. This consists of a ferrite toroid (6-8 mm OD) wound with two coils of 6-10 turns each. One coil is connected to the input of a spectrum analyzer, the other to the probes, to make a high-frequency isolating transformer.

The test is similar to the LF one but the spectrum analyzer displays noise as an amplitude-frequency plot. While this differs from time-domain information, sources of noise may be easier to identify by their frequency signatures; in addition, the use of a spectrum analyzer provides at least 60 dB more sensitivity than is possible with a broadband oscilloscope.

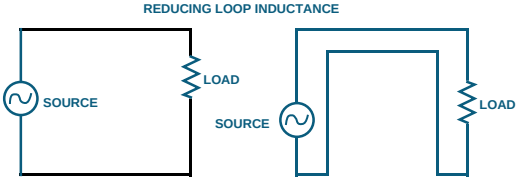
Q. What about the inductance of wires?

A. The inductance of wire- and PC-track leads should not be overlooked at higher frequencies. Here are some approximations for calculating the inductance of straight wires and runs.

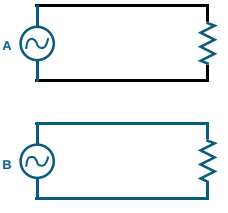
For example, 1 cm of 0.25-mm track has an inductance of 10 nH.



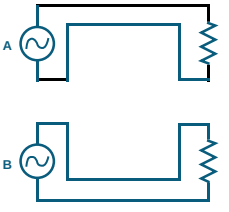
But inductive reactance is generally much less of a problem than stray flux cutting inductive loops and inducing voltages; loop area must be minimized, since voltage is proportional to it. In wired circuits this is easily done using twisted pairs.



In boards, leads and return paths should be close together; quite small changes in layout will often minimize the effect.



In this circuit, mutual inductance will couple energy from high-level source A into low-level circuit B.



Reducing area and increasing separation will minimize the effect.

Usually, all that is necessary is to minimize loop area and maximize the distance between potentially interfering loops. Occasionally magnetic shielding is required, but it is expensive and liable to mechanical damage; avoid it whenever possible.

REFERENCES

The Best of Analog Dialogue 1967-1991. Norwood MA: Analog Devices (1991), pp. 120-129, 193-195. Contains many additional references.

Mixed-Signal Design Seminar Notes. Norwood MA: Analog Devices (1991). Contains additional References.

Ask The Applications Engineer—11

by James Bryant

All analog-to-digital converters (ADCs) and digital-to-analog converters (DACs) require a reference signal, usually a voltage. The digital output of the ADC represents the ratio of the input to the reference, the digital input to a DAC defines the ratio of its analog output to its reference. Some converters have their references built-in, some require an external reference, but all must have a voltage (or current) reference of some sort.

Most early applications of data converters were in "dc" measurements of slowly varying signals, where the exact timing of the measurement was unimportant. Today most data-converter applications are in *sampled data systems,* where large numbers of equally spaced analog samples must be processed and spectral information is as important as amplitude information. Here the quality of the frequency or time reference (the "sampling clock" or "reconstruction clock") is comparable in importance to that of the voltage reference.

VOLTAGE REFERENCES

Q. How good must a voltage reference be?

A. It depends on the system. Where absolute measurements are required, accuracy is limited by the accuracy with which the reference value is known. In many systems, however, stability or repeatability are more important than absolute accuracy; and in some sampled-data systems the long-term accuracy of the voltage reference is scarcely important at all—but errors can be introduced by deriving reference from a noisy system supply.

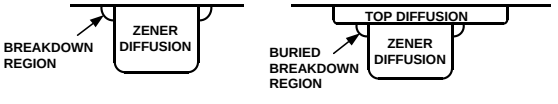
Monolithic buried-Zener references (for example the AD588 and AD688) can have initial accuracy of 1 mV in 10 V (0.01% or 100 ppm) and a temperature coefficient of 1.5 ppm/°C. They are accurate enough to use untrimmed in 12-bit systems (1 LSB = 244 ppm) but not in 14- or 16-bit systems. With the initial error trimmed to zero, they can be used in 14- and 16-bit systems over a limited temperature range. (1 LSB = 61 ppm, a 40°C temperature change in an AD588 or AD688).

For higher absolute accuracy, the temperature of the reference may need to be stabilized in a thermostatically controlled oven and calibrated against a standard. In many systems, while 12-bit absolute accuracy is unnecessary, 12-bit or higher resolution may be required; here, less accurate (and less costly) bandgap references may be used.

Q. What do you mean by "buried Zener" and "bandgap"?

A. These are the two commonest types of precision references used in integrated circuits.

The "buried" or subsurface Zener is the more stable and accurate. It consists of a diode with the correct value of reverse-breakdown voltage, formed below the surface level of the integrated-circuit chip, then covered by a protective diffusion to keep the breakdown below the surface.



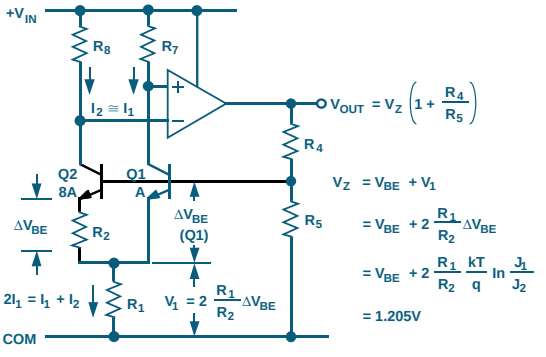
SIMPLE ZENER DIODE

BURIED (OR SUBSURFACE) ZENER DIODE

At the surface of a silicon chip there are more impurities, mechanical stresses and crystal-lattice dislocations than within the chip. Since these contribute to noise and long-term instability, the buried breakdown diode is less noisy and much more stable than surface Zeners—it is the preferred on-chip reference source for accurate IC devices.

However, its breakdown voltage is normally about 5 V or more and it must draw several hundred microamperes for optimum operation, so the technique is not suitable for references which must run from low voltage and have low power consumption. For such applications, the "bandgap" reference is preferred. It develops a voltage with a positive temperature coefficient to compensate for the negative temperature coefficient of a transistor's V_{be}, maintaining a constant "bandgap" voltage. In the circuit shown,* Q2 has 8 times the emitter area of Ql; the pair produces a current proportional to their absolute temperature (PTAT) in R1, developing a PTAT voltage in series with the V_{be} of Q1, resulting in a voltage, V_z, which does not vary with temperature and can be amplified, as shown. It is equal to the silicon bandgap voltage (extrapolated to absolute zero).

Bandgap references are somewhat less accurate and stable than the best buried-Zener references, but temperature variation of better than 3 ppm/°C may be achieved.



Q. What precautions should I take when using voltage references?

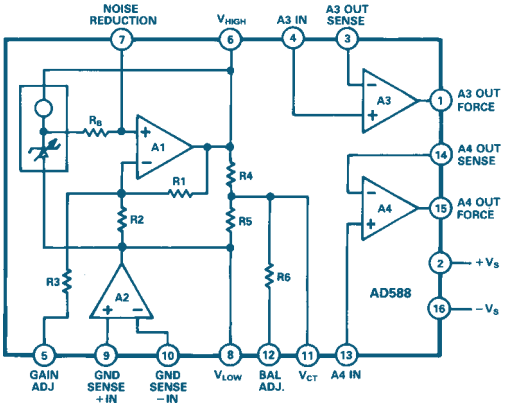
A. Remember the basics of good analog circuit design: beware of voltage drops in high impedance conductors, noise from common ground impedances, and noise from inadequately decoupled supply rails. Consider in which direction the reference current is flowing, and be careful of capacitive loads.

Q. I know about the effects of voltage drop and noise, but do references have to supply large-enough currents for voltage drop in conductors to be significant?

A. Generally, references are internally buffered; most will source and sink 5-10 mA. Some applications may require currents of this order or greater; an example is where the reference serves as the system reference; another is in driving the reference input of a high speed flash ADC which, has very low impedance. A current of 10 mA flowing in 100 mΩ will experience a voltage drop of 1 mV, which may be significant. The highest-performance voltage references, such as the AD588 and AD688, have Kelvin (force-sense) connections for both their output and output ground terminals. By closing a feedback

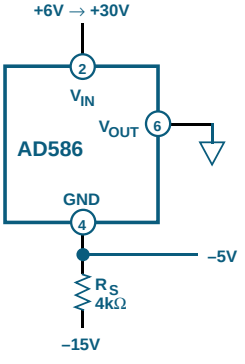
*From *Analog Dialogue* 9-1 (1975) also *The Best of Analog Dialogue,* 1967 to 1991, p. 72.

loop around the sources of error, these connections avoid the effects of voltage drops; they also correct gain and offset errors when current-buffer amplifiers are used to drive substantial loads or sink currents flowing in the wrong direction. The *sense* terminal should be connected to the output side of the buffer amplifier, preferably at the load.



Q. What do you mean by flow in the wrong direction"?

A. Consider a +5-V reference operated from a +10-V supply. If its 5-volt output terminal is loaded by a resistor to ground, current will flow out of the terminal. If the resistor is instead connected to the +10-V supply, current will flow into the terminal. Most references will allow net current flow in either direction; but some will source current but not sink it—or will sink much less than can be sourced. Such devices, identifiable by the way their output current is specified on the data sheet, may not be used in applications where substantial net current must flow *into* the reference terminal. A common example is the use of a positive reference as a negative reference.



Q. Why not just buy a negative reference?

A. Because most single voltage-output references are positive references. Two-pin active references, of course, can be used for either polarity; they are used in the same way as Zener diodes (and they are usually bandgap devices).

For a three-terminal positive reference to be used as a negative reference, it must be able to sink current. Its output terminal is connected to ground and its ground terminal (which becomes the negative-reference terminal) is connected to the negative supply via a resistor (or a constant current source). The positive supply pin must generally be connected to a positive supply at least a few volts above ground. But some devices can provide negative reference in the two-terminal mode: the positive and output terminals are connected together to ground.

R_S (or a current source) must be chosen so that for all expected values of negative supply and reference load current the ground- and output-terminal currents are within ratings.

Q. What about capacitive loads?

A. Many references have output amplifiers that become unstable and may oscillate when operated with large capacitive loads; so it is inadvisable to connect high capacitance (several µF or more) to the output of a reference to reduce noise, but 1-10 nF capacitors are often recommended—and some references (e.g., AD588) have noise-reduction terminals to which capacitance can be safely connected. If force-sense terminals are available, it may be possible to tailor loop dynamics under capacitive load. Consult data sheets and manufacturers' Application Engineers to be sure. Even if the circuit is stable, it may not be advisable to use large capacitive loads since they increase the turn-on time of the reference.

Q. Don't references turn on as soon as power is applied?

A. By no means. In many references the current that drives the reference element (Zener or bandgap) is derived from the stabilized output. This positive feedback increases *dc* stability but leads to a stable "off" state that resists startup. On-chip circuitry to deal with this and facilitate startup is normally designed to draw minimal current, so many references come up somewhat slowly (1-10 ms is typical). Some devices are indeed specified for faster turn-on; but some are even slower.

If the designer needs reference voltage very quickly after power is applied to the circuit, the reference chosen must have a sufficiently fast turn-on specification; and noise reduction capacitance should be minimized. Reference turn-on delay may limit the opportunities for strobing the supplies of data conversion systems in order to conserve system power. The problem must still be considered even if the reference is built into the converter chip; it is also important in systems of this type to consider the power-up characteristics of the converter as well [discussed in "Ask The Applications Engineer—1," *Analog Dialogue* 22-2 (1988), p. 29].

High-precision references may require an additional period of thermal stabilization after turn-on before the chip reaches thermal stability and thermally induced offsets arrive at their final values. Such effects will be mentioned on the data sheet and are unlikely to exceed a few seconds.

Q. Does using these high precision references instead of its internal reference make a converter more accurate?

A. Not necessarily. For example, the AD674B, a high-speed descendant of the classical AD574, has a factory-trimmed calibration error of 0.25% (±l0 LSB) max, with an internal reference guaranteed accurate to within +100 mV (1%). Since 0.25% of 10 V = 25 mV, full scale is 10.000 V + 25 mV. Suppose that an AD674B with a 1% high internal reference (10.1 V) had been factory-trimmed for 10.000 V full scale, by a 1% gain increase. If an accurate 10.00-V AD588 system reference were to be connected to the device's reference input, full scale would become 10.100 V, at 4 times the specified max error.

Q. Please discuss the role of the clock as a system reference.

A. Oops, we're out of space! This question introduces a topic that merits thoughtful discussion. We'll do it in a future issue.

Ask The Applications Engineer—12

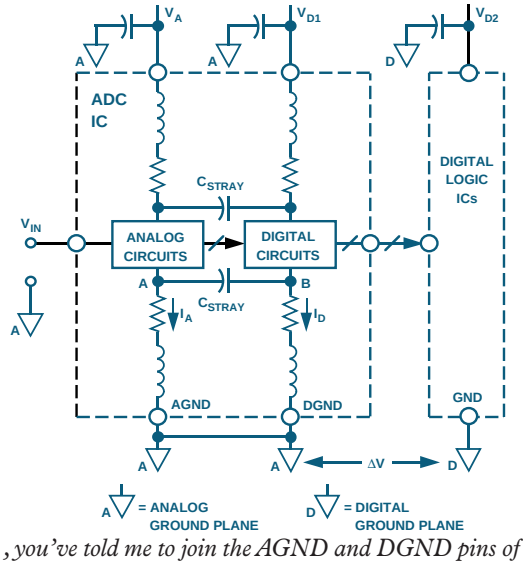
GROUNDING (AGAIN)

by Walt Kester

Q. I've read your data sheets and application notes and also attended your seminars, but I'm still confused about how to deal with analog (AGND) and digital (DGND) ground pins on an ADC. Your data sheets usually say to tie the analog and digital grounds together at the device, but I don't want the ADC to become my system's star ground point. What do I do?

A. First of all, don't feel bad that you are confused about what to do with your analog and digital grounds. So are lots of folks! Much of the confusion comes from the labeling of the ADC ground pins in the first place. The pin names, AGND and DGND, refer to what's going on inside the component itself and do not necessarily imply what you should do with them externally. Let me explain.

Inside an IC that has both analog and digital circuits, such as an ADC, the grounds are usually kept separate to avoid coupling digital signals into the analog circuits. The diagram shows a simple model of an ADC. There is really nothing the IC designer can do about the wirebond inductance and resistance associated with connecting the pads on the chip to the package pins. The rapidly changing digital currents produce a voltage at point B which will inevitably couple into point A of the analog circuits through the stray capacitance. It's the IC designer's job to make the chip work in spite of this. However, you can see that in order to prevent further coupling, the AGND and DGND pins should be joined together externally to the same low impedance ground plane with minimum lead lengths. Any extra external impedance in the DGND connection will cause more digital noise to be developed at point B; it will, in turn, couple more digital noise into the analog circuit through the stray capacitance. Though an extremely simple model, this serves to illustrate the point.



Q. O.K., you've told me to join the AGND and DGND pins of the IC together to the same ground plane—but I am maintaining separate analog and digital ground planes in my system. I want them tied together only at one point: the common point where the power supply returns are all joined together and connected to chassis ground. Now what do I do?

A. If you have only one data converter in your system, you could actually do what the data sheet says and tie your analog and digital ground systems together at the converter. Your system star ground point is now at the data converter. But this may be extremely undesirable, unless you initially planned your system with this thought in mind. If you have several data converters located on different PCBs, the concept breaks down, because the analog and digital ground systems are joined at each converter on a number of PCBs. This is a perfect invitation for ground loops!

Q. I think I've figured it out! If I must join the AGND and DGND pins together at the device, and I want to maintain separate system analog and digital grounds, I tie both AGND and DGND to either the analog ground plane or the digital ground plane on the PCB, but not both. *Right? Now, which one should it be, since the ADC is both an analog and a digital device?*

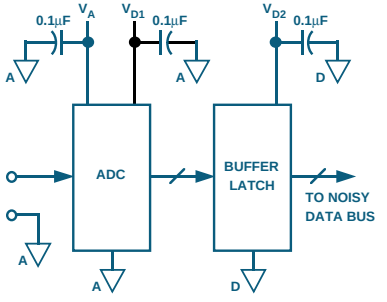
A. Correct! Now, if you connect the AGND and DGND pins both to the *digital* ground plane, your analog input signal is going to have digital noise summed with it, because it is probably single-ended and referenced to the analog ground plane.

Q. So the right answer is to connect both AGND and DGND pins to the analog ground plane? But doesn't this inject digital noise on my nice quiet analog ground plane? And isn't the noise margin of the output logic degraded because it now referenced to the analog ground plane, and all the other logic is referenced to the digital ground plane? I plan to run the ADC outputs to a backplane tristate data bus which is going to be pretty noisy to begin with so I think I need all the noise margin I can get.

A. Well, nobody ever said life was easy or fair! You have reached the right conclusion by traveling a rocky road, but the problems you suggest—digital noise on your analog ground plane and reduced noise margin on your ADC outputs—really aren't as bad as they seem; they can be overcome. It is clearly better to let a few hundred millivolts corrupt the digital interface than to apply the same corrupting signal to the analog input where the least-significant-bit for a 16-bit, 10-V-input-range ADC is only 150 µV! First of all, the digital ground currents on DGND pins can't really be that bad, or they would have degraded the internal analog parts of the ADC in the first place! If you bypass the power pins of the ADC to the analog ground plane, using a good-quality high-frequency ceramic capacitor for high frequency noise (say 0.1 µF), you will isolate these currents to a very small region around the IC, and they will have minimal effect on the rest of your system.

You will incur some reduction in digital noise margin, but it is usually acceptable with TTL or CMOS logic if it's less than a few hundred millivolts or so. If your ADC has single-ended ECL outputs, you may want to put a push-pull gate on each digital output—i.e., one with both true and complementary outputs. Tie the grounds of this gate package to the analog ground plane and connect the logic signals differentially across the interface. Use a differential line receiver at the other end which is grounded to the *digital* ground plane. The noise between the analog and digital ground planes is now common-mode—most of it will be rejected at the output of the differential line receiver. You could use the same technique with TTL or CMOS, but there is usually enough noise margin not to require differential transmission techniques.

However, one thing you said troubles me greatly. In general, it is unwise to connect the ADC outputs directly to a noisy data bus. The bus noise may couple back into the ADC analog input through the stray internal capacitance—which may range from 0.1 to 0.5 pF. It is much better to connect the ADC outputs directly to an intermediate buffer latch located close to the ADC. The buffer latch is grounded to your digital ground plane, so its output logic levels are now compatible with those of the rest of your system.



Q. I think I understand now, but why on earth didn't you just call all the ground pins of your ADC AGND in the first place; then none of this would have come up in the first place?

A. Perhaps. But what if the incoming-inspection person connects an ohmmeter between these pins and finds out that they are not actually connected together inside the package? The whole lot will probably be rejected—and the IC may be blown! Furthermore, there is a tradition associated with ADC data sheets which says we must label the pins to indicate their true function, not what we would like them to be.

Q. O. K. Now, here comes a question I've been saving as your ultimate test! I have a colleague who designed a system with separate analog and digital ground systems. My colleague says that, with the ADC's AGND pin connected to the analog ground plane and the DGND pin connected to the digital ground plane, the system is working fine! How do you explain this?

A. First of all, just because a practice is not recommended doesn't necessarily mean you can't get away with it some of the time and thereby be lulled into a false sense of security. (This is one of the lesser-known of Murphy's Laws). Some ADCs are less sensitive to external noise between the AGND and DGND pins, and it may be that your colleague picked one of those by accident. There could be other explanations—which would require that we explore your colleague's definition of "working fine"—but the point is that the ADC's specifications are not guaranteed by the manufacturer under those operating conditions. With a complex component like an ADC, it is impossible to test the device under all possible operating circumstances, especially those which aren't recommended in the first place! Your friend got lucky this time, but you can be sure that Murphy's law will ultimately catch up with him (or her) if this practice is continued in future system designs.

Q. I think I understand the ADC grounding philosophy now, but what about DACs?

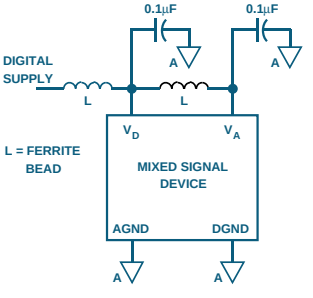
A. The same philosophy applies. The DAC's AGND and DGND pins should be tied together and connected to the analog ground plane. If the DAC has no input latches, the registers driving the DAC should be referenced and grounded to the analog ground plane to prevent digital noise from coupling into the analog output.

Q. What about mixed-signal chips which contain ADCs, DACs, and DSPs such as your ADSP-21msp5O voiceband processor?

A. The same philosophy applies. You should never think of a complex mixed-signal chip, such as the ADSP-21msp50, as being only a digital chip! The same guidelines we've just been discussing should be applied. Even though the effective sampling rate of the 16-bit sigma-delta ADC and DAC is only 8 ksps, the converters operate at an oversampling frequency of 1 MHz. The device requires an external 13-MHz clock, and an internal 52- MHz processor clock is generated from it with a phase-locked loop. So you see, successful application of this device requires an understanding of design techniques for both precision- and high-speed circuits.

Q. What about the analog and digital power-supply requirements of these devices? Should I buy separate analog and digital power supplies or use the same supply?

A. This really depends on how much noise is on your digital supply. The ADSP-21msp50, for example, has separate pins for the +5-V analog supply and the +5-V digital supply. If you have a relatively quiet digital supply, you can probably get away with using it for the analog supply too. Be sure to properly decouple each supply pin at the device with a 0.1-µF ceramic capacitor. Remember to decouple to the analog ground plane, not the digital ground plane! You may also want to use ferrite beads for further isolation. The diagram below shows the proper arrangement. A much safer solution is to use a separate +5-V analog supply. You can generate the +5 V from a quiet +15-V or +12-V supply using a three-terminal regulator, if you can tolerate the extra power dissipation.



REFERENCES [*not available from Analog Devices unless noted*]

1. Ralph Morrison, *Grounding and Shielding Techniques in Instrumentation,* Third Edition. New York: Wiley-lnterscience, 1986.
2. Henry W. Ott, *Noise Reduction Techniques in Electronic Systems,* Second Edition. New York: Wiley-Interscience, 1988.
3. *High-Speed Design Seminar,* 1996. Norwood MA: Analog Devices, Inc.
4. *Mixed-Signal Design Seminar,* 1991. Norwood MA: Analog Devices, Inc.
5. Paul Brokaw, "An I.C. Amplifier User's Guide to Decoupling, Grounding and Making Things Go Right for a Change." AN-202. Free from Analog Devices.
6. Jeff Barrow, "Avoiding Ground Problems in High Speed Circuits," *R.F. Design,* July, 1989.
7. Paul Brokaw and Jeff Barrow, "Grounding for Low- and High-Frequency Circuits," *Analog Dialogue* 23-3, 1989. Free from Analog Devices.
8. *The Best of Analog Dialogue—1967-1991.* Norwood MA: Analog Devices, Inc. Free from Analog Devices.

ODDS 'N' ENDS *(Continued from earlier issues)*

by James Bryant

TIME REFERENCES *(continued from 26-1—AA-11)*

Q. Why do you say that the clock of a system is a reference?

A. This comment does not necessarily apply to the conversion clock of an ADC; it applies principally to the sampling clock of a sampled-data system. In these systems, the signal is required to be sampled repeatedly at predictable (usually equal) intervals for storage, communication, computational analysis, or other types of processing. The quality of the sampling clock is a system-performance-limiting factor.

Q. But crystal oscillators are very stable, aren't they?

A. They have good long-term stability, but they are often used in ways which introduce short-term phase noise. Phase noise is also introduced by designers who, instead of using crystal oscillators, use R-C relaxation oscillators (such as the 555 or the 4046)—which have a great deal of phase noise.

Q. How can I ensure that my sampling clock has low phase noise?

A. Don't use the crystal oscillator circuitry in your microprocessor or DSP processor as the source of your sampling clock. If at all possible, do not use a logic gate in a crystal oscillator. Crystal oscillators made with logic gates generally overdrive the crystal; this is bad for its long term stability, and usually introduces worse phase noise than would a simple transistor oscillator. In addition, digital noise from the processor—or from other gates in the package if a logic gate is used as an oscillator—will appear as phase noise on the oscillator output.

Q. But crystal oscillators are very stable, aren't they?

A. Ideally, use a single transistor or FET as your crystal oscillator and buffer it with a logic gate. This logic gate, and the oscillator itself, should have a well-decoupled supply; the other gates in the package should not be used because logic noise from them will phase-modulate the signal. (They may be used for dc applications but not for fast-switching operations.)

If there is a divider between the crystal oscillator and the sampling clock input of the various ADCs, the divider power supply should be decoupled separately from the system logic to keep power supply noise from phase-modulating the clock.

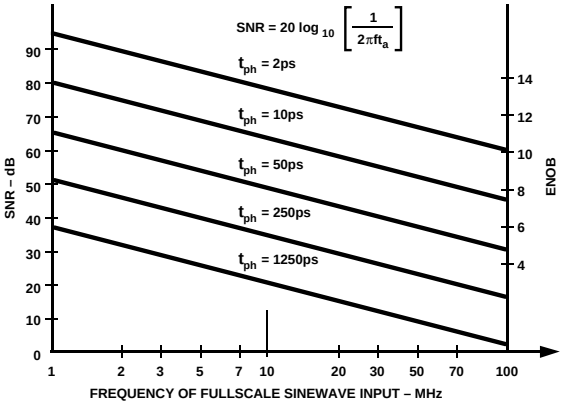
The sampling clock line should be kept away from all logic signals to prevent pickup from introducing phase noise. Equally, it should be kept away from low-level analog signals lest it corrupt them.

Q. You have told me not to use the clock oscillator of my processor as the sampling clock source. Why not? Isn't it sensible to use the same oscillator for both, since there will then be a constant phase relationship between the signals?

A. True. But in such cases, it is often better to use a single discrete low-noise oscillator to drive the processor clock input and the sampling clock divider *through separate buffers* (though they may share a package) than to use the oscillator in the processor. In medium-accuracy systems with low sampling rates it may be possible to use the processor's internal oscillator—but check with the diagram below).

Q. Just how serious is this problem of noise on a sampling clock? I hardly ever see it mentioned in articles on sampled data systems.

A. The phase noise of the sampling clock is often ignored, because the limiting factor on system performance used to be the aperture jitter of the of the sample-hold—but if we consider the system as a whole, aperture jitter is just one component of the total phase noise in the sampling clock chain. With modern sampling ADCs the aperture jitter may be less important than other components of phase noise.



The diagram shows the effect of the total phase jitter of the sampling clock on signal-to-noise ratio (SNR) or effective number of bits (ENOB). This jitter has the rms value of t_{ph}, which is made up of the root-sum-of-squares of the phase jitter on the sampling clock oscillator, the phase jitter introduced by pickup during transmission of the sampling clock through the system, and the aperture jitter of the SHA in the sampling ADC. This diagram may be somewhat unsettling, as it shows just how little phase noise is required to corrupt a high-resolution sampled-data system.

MORE ON TRIMMING

Q. I don't have enough range to adjust the offset of my circuit—and it seems to have rather more drift than I'd expected.

A. I'll bet the amplifier is a bipolar type and you are using its offset-trim terminals to trim other circuit voltages.

Q. How did you guess?

A. The range of offset adjustment of an op amp is normally 2 to 5 times the maximum expected offset of the lowest grade of the device (in some early op amps, it was much larger, but such a wide range is not ideal). If the lowest grade has a V_{OS} (max) of ±1 mV, then the likely adjustment range with the recommended circuit is ±2 to ±5 mV.

If the external voltage you are attempting to compensate for is larger than this (referred to the op amp's input), you will not be able to do so with the amplifier's offset-trim terminals.

Furthermore, if you are using a *bipolar-input* op amp, it is inadvisable to use these terminals for external offset correction because drift will be increased. Here's why: the input stage thermal drift is proportional to the internal offset; if this has been trimmed to a minimum, the drift will also be a minimum. If you then trim the amplifier to compensate for an external offset, drift will no longer be minimized. However, *FET-input* op amps have separately trimmed offset and drift, their offset adjustment terminals may thus be used for small system adjustments.

Ask The Applications Engineer—13

CONFUSED ABOUT AMPLIFIER DISTORTION SPECS?

by Walt Kester

Q. *I've been looking at your amplifier data sheets and am confused about distortion specifications. Some amplifiers are specified in terms of second- and third-harmonic distortion, others in terms of* total harmonic distortion *(THD) or* total harmonic distortion plus noise *(THD+N), still others have some of these specifications as* well as two-tone intermodulation distortion *and* third-order intercept. *Can you please clarify?*

A. Because the amplifier is fundamental to a wide range of uses, it is natural that many application-specific specifications have evolved as new amplifiers have been developed to meet those needs. So—as you so rightly pointed out—distortion may be specified in various ways; the spec depends on how distortion is defined by users for the particular application. Some distortion specifications are fairly universal, while others are primarily associated with specific frequency ranges and applications.

But there is some standardization of the basic definitions, so let's talk about them first. Harmonic distortion is measured by applying a spectrally pure sine wave to an amplifier in a defined circuit configuration and observing the output spectrum. The amount of distortion present in the output is usually a function of several parameters: the small- and large-signal nonlinearity of the amplifier being tested, the amplitude and frequency of the input signal, the load applied to the output of the amplifier, the amplifier's power supply voltage, printed circuit-board layout, grounding, power supply decoupling, etc. So you can see that any distortion specification is relatively meaningless unless the exact test conditions are specified.

Harmonic distortion may be measured by looking at the output spectrum on a spectrum analyzer and observing the values of the second, third, fourth, etc., harmonics with respect to the amplitude of the fundamental signal. The value is usually expressed as a ratio in %, ppm, dB, or dBc. For instance, 0.0015% distortion corresponds to 15 ppm, or –96.5 dBc. The unit "dBc" simply means that the harmonic's level is so many dB below the value of the "carrier" frequency, i.e., the fundamental.

Harmonic distortion may be expressed individually for each component (usually only the second and third are specified), or they all may be combined in a root-sum-square (RSS) fashion to give the *total harmonic distortion* (THD).

$$THD = \frac{\sqrt{V_2^2 + V_3^2 + V_4^2 + \cdots + V_n^2}}{V_s}$$

where

V_s = signal amplitude (rms volts)
V_2 = second harmonic amplitude (rms volts)
V_n = *nth* harmonic amplitude (rms volts)

The number of harmonics included in the THD measurement may vary, but usually the first five are enough. You see, the RSS process causes the higher-order terms to have negligible effect on the THD, if they are 3 to 5 times smaller than the largest harmonic $[\sqrt{0.10^2 + 0.03^2} = \sqrt{0.0109} = 0.104 \approx 0.10]$.

The expression for THD+N is similar; simply add the noise in root-sum-square fashion (V_{noise} = rms value of noise voltage over the measurement bandwidth).

$$THD = \frac{\sqrt{V_2^2 + V_3^2 + V_4^2 + \cdots + V_n^2 + V_{noise}^2}}{V_s}$$

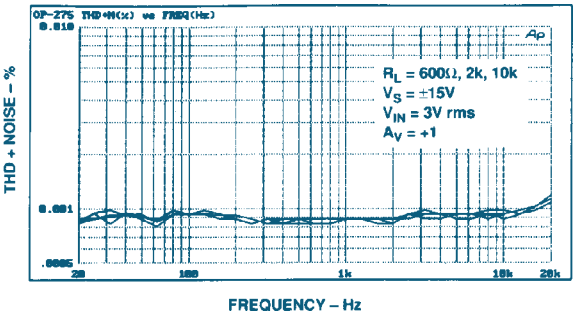
It should be evident that THD+N ≈ THD if the rms noise over the measurement bandwidth is several times less than the THD, or even the worst harmonic. It is worth noting that if you know only the THD, you can calculate THD+N fairly accurately using the amplifier's voltage- and current-noise specifics. (Thermal noise associated with the source resistance and the feedback network may also need to be computed). But if your rms noise level is significantly higher than the level of the harmonics, and you are only given the THD+N specification, you cannot compute the THD.

Special equipment is often used in audio applications for a more-sensitive measurement of the noise and distortion. This is done by first using a bandstop filter to remove the fundamental signal. The total rms value of all the other frequency components (harmonics and noise) is then measured over an appropriate bandwidth. The ratio to the fundamental is the THD + N spec.

Q. *How are the distortion specs looked at over the various frequency ranges and applications?*

A. I think the best way is to start at the low frequency end of the spectrum and work our way up; that will make it easier to see the underlying method.

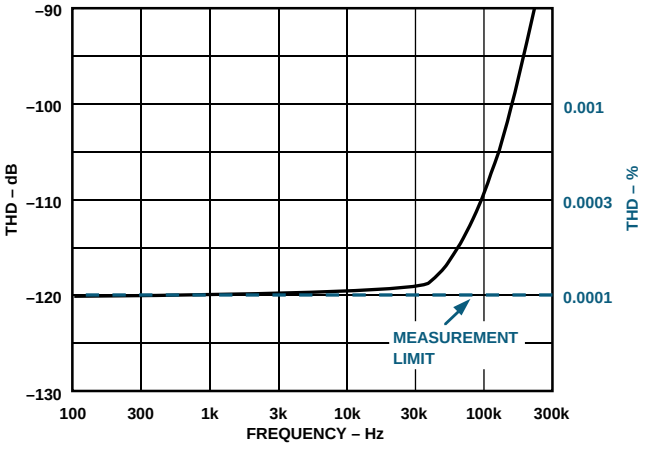
Audio-frequency amplifiers are a good place to start. Types used here (such as the OP-275*) are optimized for low noise and low distortion within the audio bandwidth (20 Hz to 20 kHz). In audio applications, total harmonic distortion plus noise (THD+N) is usually measured with specialized equipment such as the Audio Precision System One. The output signal amplitude is measured at a given frequency (e.g., 1 kHz); then, as above, the fundamental signal is removed with a bandstop filter and the system measures the rms value of the remaining frequency components, which contain both harmonics and noise. The noise and harmonics are measured over a bandwidth that will catch the highest harmonics, usually about 100 kHz. The measurement is swept over the frequency range for various conditions. THD+N results for OP-275 are plotted here as a function of frequency.



The signal level is 3 V rms, and the amplifier is connected as a unity-gain follower. Notice that a THD+N value of 0.0008% corresponds to 8 ppm, or –102 dBc. The input voltage noise of the OP-275 is typically 6 nV/√Hz @ 1 kHz and, integrated over a 100-kHz bandwidth, yields an rms noise level of 1.9 μV rms. For a 3-V rms signal level, the corresponding signal-to-noise ratio is 124 dB. Because the THD is considerably greater than the noise level, the THD component is the primary contributor.

Q. *I noticed that Analog Devices recently introduced another low-noise, low-distortion amplifier (AD797) and that it is specified in THD, not THD+N. The actual specification quoted at 20 kHz is –120 dB. What gives?*

A. Actually, we are not trying to be misleading here. The distortion is at the limits of measurement of the available equipment, and the noise is even lower—by 20 dB! Here is the measured THD of the AD797 as a function of frequency.

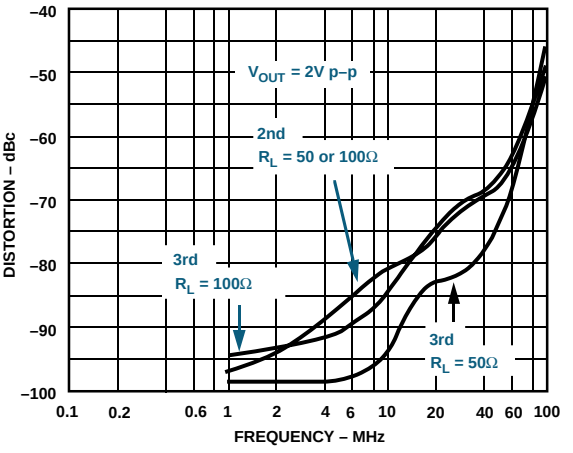


The measurement was made with a spectrum analyzer by first filtering out the fundamental sine-wave frequency ahead of the analyzer. This is to prevent overdrive distortion in the spectrum analyzer. The first five harmonics were then measured and combined in a root-sum-square fashion to get the THD figure. The legend on the graph indicates that the measurement-equipment "floor" is about –120 dB; hence at frequencies below 10 kHz, the THD may be even less.

For noise, multiply the voltage noise spectral density of the AD797 (1 nV/√Hz) by the square root of the measurement bandwidth to yield the device's rms noise floor. For a 100-kHz bandwidth, the noise floor is 316 nV rms, corresponding to a signal-to-noise ratio of 140 dB for a 3-V rms output signal.

Q. *How is distortion specified for high frequency amplifiers?*

A. Because of the increasing need for wide dynamic range at high frequencies, most wideband amplifiers now have distortion specifications. The data sheet may give individual values for the second and third harmonic components, or it may give THD. If THD is specified, only the first few harmonics contribute significantly to the result. At high frequencies, it is often useful to show the individual distortion components separately rather than specifying THD. The AD9620 is a 600-MHz (typical –3-dB bandwidth) low distortion unity-gain buffer. Here are graphs of the AD9620's second and third harmonic distortion as a function of frequency for various loading conditions.

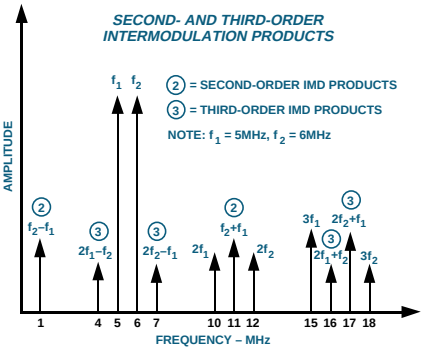


Q. *What are two-tone intermodulation products, and how do they differ from harmonic distortion?*

A. When two tones are applied to an amplifier that is non-linear, the nonlinearity causes them to modulate one another, producing intermodulation distortion (IMD) in the form of frequencies known as intermodulation products. (For the mathematical development of this concept, see Reference 1). For two tones at frequencies, f_1 and f_2 (where $f_2 > f_1$), the second- and third-order IM products occur at the following frequencies:

Second Order: $f_1 + f_2$, $f_2 - f_1$
Third Order: $2f_1 + f_2$, $2f_2 + f_1$, $2f_2 - f_1$, $2f_1 - f_2$

If the two tones are fairly close together, the third-order IMD products at the difference frequencies, $2f_2 - f_1$ and $2f_1 - f_2$, may be especially troublesome because—as the figure shows —they are hard to filter out. Notice that the other second- and third-order IMD products—which occur at substantially higher or lower frequencies—can be filtered (if the only frequencies of interest are in the neighborhood of f_1 and f_2).

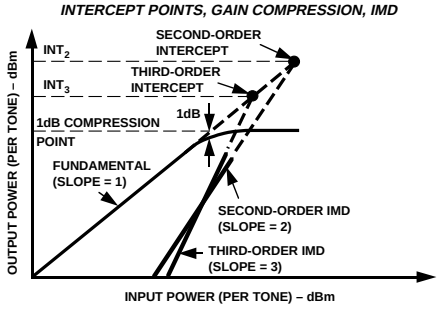


Two-tone intermodulation-distortion specifications are of especial interest in r-f applications and are a major concern in the design of communications receivers. IMD products can mask out small signals in the presence of larger ones. Although IMD has been rarely specified in op amps operating at frequencies less than 1 MHz, many of today's dc op amps are wideband types that can operate usefully at radio frequencies. For this reason, it is becoming common to see IMD specifications on fast op amps.

Q. *What are the second- and third-order intercept points, and what is their significance?*

A. Usually associated with r-f applications, these specs provide figures of merit to characterize the IMD performance of the amplifier. The higher the intercept power, the higher the input level at which IMD becomes significant—and the lower the IMD at a given signal level.

Here's how it is derived: Two spectrally pure tones are applied to the amplifier. The output signal power in a single tone (in dBm) and the relative amplitudes of the second-order and third-order products (referenced to a single tone) are plotted (and extrapolated) here as a function of input signal power.



If you go through the mathematical analysis [1], you will find that if device nonlinearity can be modeled by a simple power-series expansion, the second-order IMD amplitudes tend to increase by 2 dB for every 1 dB of signal increase. Similarly, the third-order IMD amplitudes increase 3 dB for every 1 dB of signal increase. Starting with a low-level two-tone input signal and taking a few IMD data points, you can draw (and extrapolate) the second- and third-order IMD lines shown on the diagram.

Beyond a certain level, the output signal begins to soft-limit, or compress (coinciding with the increasing visibility of IMD products). If you extend the second- and third-order IMD lines, they will intersect the extension of the output/input line; these intersections are called the second- and third-order intercept points. The projected output power values corresponding to these intercepts are usually referenced to the output power of the amplifier in dBm.

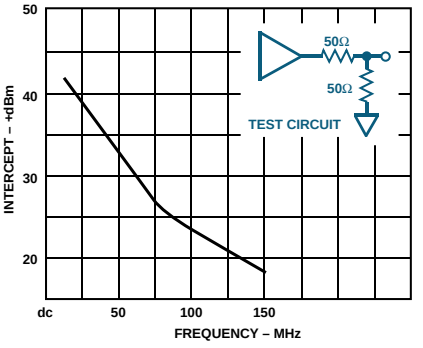
Since the slope of the third-order IMD amplitudes is known (3 dB/dB), if the intercept is also known, the third-order products at any input (or output) level can be approximated. For a higher intercept, the line moves to the right (same slope), showing lower 3rd-order products for a given input level. Many r-f mixers and "gain blocks" have 50-Ω input and output impedances. The output power is simply the power that the device transfers to a 50-Ω load. The output power is calculated by squaring the rms output voltage (V_o) and dividing by the load resistance, R_L. The power is converted into dBm as follows:

$$\text{Output power} = 10 \log_{10} \frac{\frac{V_0^2}{R_L}}{1\,\text{mW}}\,\text{dBm}$$

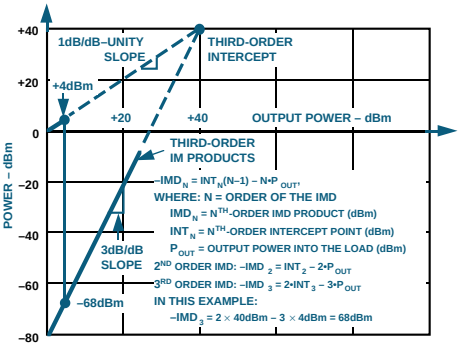
Since an op amp, on the other hand, is a low-output-impedance device, for most r-f applications, the output of the op amp must be source- and load-terminated. This means that the actual op amp output power has to be 3 dB higher than the power delivered to the load, as calculated from the above formula. In this type of application, it is customary to define the IMD products with respect to the *output power actually delivered into the 50-Ω load* rather than the actual op-amp output power.

Another parameter that may be of interest is the 1-dB compression point, shown in the figure. This is the point at which the output signal has started to limit and is attenuated by l dB from the ideal input/output transfer function.

The figure below is a plot of the third-order intercept power values for the AD9620 buffer amplifier as a function of input frequency. Its data can be used to approximate the actual value of the third order intermodulation products at various frequencies and signal levels.



Assume the op amp output signal is at 20 MHz with 2 V peak-to-peak into a 100-Ω load (50-Ω source and load terminations). The voltage into the 50-Ω load is therefore 1 volt peak-to-peak, with a power of 2.5 mW, corresponding to +4 dBm. The value of the third-order intercept at 20 MHz—from the graph—is +40 dBm. This permits a graphical solution, as shown below. For an output level of +4 dBm, the third-order IMD products, based on an extrapolation of the slope of 3 back from the intercept, amount to –68 dBm, or 72 dB below the signal.



This analysis assumes that the op-amp distortion can be modeled with a simple power series expansion as described in Reference 1. Unfortunately, op amps don't always follow simple models (especially at high frequencies), so the third-order intercept specification should primarily be used as a figure of merit, rather than a substitute for measurements. ▶

REFERENCES

1. Robert A. Witte, "Distortion Measurements Using a Spectrum Analyzer," *RF Design*, September 1992, pp. *75-84. (not available from ADI)*
2. *High Speed Design Seminar*, 1996. Norwood, MA: Analog Devices, Inc.
3. 1992 *Amplifier Applications Guide*. Norwood, MA: Analog Devices, Inc.

Ask The Applications Engineer—14

by James Bryant (ADI Europe)
with Herman Gelbach (The Boeing Company)

HIGH-FREQUENCY SIGNAL-CONTAMINATION

Q. I've heard that RF can make low-frequency circuits do strange things. What's that all about?

A. I was once summoned to France because an Analog Devices Voltage-Frequency Converter (VFC), the AD654, suffered from "unacceptable variation of accuracy." I had measured the offending parts in my own laboratory and found them to be stable and within specification, but when I returned them to the customer with my test jig he was unable to reproduce my results. While considering a site visit to confirm my suspicions, I discovered that the restaurant "La Cognette" in the town where our customer was located had three stars in the *Guide Michelin,* and the chef was a "Maitre Cuisinier de France"—a title not lightly bestowed. The visit to the customer became doubly necessary. Herman, who was in England to look at data offsets in a Boeing wind tunnel test, offered to come and help—he said it was the interesting technical problem (but just before he offered I saw him earnestly consulting the *Guide Michelin)*.

To drive from the Analog Devices office in Newbury in the South of England to the centre of France involves six hours of driving, a six hour ferry crossing of the English Channel, and a change from the correct side to the right side of the road. Nevertheless, driving is better than flying, because one can take more test gear (and the portable ham-radio station as well—we are both hams).

As we approached the customer's works we passed an enormous short-wave transmitting antenna, and then another, and yet another. We began to guess what might be wrong, and when we entered the laboratory I was carrying a hand-portable two-metre ham transceiver (an HT or "handy-talky") in my jacket pocket.

The AD654 was indeed behaving unstably, as the customer had claimed. The VFC's output frequency varied by an equivalent offset of tens of mV over the space of a few minutes. I quietly reached into my pocket and pressed the transmit button of my HT. The output frequency jumped by an equivalent of 150 mV, thus demonstrating the problem to be high-frequency pickup. More-formal measurements a little later showed that the local transmitters (of the French Overseas Broadcast organization) produced high-frequency (HF) field strengths within our customer's works of tens or hundreds of mV/m.

Many problems of instability in precision measurement circuitry can be traced to high-frequency interference, but unless there is a loudspeaker in the system that might unexpectedly burst into hard rock music from the nearby radio station, it is common for engineers to overlook this source of inaccuracy and blame the manufacturer of the amplifiers or data converters.

Furthermore, this case was unusual in that it took a high-powered signal to affect the AD654, which is single-ended and also relatively insensitive to RF—it is much more common to see with a differential amplifier in-amp. Both inputs of these types of amplifier have high input impedances to common; they are therefore far more vulnerable and are affected by low-level RF, such as radiation from a personal computer (PC). [This phenomenon is detailed in the Analog Devices system-design seminar notes, available for sale as *System Application Guide (*1993).]

An important factor is that, in instrumentation amplifiers, common-mode rejection decreases with increasing frequency, starting to roll off at quite low frequencies—and distortion increases with frequency. Thus, not only are high-frequency common-mode signals not rejected; they are distorted, producing offsets. For some applications, where RF interference is a strong possibility, the AD830 difference amplifier has wideband common-mode rejection and is designed for line-receiver applications; it may be a useful substitute for an instrumentation amplifier.

Sensors are often connected to their signal-conditioning electronics by long cables. Radio engineers have a term for such long pieces of wire; they call them *antennas.* The long feeders from sensors to their electronics will behave in the same way and will serve as antennas, even if we do not wish them to do so. It does not matter if the sensor case is grounded—at high frequencies the reactances of the case and feeders will allow the system to behave as an antenna, and any high- frequency signals (E-field, M-field, or E-M-field) which it encounters will appear across any impedances. The most likely place for them to end up is at the amplifier input. Precision low-frequency amplifiers can rarely cope with large HF signals, and the result is error—commonly a varying offset error.

Q. But this couldn't happen to me!

A. Never believe it won't happen to you! An easy free lunch can always be obtained by persuading an innocent to bet on his or her circuit being free of such problems. Using a ham radio HT on the two-metre (144-148-MHz) band, one watt at a distance of one meter for one second will win you your free lunch almost every time. But a less-dramatic test can be equally convincing.

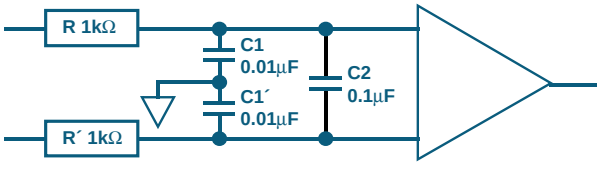
Disconnect the sensor and its leads. Short-circuit the amplifier input terminals to each other and to the amplifier circuit common (probably ground) with the shortest possible links and measure the amplifier output; observe its stability over a few minutes. Now remove the short-circuit, replace the sensor leads and place them in their normal operating environment. Disable the excitation and short-circuit the signal leads at the sensor end. Again measure the amplifier output, and its variation with time. Weep quietly.

It is often possible to see what is happening by using a high-frequency oscilloscope (or a spectrum analyzer, which is more sensitive but less easy to interpret) to measure the HF noise, both normal mode and common-mode, at the amplifier input; but normal mode measurements must be treated with some suspicion, because the oscilloscope itself—and its power- and probe leads—may themselves introduce signals and invalidate the measurement. The effect of the oscilloscope may be minimized by using a simple broadband transformer between the measurement point and the oscilloscope input, as shown in the figure; but such a transformer has fairly low impedance and will load the circuit being measured.

Common-mode signals can be observed quite easily by disabling any sensor excitation and connecting the oscilloscope ground to the ground at the board input and joining all the sensor leads together and to the oscilloscope input. All too often this signal will have an amplitude of several hundred millivolts and contain components from low frequencies to tens or hundreds of MHz.

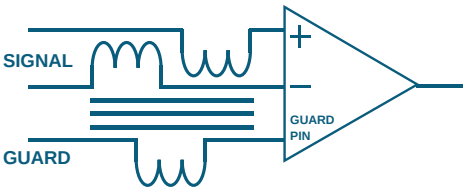
The world is full of HF noise sources: ham radio operators, police, people with portable phones, garage door openers, the sun, supernovas, switching power-supply and logic signals (e.g., PCs). Since we cannot eliminate HF noise in the environment, we must filter it out of low-frequency signals before they arrive at precision amplifiers.

The simplest type of protection can be used when the signal bandwidth is only a few Hz. A simple RC low-pass filter inserted ahead of an amplifier will afford both normal-mode and common-mode HF protection. A suitable circuit is shown in the figure. There are two important issues to be considered in the choice of components: the resistances R and R' (shown as 1 kΩ in the diagram, a value suitable for amplifier bias currents of a few nA or less) must be chosen so that they do not increase the offset appreciably as the amplifier bias current flows in them. The normal-mode time constant, $(R + R')C_2$, must be much larger than the common-mode time constants, RC_1 and $R'C_1'$, otherwise the common-mode time constants would have to be very carefully matched to avoid an imbalance that would convert the common-mode to a signal between the differential inputs.



If the signal bandwidth is wider, such simple filters will not be suitable because they remove the desired HF normal-mode signals as well as the unwanted HF common-mode signals. Large HF common-mode signals are very likely to suffer common-mode→normal mode conversion (as well as minor rectification, producing low-frequency errors) if they get to the amplifier, so it is necessary to use a filter which will reject HF common-mode signals but will pass DC and HF normal-mode signals.

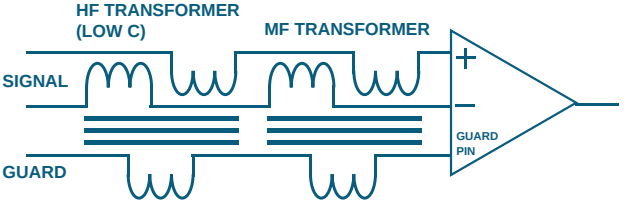
Such a filter is shown below. It was devised many years ago by Bill Gunning of Astrodata and is related to the "phantom circuit" used in long-distance telephone circuits. It uses a tightly coupled "trifilar" transformer having three windings in an accurate 1:1:1 ratio. An AC voltage across any winding will also be present on the others.



The guard line is connected to ground at the source end and at the other end to the amplifier's guard pin (or a comparable derived voltage), which represents what the amplifier "thinks" is common mode, via a capacitor. The high-frequency common-mode signal will appear (by definition) across the bottom winding, and will induce an equal common-mode voltage in the other two, subtracting the common-mode voltage in series with each line and effectively cancelling the HF common-mode signal at the amplifier inputs.

There are, of course, potential problems. A capacitor in series with the transformer is almost essential in the guard circuit to block DC and LF and prevent transformer core saturation by low-frequency currents in the guard circuit. The impedance looking into the amplifier guard terminal must be much lower than the impedance of the transformer windings; and at very high frequencies the capacitances of the transformer will allow signal leakage or may cause phase shift. These issues set incompatible constraints on the design of the transformer, if it must deal with a very wide range of common-mode frequencies.

In such a case double cancellation using two separate transformers as shown might be considered—the one nearer the amplifier having high inductance (and correspondingly high capacitance) and the other having good VHF efficiency.



Other approaches are also possible: the amplifier can be sited closer to the sensor and the long leads be replaced by leads (or optical fibre) carrying digital data, which is less vulnerable; more shielding is often (but not always) helpful; and sometimes (but rarely) it is possible to reduce the possibility of unexpected HF signals (even if you keep away the hams and police, there is always the possibility of the unexpected pizza delivery truck radioing to its base . . .)

The most important consideration, though, is awareness of the possibility of HF interference and readiness to tackle it. If designs are always made in the expectation of unwanted HF, chances are excellent that precautions will be adequate—it's when you don't expect it that the trouble starts.

Q. How did it work out with the French customer?

A. His problem was cured with two resistors, three capacitors and a piece of grounded copper foil. We went off to "La Cognette" to celebrate.

A Reader Notes

HIGH-FREQUENCY SIGNAL CONTAMINATION

by Leroy D. Cordill*

I found your article on high-frequency signal contamination ("Ask The Application Engineer—14," *Analog Dialogue* 27-2, 1993) interesting and would like to offer some additional comments.

EMI/EMC requirements are becoming more important to designers of industrial equipment as analog signal sensitivities are increased while more "RF generators" (higher-frequency digital circuits) are incorporated into the same equipment. Therefore, I would like to see a *good* application note relating to the issue of RF susceptibility produced by someone such as Analog Devices. By "good", I feel it should cover:

a. rules of thumb about the types of circuits where you will likely have trouble

b. some explanation of the phenomenon

c. general grounding/shielding approaches for equipment

d. "fix" type approaches to minimizing the effects when items from (c) can't be implemented

e. bench-level testing techniques.

(At least I'm not aware of any such application note in existence; maybe one exists and I haven't found it.) Based on my own experience, I offer the following comments on the above five areas:

Regarding (a), I generally see the problem with low-level input or preamp circuits involving a voltage gain of 50 V/V or more. In my case, the signals are usually from thermocouples, RTDs, pressure sensors, etc., and the required signal bandwidth is less than 100 Hz. And I'm trying to maintain signal integrity suitable for conversion by a 10-to-14-bit A/D converter.

For (b), my "model" of the effect is that the error gets created by rectification of the rf at the base-emitter junctions at the inputs of the op amp, and essentially becomes a large input offset voltage for the op amp. This introduces errors into dc-coupled circuits that cannot be corrected for by any usual low-pass filtering of the signal.

One observation I have made regarding this susceptibility problem is that it is primarily related to bipolar-type op amps (741, 5558, OP05, OP07, OP27, AD708, OP220, etc.) If I swap to a FET-input op amp (TL082, TL032, OP80, OP42, AD845) the error will largely disappear. (Due to other considerations, this is not usually a permanent solution, but helps to identify error sources during EMC testing.)

Also involved is the RF impedance at the two input nodes of the op amp. If (in a typical inverting configuration) the feedback path has a capacitor for low-pass filtering, it aggravates the problem as one input node of the op amp sees more of the RF than the other. If this is the situation, I'm not sure a wide-bandwidth op amp would help (regarding suggestions for using an AD830). Even without an intentional discrete capacitor in the feedback loop, PC-board layout makes it difficult to count on matched impedances at the two inputs.

*RR 3 Box 8910, Bartlesville OK 74003. Leroy Cordill, a design engineer with Applied Automation, Inc., has been involved in designing process gas chromatographs for about 20 years. His areas of design have included system architecture, analog, digital, and serial communication circuits, as well as GC detectors and valves.

Regarding (c), a good RF ground to the chassis is important for the signal common; but I find the shielding/grounding aspects of the equipment design relate more to the ESD requirements than RF (continuous-wave) susceptibility problems. I also try not to rely on these (shielding/grounding) to a great extent, since I find them very uncontrollable during the life of a piece of field-customizable equipment.

For (d), my best, most consistent prescription is placing a small capacitor directly across the input pins on bipolar op amps. I have used 100-1000 pF for this purpose in various circuits; it usually significantly reduces or eliminates the problem up to the level of interference that I plan for. I have found that with this in place on the critical parts of the circuit, the requirements for extreme care in grounding and shielding of cables are greatly reduced.

Regarding (e), I agree that a small walkie-talkie is useful, but primarily as a go/no-go test on the equipment when it is all assembled, in the enclosure, etc. However, for pc board or circuit-level work, I have two problems with the walkie-talkie technique: (1) you will get many unkind remarks from the guy on the next bench over if he's trying to breadboard a low-level circuit and is not ready for EMI testing yet; and (2) if you start attaching leads to various points in the circuit to determine where the problems are, and then apply RF in a radiated fashion, you have so many antennae, both to your circuit and to the various test gear, that you will have *no* idea what is happening.

I prefer to use an RF signal generator and apply the interference in a conducted fashion. This allows much better control of which items get RF applied to them. I don't use a lot of RF power, as I usually connect the output of the generator directly to some connector or cable supplying the low-level signal of interest, or in some cases the body of a sensor. A few hundred millivolts of RF signal is generally sufficient to identify problem circuits. I manually sweep from about 10 MHz to 100 MHz. While this is not a quantitative type of test, it is a very useful qualitative technique.

Some of the RF generators I have used for this are older model units—usually acquired at garage sales for $5 to $20 each:

RCA WV-50B
Advance Schools, Inc., Model IGB-102
Heathkit Model IG-102 (same as above)
Precise Model 630

I hope this may be useful, and, as I mentioned would like to see a good application note put together on this subject by someone who can add some additional information regarding performance implications of adding a capacitor on the op-amp inputs for various circuit configurations.

Thanks to Mr. Cordill for a useful contribution to the Dialogue, and for throwing down the gauntlet to our Application Engineers. They have accepted the challenge; so keep your eyes on the "Worth Reading" page in future issues. Having said that, we feel obliged to point out that the challenge is to get it together in one place; much of the material he suggests already exists in the Analog Devices literature (and elsewhere). For example, the System Applications Guide devotes pages 1-13 thru 1-55 to remote sensor application problems—including an exhaustive discussion of RFI rectification in high-accuracy circuits. Other good sources include the Applications Reference Manual, Chapter 3 and Bibliography of the Transducer Interfacing Handbook, and Part 5 and Bibliography of the Analog-Digital Conversion Handbook.

Ask The Applications Engineer—15

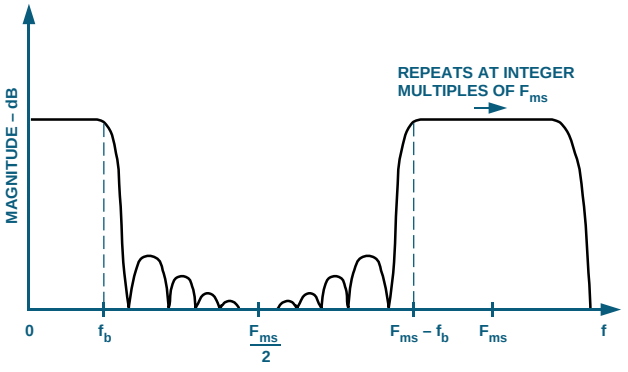
by Oli Josefsson

USING SIGMA-DELTA CONVERTERS—PART 1

Q: I'd like to use sigma-delta A/Ds but have some questions because they seem markedly different from what I've been using. To start with, what issues do I need to consider when designing my antialiasing filter?

A: A major benefit of oversampling converters is that the filtering required to prevent aliases can be quite simple. To understand why this is the case and what the filter constraints are, let's look at the basic digital signal processing that takes place in such a converter. For the purpose of anti-alias filter design we can think of a sigma-delta converter as a conventional high-resolution converter, sampling at a rate much faster than the Nyquist sampling rate, followed by a digital decimator/filter; the fact that the input into the digital decimator is 1-bit serial with a noise-shaping transfer function is irrelevant.

The input signal is sampled at F_{ms}, the modulator input sampling rate, which is much faster than twice the maximum input signal frequency (the Nyquist rate). The figure shows what the frequency response of a decimation filter may look like; frequency components between f_b and $F_{ms}–f_b$ are greatly attenuated. Thus, the digital filter can be used to filter out all energy from the converter within $[0, F_{ms}–f_b]$ that does not fall within the bandwidth of interest $[0,f_b]$. However, the converter can not distinguish between signals appearing at the input that are in the range $[0,\pm f_b]$ and those in the ranges, $[kF_{ms}\pm f_b]$, where k is an integer. Any signals (or noise) in those ranges get aliased down to the bandwidth of interest $[0,fb]$ via the sampling process; the decimation filter, which works only on the digitized samples, cannot be of any help attenuating these signals.



Thus it is the input noise energy in these bands $[kF_{ms}\pm f_b]$ that must be removed by the antialiasing filter before the input signal is sampled by the converter.

Q: So if I were to use the AD1877, which has a dynamic range of 90 dB, the antialiasing filter will need attenuation well above 90 dB at $F_{ms} – f_b$ (≈ 3 MHz)?

A: Not quite. You are assuming that the A/D has full-scale input at frequencies close to the modulator sampling rate; this is simply not the case in most systems. The only signal input of concern for aliasing is normally just noise from sensors and circuitry preceding the converter. The noise is usually low enough for a simple RC filter to suffice as an antialias filter.

Q: How do I make sure that a one-pole RC filter will suffice for my application—and establish the time constant of the filter?

A: Your application will typically specify a maximum allowable attenuation of an input signal that falls within the bandwidth of interest. This in turn puts a minimum on the 3-dB point of the RC filter. Let's take a look at an example using the AD1877 to illustrate this point further and to show how one might verify that a single-pole filter will provide enough filtering.

Let's assume that we have an application where the bandwidth of interest is 0 to 20 kHz, and signals in this range must not be attenuated more than 0.1 dB, or a ratio of 0.9886 [dB = 20 $\log_{10}$ (*ratio*) for voltage and 10 $\log_{10}$ (*ratio*) for power]. From the formula for attenuation of a single-pole filter,

$$ratio = \frac{1}{\sqrt{1+\left(2\pi\, fRC\right)^2}} > 0.99 \text{ at } f = 20 \text{ kHz}$$

$$RC \le \sqrt{\frac{1-\left(ratio\right)^2}{\left(2\pi\, f\right)^2\left(ratio\right)^2}} \approx 1.21\times10^{-6} \text{ s}$$

Choosing RC = 1.0 µs, to allow for component tolerances, the –3-dB frequency will be 159 kHz. We can now calculate the attenuation the filter will provide in the frequency bands, $kF_{ms}\pm f_b$, that alias down to the baseband. Assuming that the AD1877 has a modulator sampling rate of 3.072 MHz (and output sampling rate of 48 kHz), the first frequency band occurs at 3.052 MHz to 3.092 MHz. The attenuation of the RC filter at these frequencies is approximately 25.7 dB (about 0.052) over the whole band. Over the second band (6.124 MHz to 6.164 MHz), the attenuation is 31.8 dB (0.026). We know that the noise in these two bands (and all higher bands up the scale) that escapes through the filter to the A/D input will be aliased down to the baseband and get added as root sum-of-the-squares (rss) of their rms values, i.e., $\sqrt{n_1^2+n_2^2+\ldots+n_i^2}$. For values given in dB, the formulas shown the Appendix can provide results directly in dB, avoiding the intermediate step of computing the ratios.

For white noise, the noise spectral density is constant as a function of frequency, and each frequency range has the same bandwidth, so each band contributes an equal amount of noise to the input of the filter. We can therefore find the effective attenuation of the RC filter by adding the attenuation of the different frequency bands in rss fashion. The noise contribution from the first two bands, for example, is the same as the contribution from a single frequency band with attenuation of $\sqrt{0.052^2+0.026^2}$ = 0.058, or 24.7 dB, compared with 25.7 dB for the first band. How many bands do we need to consider when calculating the total aliased noise? For this case, the rss sums of the first 3, 4, 5, and 6 bands are, respectively, –24.2, –24.0, –23.9, –23.8 dB. The first band is therefore quite dominant; its attenuation is within 2 dB of the attenuation for all bands. It is usually sufficient to take only the first band into account unless the noise is exceptionally large or has a non-white spectrum; in addition, the A/D itself, though fast, has limited bandwidth; it tends to reject high-order bands.

Now that the attenuation is in hand, we can consider the noise magnitude itself: Let's be conservative (by about 50%) and take the effective filter attenuation to be 20 dB (i.e., 0.1 V/V). To be able to calculate the maximum allowed noise spectral density when using a single pole filter, an estimate should be made of the maximum performance degradation that aliased noise can contribute. From the dynamic specs of the AD1877 we find that the total noise power internal to the converter is 90 dB below (32 ppm of) full-scale input. If the whole system is to be within, say, 0.5 dB of this spec, the total aliased noise power can't exceed the rss difference between –90 dB and –89.5 dB or –99.1 dB (11.1×10^{-6}). Using this information, and the fact that the input scale of the AD1877 is 3 V p-p, we find that aliased noise must not exceed $3/(2\sqrt{2})$ V $\times 11.1 \times 10^{-6}$ = 11.8 μV rms. If all this noise were assumed lumped in a single aliased band, and noting that rms noise = noise spectral density (N.S.D.) $\times \sqrt{BW}$,

$$N.S.D. < \frac{11.8\,\mu V}{\sqrt{3.092\ MHz - 3.052\ MHz}} = 59\ nV/\sqrt{Hz}$$

This is the maximum post-filter spectral density allowed. To find the maximum prefilter spectral density (MPSD), with the effective filter attenuation of 20 dB (i.e., × 0.1) established previously, M.P.S.D. = 10 × 59 nV/$\sqrt{Hz}$ = 0.59 μV/$\sqrt{Hz}$.

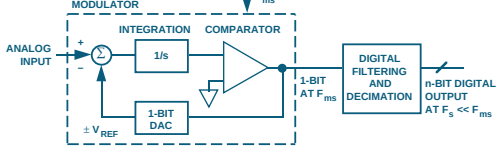
Clearly your system has to be pretty noisy in the 3-6-9-12-MHz regions in order for a simple RC filter not to suffice; however, as always, one must be careful of ambient rf pickup.

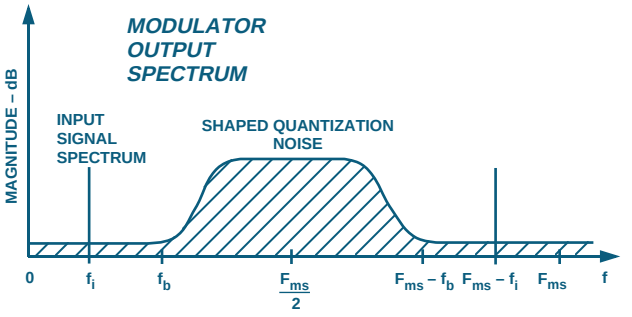
Q: As I understand it, the noise floor of sigma-delta converters may exhibit some irregularities. Any thoughts on that?

A: Most sigma-delta converters exhibit some spikes in the noise floor, called *idle tones*. In general, these spikes have low energy, not enough to substantially affect the S/N of the converter. Despite that, however, many applications cannot tolerate spikes in the frequency spectrum that extend much beyond the white noise floor. In audio applications, the human ear, for example, does an excellent job of detecting tones in the absence of large input signals even though the tones are well below the integrated (0-20-kHz) noise of the system.

There are two sources of idle tones. Their most common cause is voltage-reference modulation. To understand this mechanism a basic understanding of sigma-delta converters is needed. Here is a one minute crash course on sigma-delta converters (to probe further please consult[1]).

As the block diagram shows, a basic sigma-delta A/D converter consists of an oversampling modulator, followed by a digital filter and a decimator. The modulator output swings between two states (high and low, or 0 and 1, or +1 and –1), and the average output is proportional to the magnitude of the input signal. Since the modulator output always swings full-scale (1 bit), it will have large quantization errors. The modulator, however, is constructed so as to confine most of the quantization noise to the portion of the spectrum beyond f_b, the bandwidth of interest.



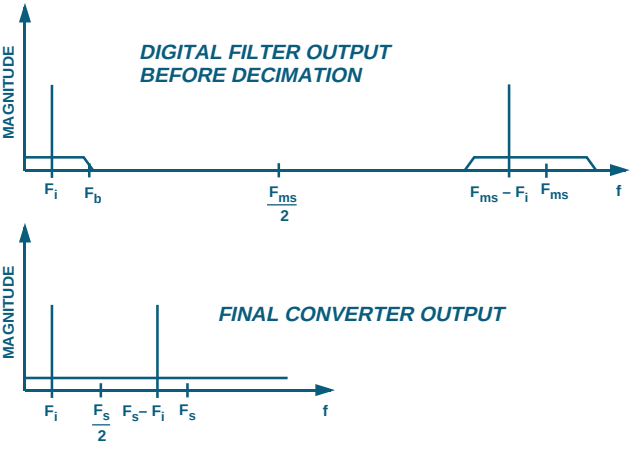


As shown, the spectral "sticks" (single frequencies) at f_i *and* $F_{ms} - f_i$ correspond to an input signal, while the shaded area shows how the quantization noise has been pushed (shaped) beyond the bandwidth of interest, f_b.

The digital filter, which is often an *n*-tap FIR filter, takes the high-speed low-resolution (1-bit) modulator output and performs a weighted average of *n* modulator outputs in a manner dictated by the desired filter characteristics. The output of the filter is a high-resolution word, which becomes the A/D output. The digital filter is designed to filter out "everything" between f_b and $F_{ms}–f_b$, *where* F_{ms} is the sampling rate of the modulator. Cleaning out all the noise in between f_b and $F_{ms}–f_b$ makes it possible to reduce the sampling rate to values between F_{ms} and $2f_b$ without causing any spectra to overlap (i.e., aliasing).

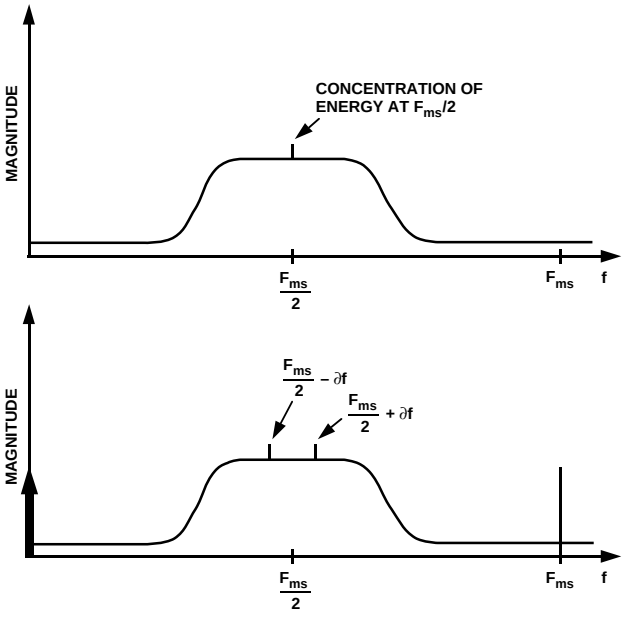
Conceptually, reducing the sample rate, i.e., *decimation*, can be thought of as only sending every *d*th digital filter output to the A/D output, where *d* is the decimation factor. This will bring the spectral images close together, as shown in the figure, which makes the output look like an output from a non oversampled converter. The upper figure shows the output of the modulator after digital filtering but prior to decimation. The lower figure shows the spectral output after decimation—the final A/D output.

In real converters, digital filtering and decimation are intimately combined for economy in design and manufacture. Thus, the terms "digital filter" and "decimator" are used interchangeably to describe the digital circuitry processing the modulator output to produce the output of the converter.



O.K., now back to "idle tones". Let's start by looking at the output of the modulator when a dc signal is applied to the input. For an exact mid-scale dc input level, the output of the modulator is equally likely to be high (1) or low (0), in other words,

the pulse density is 0.5, very likely to result in bitstream patterns like 010101. These regular patterns mean that the output spectrum will have a spike at $F_{ms}/2$ (upper figure). If the dc input now moves somewhat off midscale, the modulator output bit pattern will change accordingly. The spectrum of the modulator output will now show spikes at $F_{ms}/2 - \partial F$ and $F_{ms}/2 + \partial F$, with ∂F proportional to the dc change from midscale (lower figure).



With effective digital filtering, how can such tones possibly find their way down to baseband? The answer is via the *voltage reference*. The digital output is a measure of the ratio of the analog input to the voltage reference. An x% change in the magnitude of the voltage reference will result in a –x% change in the magnitude of the digital output word. Voltage-reference change will, in effect, amplitude modulate the A/D output. Now, we have clocks internal to the converter, and possibly also externally, running at $F_{ms}/2$. If small amounts of these clock pulses get coupled onto the voltage reference line, they will change it slightly and, in effect, modulate the tones at $F_{ms}/2 - \partial F$ and $F_{ms}/2 + \partial F$. One of the difference frequencies created by this modulation is at ∂F, and it is clearly in the bandwidth of interest. Nonlinearities may also create tones at multiples of ∂F.

Q: From your explanation it seems that if I apply an ac signal to the converter I do not have to worry about idle tones?

A: Well, any ac signal generally has a dc component associated with it, which will have to be represented by the modulator output, so the explanation above still applies. But if the total dc input offset (i.e., internal converter offset plus external offsets) in your system is exactly 0, the tones will be at dc (0 Hz).

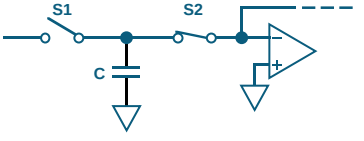
There is another source of idle tones in lower-order (<3rd-order) modulators. The order of the modulator (number of integrations) is a measure of how much quantization-noise shaping takes place. Second-order modulators can actually exhibit bit patterns that show up directly in the baseband, even if voltage-reference modulation is not occurring. This is one of the reasons why sigma-delta converters from Analog Devices that are designed for ac applications use higher-order (≥3) sigma-delta modulators.

Q: So what can I do to minimize the chances of idle tones interfering with my A/D conversion?

A: Follow the layout recommendations and bypassing schemes recommended by the manufacturer of the converter. This applies not only to the voltage reference, but to power supplies and grounding as well. It is the manufacturer's responsibility to minimize the voltage-reference corruption that takes place *inside* the converter, but it is up to the system designer to minimize the *external* coupling. By following those guidelines, the user should be able to reduce the coupling to a negligible level. If, despite the proper design precautions, idle tones are still an issue, there is yet another option that can be pursued. As I explained previously, frequency of the idle tones is a function of the dc input. This opens up the possibility of introducing enough dc offset on the A/D input to move the idle tones out of the bandwidth of interest to where they will be filtered out by the decimation filter. If the user does not want the dc offset to propagate through the system it can be subtracted out by the processor that handles the data from the A/D.

Q: What kind of a load does the input of sigma-delta converters present to my signal conditioning circuitry?

A: It depends on the converter. Some sigma-delta converters have a buffer at the input, in which case the input impedance is very high and loading is negligible. But in many cases the input is connected directly to the modulator of the converter. A switched-capacitor sigma-delta modulator will have a simplified equivalent circuit like that shown in the figure.



Switches S1 and S2 are controlled by the two phases of a clock to produce alternating closures. While S1 is closed, the input capacitor samples the input voltage. When S1 is opened, S2 is closed and the charge on C is dumped into the integrator, thus discharging the capacitor. The input impedance can be computed by calculating the average charge that gets drawn by C from the external circuitry. It can be shown that if C is allowed to fully charge up to the input voltage before S1 is opened then the average current into the input is the same as if there were a resistor of $1/(F_{sw}C)$ ohms connected between the input and ground, where F_{sw} is the rate at which the input capacitor is sampling the input voltage. F_{sw} is directly proportional to the frequency of the clock applied to the converter. This means that the input impedance is inversely proportional to the converter output sample rate.

Sometimes other factors, such as gain, can influence the input impedance. This is the case for the 16/24-bit AD771x family of signal conditioning A/Ds. The inputs of these converters can be programmed for gains of 1 to 128 V/V. The gain is adjusted using a patented technique that effectively increases F_{sw} (but keeps the converter output sample rate constant) and combines the charges from multiple samples. The input impedance of these converters is, for example, 2.3 MΩ when the device's external clock is 10 MHz and the input gain is 1. With input gain of 8, the input impedance is reduced to 288 kΩ.

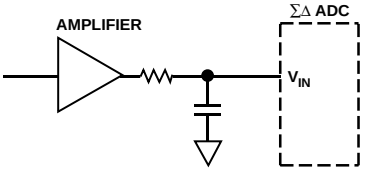
These impedances, as noted earlier, represent the average current flow into or out of the converters. However, they are not the impedances to consider when determining the maximum allowable output impedance of the A/D driver circuitry. Instead, one needs to consider the charging time of the capacitor, C, when S1 is closed. For dc applications the driver circuit impedance has only to be low enough so that the capacitor, C, will be charged to a value within the required accuracy before S1 is opened. The impedance will be a function of how long S1 is closed (proportional to the sampling rate), the capacitance, *C* and C_{EXT} in parallel with the input (unless C_{EXT} >> C). The table shows allowable values of external series resistance with f_{CLKIN} = 10 MHz which will avoid gain error of 1 LSB of 20 bits—for various values of gain and external capacitance on the AD7710.

Typical external series resistance which will not introduce 20-bit gain error

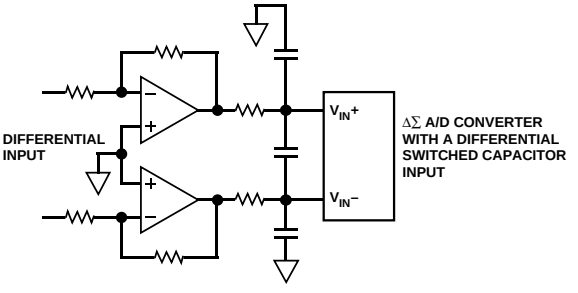
	External Capacitance (pF)					
Gain	**0**	**50**	**100**	**100**	**500**	**5000**
1	145 kΩ	34.5 kΩ	20.4 kΩ	5.2 kΩ	2.8 kΩ	700 Ω
2	70.5 kΩ	16.9 kΩ	10 kΩ	2.5 kΩ	1.4 kΩ	350 Ω
4	31.8 kΩ	8.0 kΩ	4.8 kΩ	1.2 kΩ	670 Ω	170 Ω
8-128	13.4 kΩ	3.6 kΩ	2.2 kΩ	550 Ω	300 Ω	80 Ω

For ac applications, such as audio, where the modulator sample rate is around 3 MHz for 64× oversampling, the input capacitor voltage may not have enough time to settle within the accuracy indicated by the resolution of the converter before the capacitor is switched to discharging. It actually turns out that as long as the input capacitor charging follows the exponential curve of RC circuits, only the gain accuracy suffers if the input capacitor is switched away too early.

The requirement of exponential charging means that an op amp can not drive the switched capacitor input directly. When a capacitive load is switched onto the output of an op amp, the amplitude will momentarily drop. The op amp will try to correct the situation and in the process hits its slew rate limit (non linear response), which can cause the output to ring excessively. To remedy the situation, an RC filter with a short time constant can be interposed between the amplifier and the A/D input as shown in the figure. The (low) resistance isolates the amplifier from the switched capacitor, and the capacitance between the input and ground supplies or sinks most of the charge needed to charge up the switched capacitor. This ensures that the op amp will never see the transient nature of the load. This additional filter can also provide antialiasing.



For converters that have a differential input, a differential version of this circuit may be used, as shown in the figure below. Since one input is positive with respect to ground while the other is negative, one input (the negative one) needs to be supplied negative charge while the other needs to get rid of negative charge when the input capacitors are switched on line. Connecting a capacitor between the two inputs enables most of the charge that is needed by one input to be effectively supplied by the other input. This minimizes undesirable charge transfers to and from the analog ground.



To be continued. Topics to be covered in the next installment include multiplexing, clock signals, noise, dither, averaging, spec clarifications

APPENDIX

RSS addition of logarithmic quantities: The root-sum square of two rms signals, S_1 and S_2, has an rms value of $\sqrt{S_1^2 + S_2^2}$. One often needs to calculate the rss sum of two numbers that are expressed in dB relative a given reference. To do this one has to take the antilogs, perform the rss addition, then convert the result back to dB. These three operations can be combined into one convenient formula: If D_1 and D_2 are ratios expressed in dB [negative or positive] their sum, expressed in dB, is

$$10 \log_{10}\left(10^{D_1/10} + 10^{D_2/10}\right)$$

Similarly, to find the difference between two rms quantities,

$$x = \sqrt{S_2^2 - S_1^2}$$

the result, *x*, expressed in dB, is

$$10 \log_{10}\left(10^{D_2/10} - 10^{D_1/10}\right)$$

References (not available from Analog Devices):

[1] *Oversampling Delta-Sigma Data Converters—Theory, Design, and Simulation,* edited by J.C. Candy and G.C. Temes, IEEE Press, Piscataway, NJ, 1991.

[2] J. Vanderkooy and S.P. Lipshitz, "Resolution Below the Least Significant Bit in Digital Systems with Dither," *J. Audio Eng. Soc.*, vol. 32, pp. 106-113 (1984 Mar.); correction ibid., p.889 (1984 Nov.).

[3] A.H. Bowker and G.J. Lieberman, *Engineering Statistics,* Prentice Hall, Englewood Cliffs, NJ, 1972.

Ask The Applications Engineer—16

by Oli Josefsson

USING SIGMA-DELTA CONVERTERS—PART 2

This is a continuation of a discussion of sigma-delta converters begun in the last issue. We covered antialiasing requirements, idle tones, and loading on the signal source.

Q: What happens if my input signal is beyond the input range of the sigma-delta converter? I remember hearing something about the converter becoming unstable?

A: The modulator can become temporarily unstable if it is driven with inputs outside the recommended range. However, this instability is invisible to the user, since decimators are generally designed to simply clip the digital output and show either negative or positive full scale, just as one would expect with a conventional converter.

Q: The specifications for sigma-delta converters assume a certain input clock rate and therefore a specific sampling rate. Can I safely use the converter with a higher or lower clock frequency?

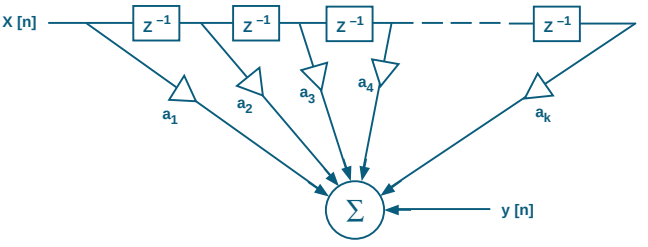
A: While the specs are measured at a particular sampling frequency, we often specify a range of input clock frequencies that the device can be operated with. This translates into a range of possible sampling rates. If you plan to go much beyond that range you can expect some performance degradation. If you sample at higher rates than specified, the internal switched-capacitor circuits may not be able to settle to the required accuracy before a new clock edge comes along. With too slow a sampling rate, capacitor leakage will degrade performance.

The digital filter characteristics of the converter (group delay, cutoff frequency, etc.) scale with sampling rate; so too do the input impedance (unless the input is buffered) and power consumption.

Q: I am planning to use a sigma-delta converter to digitize several signals by using a multiplexer at the input of the converter. Is that a problem?

A: While sigma-delta converters have a certain appeal due to their ease of antialiasing, they do not lend themselves well to applications for multiplexed ac signals. The reason for this is that the output of a sigma-delta converter is a function not only of the latest analog input but also of previous inputs. This is mostly due to the memory that the digital filter has of previous inputs, but the modulator has some memory as well. In a multiplexing application, after switching from one input to another, all information the filter has about the old input needs to be flushed out before the converter output word represents the new input.

Most decimation filters in sigma-delta converters intended for ac applications are FIR filters, principally because of their linear phase-response. For FIR filters, it is easy to calculate the time it takes to rid the filter of any information about the old input. The figure shows the structure of a FIR filter; the number of clock cycles required to clock all old data points out (i.e., the filter settling time) is equal to k, the number of taps in the filter. While data corresponding to a new input is propagating through the filter and replacing the earlier data, the output of the filter is calculated from a combination of the old data and the new data. The AD1879, for example, an 18-bit audio A/D converter, has a 4096-tap FIR filter which, when running at 3.072 MHz, has a 1.33-ms settling time.



$$y(n) = a_1\,x(n) + a_2\,x(n-1) + a_3\,x(n-2) + \ldots + a_k\,x(n-k+1)$$

The effective sampling rate for sigma-delta converters in multiplexed applications is quite low because of this need to wait for the old signal to be flushed out before capturing a valid data point for the new input. Traditional converters, which convert directly, or in a small number of stages, are therefore a much better choice in applications requiring the capture of multiple ac channels.

For a multichannel dc application where time is available to wait after switching between channels, or if the application does not require frequent changes between channels, the use of a sigma-delta converter can be very feasible. In fact, Analog Devices offers 16-24-bit converters with multiplexers on the input (AD771x family) specifically for such applications.

Q: Does this also explain why sigma-delta converters are not suitable for some control applications?

A: Yes. Since delays in control loops must be minimized for stability, sigma-delta converters are not suitable for control applications where they add a relatively long time delay. However, the actual delay is predictable; in applications that involve relatively slow signals, the converter phase delay, and therefore the effect on pole and zero locations of the control loop, may be negligible. However, even if this is the case, a traditional non-oversampling converter may still be a much better choice for the application, because a sigma-delta converter would need to run at a much faster sampling rate than a traditional converter in order to have the same phase delay. This will unnecessarily burden the circuitry that processes the A/D data.

Q: Are there any other issues I should be aware of when using sigma-delta converters?

A: In addition to the general guidelines on grounding, power supply bypassing, etc., that apply to all converters, there are a couple of points worth remembering when designing with sigma-delta converters. The first issue involves their input. As mentioned earlier, some sigma-delta converters (such as the AD1877) have buffers on the input; others (such as the AD1879), without a buffer, present a switched-capacitor load, which needs periodic current transients to charge the input capacitor. It is important that the circuitry driving the converter be as close to the converter as possible to minimize the inductance in the leads between the external circuitry and the switched-capacitor node. This reduces the settling time of the input and minimizes radiation from the input to other parts of the circuit board.

Another issue has to do with interference from clock signals affecting the A/D conversion. As I noted earlier, the digital decimation filter can't provide any filtering of signals whose frequencies are close to multiples of the modulator sampling rate. To be precise, the passbands are $[kF_{ms} \pm f_b]$s *where* k is an integer, F_{ms} is the modulator sampling rate, and f_b is the decimator cutoff frequency.

Besides the consequences for anti-aliasing discussed earlier, the decimator cutoff frequencies have a bearing on the selection of clock frequencies for devices that operate in the same system as the converter. These frequency bands (i.e., the passbands) embody the converter's greatest vulnerability to interference (inductive or capacitive coupling, power supply noise, etc.), because any signals in these frequency bands that manage to get into the modulator will not be subjected to attenuation in the filter. Therefore one is wise to avoid using clock frequencies that fall in these bands to minimize the possibility of interfering with the conversion—unless they are synchronous with the converter clock.

QUESTIONS ON NOISE IN CONVERTERS

Q: I recently evaluated a dual-supply A/D converter; one of the tests I did was to ground the input and look at the output codes on a LED register. To my big surprise I got a range of output codes instead of a single code output as I expected?

A: The cause is *circuit noise*. When the dc input is at the transition between two output codes, just a little circuit noise in even the finest dc converters will ensure that two codes will appear at the output. This is a fact of life in the converter world. In many instances, as in your case, the internal noise may be large enough to cause several output codes to appear. Consider, for example, a converter with peak-to-peak noise of just over 2 LSB. When the input of this converter is grounded, or a clean dc source is connected to the input, we will always see three—and sometimes even four—codes appear at the output. The circuit noise prevents the voltage being sampled from being confined to a voltage bin that corresponds to one digital code. Any external noise on the A/D input (including a noisy signal), on the power supplies, or on the control lines will add to the internal circuit noise—and possibly result in more bits toggling.

Q: Is there a way to determine how many codes I can expect to appear when I apply a dc signal to a converter?

A: It would not be hard in the ideal case where you knew the noise distribution, the exact size of the codes where the dc input is at and where within a code quantum the input lies (in the center, on the edge of two codes, etc.). But in reality you don't have this information. However, knowing some of the ac specifications (S/N, dynamic range, etc.) of the converter, you can make an *estimate*. From these specs you can find the magnitude of the rms converter noise relative to full scale. The noise will in all likelihood have a Gaussian amplitude distribution, so the standard deviation (sd) of the distribution equals the rms value. This also means that the codes that appear will not have equal probability of occurring. Using the fact that 99.7% of a Gaussian distribution occurs within ±3 standard deviations from the mean, we can estimate the peak-to-peak noise voltage at six times the standard deviation.

If N_{rms} is the rms value of the converter noise and V_{LSB} is the size of the LSB in volts (= $V_{span}/2^b$, where b is the number of bits in the output word) the peak to peak noise in terms of LSBs, N_B, is

$$N_B = \frac{6 \times N_{rms}}{V_{LSB}} = \frac{6 \times 2^b \times N_{rms}}{V_{\text{span}}}$$

If the signal-to-noise ratio of a converter expresses noise power relative to full scale, rms signal $\left(V_{span} / (2\sqrt{2})\right)$, we have

$$N_B = \frac{3}{\sqrt{2}} \times 2^b \times 10^{-\text{SNR}/20}$$

How many *codes* show up at the output depends where the mean of the input, i.e., the dc input value, is with respect to code transitions. If the mean is close to the boundary between two output codes, more codes are likely to appear than if the mean is half way between two output codes. It can easily be shown that N_C, the number of codes appearing for a particular value of N_B, is either INT(N_B)+1 or INT(N_B)+2, depending on the dc input value [INT(N_B) is the integer portion of N_B]. And don't be surprised to see even more codes from the less-probable noise amplitudes >±3 standard deviations.

How many *bits* will N_C cause to toggle on the output? The number of bits needed to represent N_C codes is

$$\text{INT}\left(\frac{\log N_C}{\log 2} + 0.5\right)$$

We can, however, see many more bits toggle, since the number of bits toggling is a function of the actual value of the converter's dc input. Consider, for example, that a one-code transition from an output word of –1 to 0 on a 2s-complement-coded converter involves inverting all the output bits.

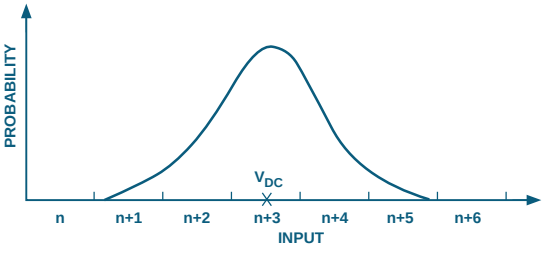
Lets look at an example using the AD1879, an 18-bit sigma-delta converter with dynamic range of 103 dB. From the definition of dynamic range we have

$$103 = 20 \log \frac{S}{N_{rms}}$$

From the AD1879 data sheet, we find that the rms value of a full-scale input signal, S, is $6/\sqrt{2}$ V rms. This allows us to solve for N_{rms} which turns out to be 30 µV. We next find the LSB size by dividing the full input range by the number of possible output codes:

$$V_{LSB} = \frac{12}{2^{18}} = 45.8\,\mu\text{V}$$

Thus N_B is 3.9. We can therefore expect either 4 or 5 different codes to appear at the AD1879 output when the input is grounded (ground corresponds to a midscale input for the AD1879).

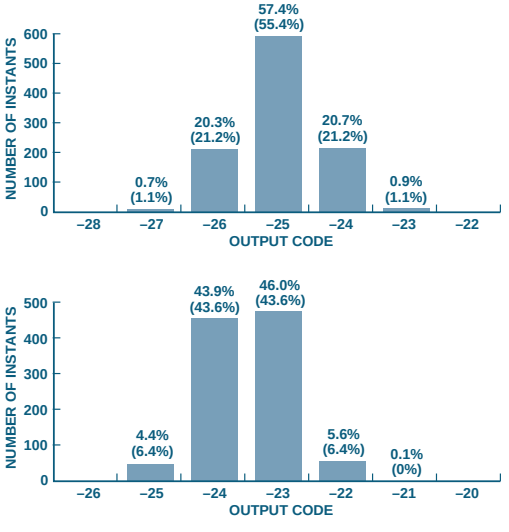


One can take this estimation one step further: If the standard deviation (the rms value) of a Gaussian distribution and the mean (the mean of the noise is 0 in this case) are known, one can use standard tables for the Gaussian distribution to calculate what percent of the time the noise will fall into a voltage interval corresponding to a specific output code. A histogram can be estimated, showing the distribution of codes at the output. Also the process can be reversed: a histogram showing the distribution of noise codes at a given value of dc output permits one to estimate the S/N ratio for a converter.

To make all this real, let's continue our example involving the AD1879. Consider two cases, one where the input lies midway between two output codes and one when the input is on the transition between two codes. From the calculations above, we found that the standard deviation (sd) of the noise (the rms value) was 30 μV. The size of one LSB in terms of sd is

$$\frac{45.78\,\mu V}{30.0\,\mu V} = 1.524$$

In the case where the dc input is midway between code transitions, as shown below, it is clear that any noise that falls within –0.5 LSBs to +0.5 LSBs from the input will result in the correct code at the A/D output. This corresponds to the noise being confined to a range of (–0.5 × 1.524) sd to (+0.5 × 1.524) sd from its mean (0). From standard tables one can find that the noise will fall in this range 55.4% of the time. If the noise falls within 0.5 LSBs to 1.5 LSBs, the output will be one code too high. Again from standard tables one can find that this will occur 21.2% of the time. Continuing in this manner one can calculate the whole histogram showing the distribution of output codes.



The upper figure shows an actual measurement where the dc input happened to be –25 LSBs. Five output codes, ranging from –27 to –23, appeared. 1024 measurements were taken and the percentage distribution of each code is shown on top of each column. The calculated distribution is listed in brackets on top of each column. As can be seen, the experimental results agree well with the calculated values. The lower figure shows a case where the dc input is close to the boundary between two codes. By following a similar procedure, one can calculate how the histogram should look. Again the experimental and calculated values are in excellent agreement. Note that the actual applied dc input is slightly above the border between the two codes, whereas the calculations assume it is exactly on the border.

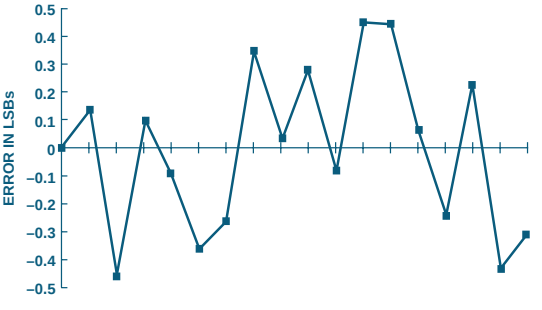
The biggest weakness of this estimating technique is the fact that in conventional converters the code width (the amount the dc input has to be increased to increase the digital output by one bit) varies from code to code. This means that if the dc input is in an area where codes are narrow, we can expect more bits to be toggling than in an area where the codes are wide. This method also assumes that the circuit noise within the converter stays constant, whether the applied signal is ac or dc. This is not exactly true in many cases.

The estimate will probably be more accurate when used with sigma-delta converters (except for "dead bands"), because neither of the two factors mentioned above is an issue in such converters.

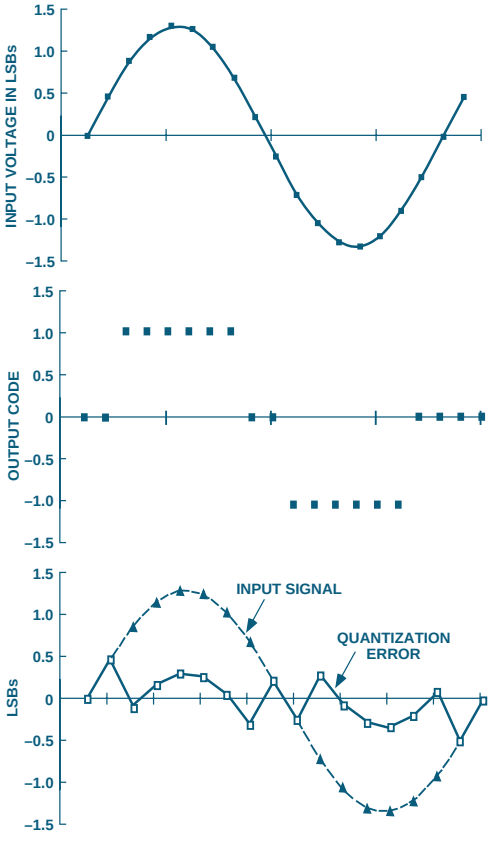
Q: *Ah, now I understand why there are multiple codes at the output. But why not discard the bits that toggle and only bring out the bits that stay steady, since the others are really indeterminate? Isn't that the real resolution of the converter?*

A: Many converters are designed for ac or dynamic applications where THD (total harmonic distortion) and THD+N (total harmonic distortion+noise) are the most important specs. The design therefore focuses on minimizing harmonic distortion for high- *and* low-level input signals, while keeping the noise to acceptable levels. As it turns out, these requirements somewhat contradict the requirements for a good dc converter, which is optimized for precision conversion of slow moving signals where harmonic distortion is not an issue. It is actually desirable to have some noise (called dither) superimposed on the input signal to minimize distortion at very low input signal levels; dither can also be used to improve dc accuracy where repeated measurements can be made.

To understand how this may be, let's start by looking at quantization noise. The output of an ideal A/D converter has finite accuracy because of the finite number of bits available to represent the input voltage. Each one of the 2^b quanta represents with one single value all values in the analog range from –0.5 LSB to +0.5 LSB of its nominal input value. The A/D output can therefore be thought of as a discrete version of the analog input plus an error signal (quantization noise). When a large and varying input signal (dozens, hundreds, or thousands of LSBs in amplitude) is applied to a converter, the quantization noise has very little correlation with the input signal. It is, in other words, approximately white noise. The figure shows the quantization noise of a perfect A/D converter at various instants of time when the input signal is a sinusoid of about 100 LSBs in amplitude.



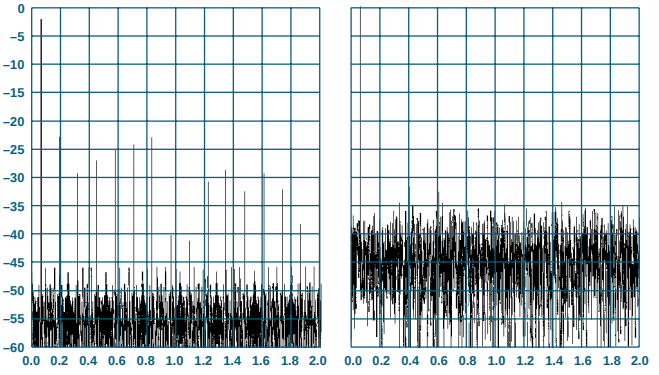
When the A/D input is very low in amplitude, so that the amplitude does not change more than a fraction of a LSB between samples, the samples stay in the same quantum, and are therefore constant for a few sample periods. This is depicted in the figure below, which shows a sinusoidal input signal that has an amplitude of only 1.5 LSBs, the A/D output and the quantization noise. Note that the quantization error follows the input waveform exactly while the samples are staying constant. The longer the samples stay constant, the more the quantization noise looks like the input waveform, i.e., the correlation between the input signal and the quantization noise increases. While the rms of the quantization error may not have changed, the quantization error will take on a non-uniform spectral shape. In fact, the correlated quantization noise shows up as harmonics in the A/D spectrum.



Another way to look at this phenomenon is to consider the case when the (sinusoidal) input signal is only around 1 LSB in size and the digital output resembles a square wave. Square waves are rich in harmonics! The harmonics, or noise modulation products, are very objectionable in many converter applications, especially audio.

To get around this problem, a technique called dithering is used to trade correlated quantization noise for white noise, which is less offensive to the human ear than correlated noise. Dithering is done by using circuit elements to *add* random noise to the input signal. While this will result in an increase of the total converter noise, the added noise breaks up the simple square wave patterns in the output code. The quantization error will not be a function of the input signal but of the instantaneous value of the dither noise. Thus the dither decorrelates the quantization noise and the input signal. The size of the dither signal is often about 1/3 LSB rms (2 LSBs peak-to-peak if the noise is Gaussian). Clearly, this will result in a converter that will have more than two codes at the output when the input is grounded. We saw an example earlier involving the AD1879 which had either four or five codes appear on the output depending on the dc input level.

The figure below shows the simulated output of an A/D converter with an undithered low level input signal. The quantization noise is a function of the input signal magnitude at the sample instant. This correlation between the quantization noise and the input signal shows up as a cluster of harmonically related sticks in the A/D output spectrum. Note that the magnitude scale in the figure is referenced to the input signal (not full scale input).



The right-hand figure shows the A/D output after a dither signal that is 4 dB above the quantization noise floor is added to the input. In this case the quantization noise depends on the magnitude of the dither signal at the instant when a sample is taken. Since the value of the dither doesn't depend on the input signal, the quantization noise becomes uncorrelated to the input and the harmonics in the A/D spectrum are eliminated, but at the cost of an overall increase in the noise floor.

Instead of actually adding noise to the A/D input, dithering can be accomplished by using the thermal noise of the converter as the dither signal and calculating enough output bits to ensure a decorrelated quantization noise.

Though I have used A/D converters in my examples, the idea of using dither also applies to D/A converters as well. Dither is applied to D/A converters by adding the output of a digital noise generator to the digital word sent to the D/A.

Q: *But in dc applications, I want to make an accurate measurement each time and may not be able to tolerate the uncertainty of having a few LSBs of error in a particular measurement.*

A: If you need *n*-bit dc accuracy in each conversion and you have problems finding a suitable *n*-bit converter, you have two options. One is to use an (*n*+2)-bit converter and simply ignore the two LSBs. However, if your hardware has the capability (and time) to do some signal processing, you can enhance the resolution of a noisy (dithered) dc converter and, in fact, get more than *n*-bit accuracy out of an *n*-bit converter if the accuracy is limited by noise.

To understand why this may be so, think of an ideal *n*-bit converter. For a particular value of dc input, you will get one digital code at the output. However, you do not know where the input lies within the code quantum (i.e., in the middle,

close to the upper transition, etc.). That may be sufficiently accurate for your application, but if you add noise to the input of the converter—so that several codes can appear at the output—you will find that the code distribution contains information to place the dc value of the input more exactly.

In the earlier examples involving the AD1879, we saw how the code distribution looks when the input is in the vicinity of a code transition; the two most-frequent output codes are the ones on either side of the transition. Their average is therefore a good estimate of where the input lies. In fact, taking the average of a lot of conversions, while the input stays put, is an excellent way of enhancing the resolution of the converter. One has to be careful, when processing the converter output, to allow the output word length to grow without introducing roundoff errors. Otherwise one actually injects unwanted noise—called *requantization noise*—into the final output. Note that filtering out the noise is only just that; it will have no effect on other error sources of the converter, such as integral and differential nonlinearity.

This concept of resolution enhancement is an interesting one and is not restricted to the dc domain. One can actually trade resolution for bandwidth in the ac domain and combine the outputs of several converters or to construct a more-accurate output. The basic principle is that signal repetitions (which are self-correlated) add linearly, while repetitions of random noise produce root-square increases. Thus, a fourfold increase in number of samples increases S/N by 6 dB.

Q: You mentioned a couple of converter ac specifications above. I am somewhat confused about how S/N, THD+N, THD, S/THD, S/THD+N, and dynamic range are measured on A/D and D/A converters and how they relate to each other. Can you shed any light on this?

A: Your confusion is quite understandable. There is unfortunately no industry standard on exactly how these quantities are measured and therefore, what exactly they mean. Sometimes manufacturers are guilty of choosing the definition that portrays their part favorably.

Most often data sheets include a note on the testing conditions and how the different specs were calculated. The best advice I can give is to read these very carefully. By simple calculations you can often convert a specification for one part to a number that allows a fair comparison to a specification for another part.

Most specifications are not expressed in absolute units, but as relative measurements or ratios. Noise, for example, is not specified in rms volts, but as *SNR*, or the ratio between signal power and noise power under particular test conditions. These ratios are usually expressed in decibels, dB, and occasionally as percentages (%). A power ratio, x, expressed in bels, is defined as $\log_{10}x$; multiply by 10 if expressed in decibels (one tenth of a bel): $10\log_{10}x$. *SNR* is therefore equal to $10\log_{10}$ (signal power/noise power) dB. Evaluated in terms of rms voltage quantities, $SNR = 20\log_{10}(V_{signal}/V_{noise})$.

Armed with this knowledge, let's see whether we can make sense out of the multiple specifications you mentioned above (many of which are redundant). Those specifications seek to describe how the imperfections of the converter affect the characteristics of an ac signal that gets processed by the converter. For dc applications, a listing of the magnitude of the actual imperfections suffices, but these can only suggest ac performance. For example, integral nonlinearity is a major factor in determining large-signal distortion (along with glitch energy for D/As) while differential nonlinearity governs small-signal distortion. To accurately determine the ac performance, at least two types of tests are performed in the case of A/Ds. The tests are as follows:

a) Full-scale sine

A sinusoidal signal approaching full-scale is applied to the converter. The signal is large enough so that converter's imperfections cause significant harmonic components to occur at multiples of the input signal frequency. The harmonics will show up in the output spectrum, along with noise. A common performance measure is the relative magnitude of the harmonic components, usually expressed in dB. Relative to what? Two possibilities are the applied input signal and the full scale of the converter (which in most cases is different from the applied input signal). Referring the harmonics to full scale will clearly yield a lower (more attractive) number than referring them to the rms value of the actual input signal. This reference issue causes a lot of confusion when dynamic specifications are evaluated, because there is no universally accepted standard for what each performance measure should be referred to. The best advice I can give you is: never assume anything; read manufacturers' data sheets very carefully.

Sometimes the magnitudes of the individual harmonics are specified, but most often only the total harmonic distortion (THD) is specified. The THD measures the total power of the harmonics and is found by adding the individual harmonics in rss fashion. The formula then for *THD* when referred to the input signal is

$$20\log_{10}\left[\frac{\sqrt{\sum_{i=2}^{m}H^2(i)_{rms}}}{S}\right] \text{ or } 10\log_{10}\left[\frac{\sum_{i=2}^{m}H^2(i)_{rms}}{S^2}\right]$$

where $H(i)_{rms}$ refers to the rms value of ith harmonic component and S to the rms value of the input signal. Usually, harmonics 2 through 5 are sufficient. Note that the input-frequency, or *fundamental*, component is the first harmonic. To refer any harmonic to full scale, add x dB to the formula above, where x is the magnitude of the input signal relative to full scale. This simple conversion formula can be applied to other specifications, but take care to observe proper polarity of the log quantities.

Nowadays, clear distinction is usually made between total harmonic distortion plus noise (*THD+N*) and *THD*. This has not always been the case. *THD+N* includes not only the harmonics that are generated in the conversion, but also the noise. The formula for *THD+N* when referred to the input signal is:

$$20\log_{10}\left[\frac{\sqrt{N^2_{rms}+\sum_{i=2}^{m}H^2(i)_{rms}}}{S}\right]$$

or

$$10\log_{10}\left[\frac{N^2_{rms}+\sum_{i=2}^{m}H^2(i)_{rms}}{S^2}\right]$$

where N_{rms} is the rms value of the integrated noise in the bandwidth specified for the measurement.

Another commonly used specification is signal to noise-plus-distortion ($S/[N+D]$, or $S/[THD+N]$), also called *sinad*. This is essentially the inverse of *THD+N*, when referred to the signal; its dB number is the same, but with opposite polarity.

Another performance measure describing the test results is the signal to noise ratio, *S/N* or *SNR*, which is a measure of the relative noise power, most useful for estimating response to small signals in the absence of harmonics. If *S/N* is not specified, but *THD* and *THD+N* are provided, relative to the input signal, *THD* can be rss-subtracted from *THD+N* to obtain the noise to signal ratio [= $1/(S/N]$. If the numbers are given in dB, the rss subtraction formula for logarithmic quantities in the Appendix can be used as follows

$$SNR = -10\log_{10}\left(10^{(THD+N)/10} - 10^{THD/10}\right)$$

to yield the input signal power relative to noise power expressed in dB.

b) Low-level sine

The second test usually performed is to apply a sinusoidal signal well below full scale to the converter (usually –60 dB). At this input level, sigma-delta converters usually exhibit negligible nonlinearities, so only noise (no harmonic components) appears in the spectrum. At this level, $S/N = S/N+D = -THD+N = -THD$, when all are referred to the same level. As a result, one specification indicating the noise level suffices to describe the result of this test. This specification called *dynamic range* (inversely, dynamic-range *distortion*), specifies the magnitude of the integrated noise (and harmonics if they exist) over a specific bandwidth relative to full scale, when a –60-dB input signal is applied to the converter.

Conventional (i.e. not sigma-delta) converters can exhibit harmonics in their output spectrum even with low-level input signals because all the codes may not have equal width (differential nonlinearity). In some such instances, the *S/N*, which ignores harmonics, measured with a –60-dB input signal, is different from dynamic range.

Frequently one sees *THD+N* at –60-dB and dynamic range specified for the same converter. These really are, as explained above, redundant since they only differ in the reference used. The only twist on dynamic range is that sometimes, when audio converters are specified, a filter that mimics the frequency response of the human ear is applied to the converter output. This processing of the converter output is called A-weighting (because an A-weighting filter is used); it will effectively decrease the noise floor, and therefore increase the signal-to-noise ratio, if the noise is white.

Everything discussed above applies to both A/D and D/A converters, with the possible exception of signal to noise ratio. Sometimes (particularly for audio D/A converters) *S/N* is a measure of how "quiet" the D/A output is when zero (midscale) code is sent to the converter. Under these conditions, the S/N expresses the analog noise power at the D/A output relative to full scale output.

It's important to note that the performance measures above are affected by: *bandwidth* of the measurement, the *sampling frequency*, and the input *signal frequency*. For a fair comparison of two converters, one has to make sure that these test conditions are similar for both.

Image Filtering Question

Q: I intend to use Analog's AD1800 family of audio D/A converters for a digital audio playback application. I understand that using an interpolator ahead of the D/A will make it easier to filter the D/A output, assuming I want to get rid of all the images at the D/A output. But is it really necessary to filter the output, since all the images will be above the audible range as long as sampling is at >40 kHz?

A: Good question. The audio equipment (audio amplifiers, equalizers, power amplifiers, etc.) that may eventually receive the output of your D/As are typically built to handle 20-Hz to 20-kHz signals. Since they are not intended to respond at frequencies much beyond 20 kHz—and in effect themselves function as filters—they may not have the necessary slew rate and gain to handle incoming signals from an unfiltered D/A output having significant energy well above 20 kHz. With their slew-rate and gain limitations, the amplifiers are driven into nonlinear regions, generating distortion. These distortion products are not limited to high frequencies but can affect the 20-Hz to 20-kHz range as well. Attenuating the high frequency signals at the DAC will therefore reduce the possibility of distortion. CD players often include filters steep enough to reduce the total out-of-band energy to >80 dB below full scale.

APPENDIX

RSS addition of logarithmic quantities: The root-square sum of two rms signals, S_1 and S_2, has an rms value of $\sqrt{S_1^2 + S_2^2}$. One often needs to calculate the rss sum of two numbers that are expressed in dB relative a given reference. To do this one has to take the antilogs, perform the rss addition, then convert the result back to dB. These three operations can be combined into one convenient formula: If D_1 and D_2 are ratios expressed in dB, their sum, expressed in dB, is

$$10\log_{10}\left(10^{D_1/10} + 10^{D_2/10}\right)$$

Similarly, to find the difference between two rms quantities,

$$x = \sqrt{S_2^2 - S_1^2}$$

the result, x, expressed in dB, is

$$10\log_{10}\left(10^{D_2/10} - 10^{D_1/10}\right)$$

References:

[1] *Oversampling Delta-Sigma Data Converters—Theory, Design, and Simulation*, edited by J.C. Candy and G.C. Temes, IEEE Press, Piscataway, NJ, 1991.

[2] J. Vanderkooy and S.P. Lipshitz, "Resolution Below the Least Significant Bit in Digital Systems with Dither," *J. Audio Eng. Soc.*, vol. 32, pp. 106-113 (1984 Mar.); correction ibid., p.889 (1984 Nov.).

[3] A.H. Bowker and G.J. Lieberman, *Engineering Statistics*, Prentice Hall, Englewood Cliffs, NJ, 1972.

Ask The Applications Engineer—17

MUST A "16-BIT" CONVERTER BE 16-BIT MONOTONIC AND SETTLE TO 16 PPM?

by Dave Robertson and Steve Ruscak

Q. *I recently saw a data sheet for a low-cost 16-bit, 30 MSPS D/A converter. On examination, its differential nonlinearity (DNL) was only at the 14 bit level, and it took 35 ns (1/28.6 MHz) to settle to 0.025% (12 bits) of a full scale step. Isn't this at best a 14 bit, 28 MHz converter? And if the converter is only 14-bit monotonic, the last two bits don't seem very effective; why bother to keep them? Can I be sure they're even connected?*

A. That's a lot of questions. Let's take them one at a time, starting with the last one. You can verify that the 15th and 16th bits are connected by exercising them and observing that 0..00, 0..01, 0..10, and 0..11 give a very nice 4-level output staircase, with each step of the order of 1/65,536 of full scale. You can see that they would be especially useful in following a waveform that spent some of its time swinging between 0..00 and 0..11, or providing important detail to one swinging through a somewhat wider range. This is the crux of the *resolution* spec, the ability of the DAC to output 2^{16} individual voltage levels in response to the 65,536 codes possible with a 16-bit digital word.

Systems that must handle both strong and weak signals require large dynamic range. A notable example of this is the DACs used in early CD player designs. These converters offered 16-20 bits of dynamic range but only about 14 bits of differential linearity. The somewhat inaccurate representation of the digital input was far less important than the fact that the dynamic range was much wider than that of LP records and allowed both loud and soft sounds to be reproduced with barely audible noise—and that the converters' low cost made CD players affordable.

The resolution is what makes a 16-bit DAC a "16-bit DAC". Resolution is closely associated with *dynamic range*, the ratio of the largest signal to the smallest *that can be resolved*. So dynamic range also depends on the noise level; the irreducible "noise" level in ideal ADCs or DACs is *quantization noise*.

Q. *What is quantization noise?*

A. The sawtooth-wave-shaped quantization noise of an ideal n-bit converter is the difference between a linearly increasing analog value and the stepwise-increasing digital value. It has an rms value of $1/(2^{n+1}\sqrt{3})$ of span, or $-(6.02\,n + 10.79)$ dB (below p-p full scale). For a sine wave, with peak-to-peak amplitude equal to the converter's span, rms is $\sqrt{2}/4$, or –9.03 dB, of span, so the full-scale signal-to-noise ratio of an ideal n-bit converter, expressed in dB, becomes the classical

$$6.02\,n + 1.76 \text{ dB}. \qquad (1)$$

As the analog signal varies through a number of quantization levels, the associated quantization noise resembles superimposed "white" noise. In a real converter, the circuit noise produced by the devices that constitute it adds to quantization noise in root-sum-of-squares fashion, to set a limit on the amplitude of the minimum detectable signal.

Q. *But I still worry about that differential nonlinearity spec. Doesn't 14-bit differential nonlinearity mean that the converter may be non-monotonic at the 16-bit level, i.e., that those last two bits have little influence on overall accuracy?*

A. That's true, but whether to worry about it depends on the application. If you have an instrumentation application that really requires 16-bit resolution, 1/2-LSB accuracy for all codes, and 1-LSB full-scale settling in 31.25 ns (we'll get to that discussion shortly), this isn't the right converter. But perhaps you really need 16-bit dynamic range to handle fine structure over small ranges, as in the above example, while high overall accuracy is not needed—and is actually a burden if cost is critical.

What you need to consider in regard to DNL in signal-processing applications is 1) the noise power generated by the DNL errors and 2) the types of signals that the D/A will be generating. Let's consider how these might affect performance.

In many cases, DNL errors occur only at specific places along the converter's transfer function. These errors appear as spurious components in the converter's output spectrum and degrade the signal-to-noise ratio. If the power in these spurs makes it impossible to distinguish the desired signal, the DNL errors are too large. Another way to think about it is as a ratio of the quantity of good codes to bad codes (those having large DNL errors). This is where the type of signal is important.

The various applications may concentrate in differing portions of the converter's transfer function. For example, assume that the D/A converter must be able to produce very large signals and very small signals. When the signals are large, there is a high proportion of DNL errors. But, in many applications, the signal-to-noise ratio will be acceptable because the signal is large.

Now consider the case where the signal is very small. The proportion of DNL errors that occur in the region of the transfer function exercised by the signal may be quite small. In fact, in this particular region, the spurs produced by the DNL errors could be at a level comparable to the converter's quantization noise. When the quantization noise becomes the limiting factor in determining signal-to-noise ratio, 16 bits of resolution will really make a difference (12 dB!) when compared to 14 bits.

Q. *OK, I understand. That's why there's such a variety of converters out there, and why I have to be careful to interpret the specs in terms of my application. In fact, maybe data sheets that have a great number of "typical" plots of parameters that are hard to spec are providing really useful information. Now, how about the settling-time question?*

A. *Update rate* for a D/A converter refers to the rate at which the digital input circuitry can accept new inputs, while *settling time* is the time the analog output requires to achieve a specified level of accuracy, usually with full-scale steps.

As with accuracy, time-domain performance requirements differ widely between applications. If full accuracy and full-scale steps are required between conversions, the settling requirements will be quite demanding (as in the case of offset correction with CCD image digitizers). On the other hand, waveform synthesis typically requires relatively small steps from sample to sample. The solid practical ground is that full-scale steps in consecutive samples mean operation at the Nyquist rate (half the sampling frequency), which makes it extremely difficult (how about "impossible"?) to design an effective anti-imaging filter.

Thus, DACs used for waveform reconstruction and many other applications* inevitably oversample. For such operation, full-scale settling is not required; and in general, smaller transitions require less time to settle to a given accuracy. Oversampled waveforms, taking advantage of this fact, achieve accuracy and speed greater than are implied by the full-scale specification.

*The AD768 is an example of such a DAC.

Ask The Applications Engineer—18

SETTLING TIME

by Peter Checkovich

Q. *Why is settling time important?*

A. Op amp settling time is a key parameter for guaranteeing the performance of data acquisition systems. For accurate data acquisition, the op amp output must settle before the A/D converter can accurately digitize the data. However, settling time is generally not an easy parameter to measure.

Over the years, the techniques and equipment used to measure the settling time of op amps have been barely able to keep up with the performance of the devices themselves. As each new generation of op amps settles to better accuracy in shorter time, greater demands have been placed on test equipment, its designers, and its users. A major dilemma, often causing disagreement among engineers, is whether some combination of techniques and equipment actually measures the device under test (DUT) or just some limiting property of the test setup. So there is continual development of new test equipment and techniques in an effort to specify this ever-demanding parameter.

In a data-acquisition system, the output of an op amp should settle to within 1 LSB [i.e., $2^{-n}FS$] of final value of the A/D that it drives within a time period dictated by the sampling rate of the system. To settle within 1 LSB of full scale implies the settling accuracy of the A/D is ±1/2 LSB. Thus, a 10-bit system will require the op amp to settle to half of one part in 1024, or approximately 0.05%. A 12-bit system will require settling to half of one part in 4096 (0.01%). The requirements for 14-bits and greater are yet more demanding. Settling-time values such as 0.1% and 0.01% are the most widely specified.

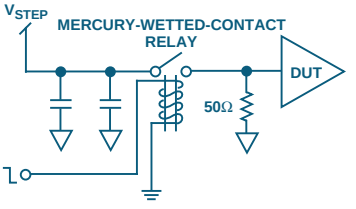
Although a larger full-scale signal range will increase the size of the LSB, easing the problem somewhat, it is not a feasible approach for high-frequency systems. Most high frequency A/Ds have a full-scale span of 1 V or, at most, 2 V. For a 10-bit system with a 1-V full scale signal, an LSB is about 1 mV. For a 12-bit system, an LSB is approximately 250 µV. To resolve the settling characteristics for a full-scale transition, dynamic ranges approaching four orders of magnitude must be handled. With settling times of new op amps [e.g., the AD9631 and AD9632] dropping to the 20 ns to 10 ns range, the measurement of settling time presents quite a challenge.

Q. *How is settling time measured?*

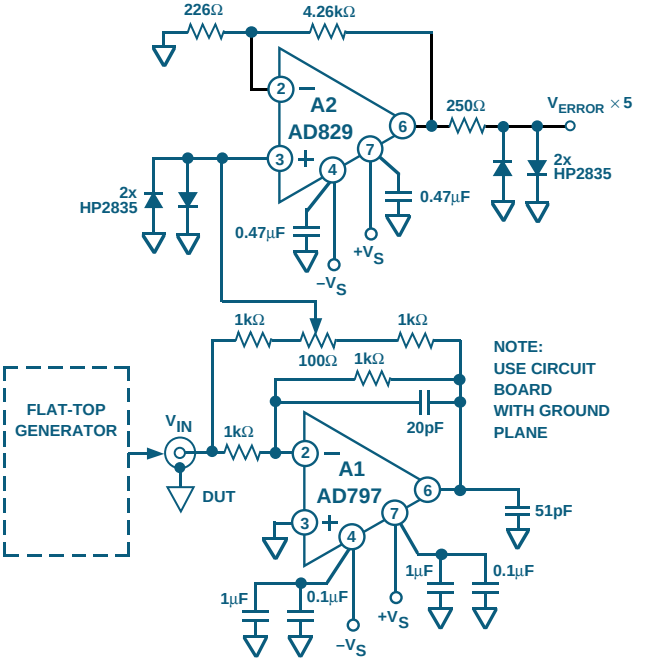
A. A key requirement over the years has been the need to drive the input of the op amp with a fast, precise signal source, often referred to as a flat-top generator. As the name implies, such a generator would have a sharp transition between two levels of known amplitude at time, t_0, should have minimal overshoot (or undershoot) and then remain flat for the remainder of the measurement time. In this case "flat" means significantly flatter than the error to be measured in the amplifier.

The great accuracy is required to be certain that any output signal from the op amp is entirely due to its settling response and not its response to a signal that is present at the input after the step transition. Any active device in the path of this signal would require better settling characteristics than the DUT.

Such generators are in practice very difficult to develop. A rather "low-tech" device has served for quite some time as a means for generating a flat-top transition; the contact opening of a mercury-wetted-contact relay connected to a stable low-impedance voltage source can be used to produce a rather clean (and surprisingly fast) flat-topped pulse. The figure shows a simple circuit that performs this function. For a negative-going transition, with the relay closed, a dc voltage, VSTEP, is applied to the input of the DUT and a 50-Ω resistor to ground. When the relay opens, the input node rapidly discharges to ground, creating the input transition. The open relay contact ensures that all other elements are totally isolated from the amplifier input; the input level is held constant (grounded through 50 Ω) for as long as the relay remains open.



Next problem: directly measuring the output requires handling a large dynamic range. If the DUT is configured as an inverter, a subtractor circuit can be created that only looks at the error signal and does not have to handle the entire dynamic range of the output. This figure shows a circuit used for measuring the 16-bit settling time of the AD797—800 ns typical to 0.0015%.



A1, the DUT in this circuit, is configured for a gain of –1. The voltage divider from input to output forms a second "false" summing node that will replicate the signal at the amplifier's summing node. The 100-Ω potentiometer is used to null the dc voltage. The wiper of the potentiometer is clamped by the diodes at the input of A2 to limit saturation effects in this amplifier. The output is also similarly clamped.

Since the pre- and post-transition voltages at the output of A2 will be the same (i.e., the difference will be zero), the settling

characteristics of this amplifier due to a step change are not important for measuring A1. Thus, the output of A2 can be measured to find the settling time of A1.

This technique requires that the DUT be configured as an inverting amplifier. The circuit can be made to work at other gains, but the resistor values and setting of the dc balance potentiometer will have more influence on the measurement.

Q. *Any other techniques?*

A. Another technique for measuring settling time uses the computing power of a digital oscilloscope. It calculates a waveform that represents the settling error as the instantaneous difference between the acquired input and output signals of the DUT and compares them with the values for an ideally settling device. The resulting waveform is the error of the DUT.

If there is a gain error in this system, it will show up as a dc offset in the error waveform. The calculation can be adapted for a DUT with any gain, either inverting or non inverting. It also can compensate for a signal generator that itself has a low frequency settling tail. The DUT response to a low frequency input will not be influenced by that settling time.

Because such oscilloscopes are designed primarily for speed, in order to determine errors at higher resolutions, averaging must be used. For example, if the A/D used in the oscilloscope has only 8 bit resolution, but accuracy better than 8 bits, a number of cycles can be averaged to increase the effective resolution of the measurement.

Q. *Any more?*

A. Yet a third way to measure settling time is to look at the output directly. A Data Precision Data 6000 can directly digitize signals of up to 5 V with 16-bit accuracy and 10-ps resolution. The only fly in the ointment is that the instrument relies on repetitive sampling with a comparator probe. The waveform is built up one bit at a time for each of the sample points. As a result, obtaining a settling characteristic can be very time consuming. This is especially so when using a relay-type flat top generator with a 1-kHz upper frequency.

Q. *Why do data sheets sometimes define short term and long term settling characteristics?*

A. The traditional definition of settling time is the time from the input transition to the time when the amplifier output enters the specified error zone and does not leave again. This concept is relatively uncomplicated and straightforward. However, there are some cases where the initial settling is fast, followed by an extended period of settling to the final value. Single-supply amplifiers may exhibit this characteristic in the vicinity of the lower rail. Of greater prevalence for large transients, a "thermal tail" is a slow drift that continues for a relatively long time after rapid settling to apparently excellent initial accuracy.

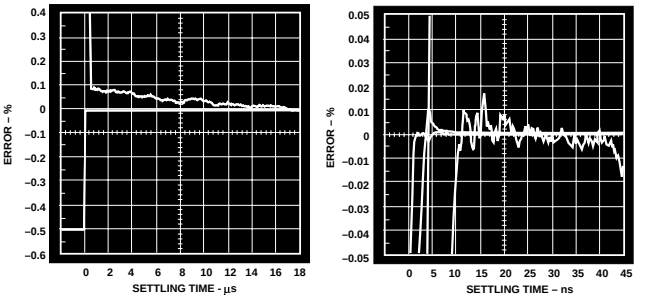
Thermal tails are produced when voltage level changes within the op amp caused by a step transition create temperature gradients among the transistors. Matched transistors will not track well while they are at temporarily different temperatures. The thermal time constant of the chip determines how long it takes for equilibrium to return. Op amps are designed to prevent or reduce these effects by careful placement of devices and strategies to produce thermal symmetry, but this is easier for low-level high-precision devices than those designed for high-speed, because of the large, rapid swings of power that occur.

In particular, the new dielectrically isolated processes (like XFCB) that have worked wonders for improving the raw speed of the op amps can have some difficulty in minimizing the presence of thermal tails. This is because the process provides each transistor a separate dielectric "tub." While this dielectric isolation reduces the parasitic capacitance and greatly speeds up electrical performance, it also provides thermal insulation that slows the dissipation of heat to the substrate.

The seriousness of long tails depends on the application. For example, some systems sample at rates compatible with the initial short-term settling time and are not seriously affected by longer term drifts. Communication systems and others, where the frequency domain properties of the converted signal are most important, are examples of such systems. Although long-term settling errors can produce variations in gain and offset, the long-term thermal tails will have minimal contribution to the distortion products of the digitized signal. For these systems, frequency domain measurements—such as distortion products—are more important than time domain measurements, such as settling time.

On the other hand, systems such as video and scanners might produce a step input, followed by a long-duration plateau of constant value. During this time, repetitive A/D conversions of the op amp output signal will track the long-term settling characteristic. For these systems it is important to understand the long term settling characteristics of the op amp.

The figures below illustrate the long- and short-term settling patterns for the AD8036, a unity-gain-stable high-speed clamp amp that is a good candidate for an A/D driver in high speed systems. The figure at left shows that after the initial large transition, the output is still about 0.09% from its long-term final value. However, the right-hand figure shows, on a 300× faster scale, that after about 16 ns the output has entered a local 0.01% short-term settling region which can be usefully sampled by some systems. The distortion of the AD8036 is extremely low (2nd and 3rd harmonics down by more than 65 dB with 500-Ω load) so it would be a good candidate in systems where this kind of performance is critical.



Reference: Demrow, Robert, "Settling time of operational amplifiers," in *The Best of Analog Dialogue*, 1967 to 1991, pages 32-42.

Analog-Digital Conversion Handbook. Norwood, MA; Analog Devices, 1986, pp. 312-317 and 436-439 (DAC settling time).

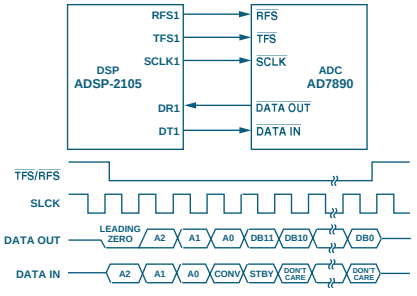
Ask The Applications Engineer—19

INTERFACING TO SERIAL CONVERTERS—I

by Eamon Nash

Q. I need data converters to fit in a tight space, and I suspect that a serial interface will help. What do I need to know to choose and use one?

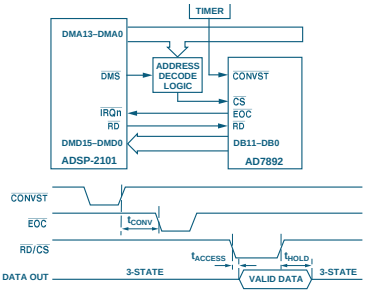
A. Let's start by looking at how a serial interface works and then compare it to a parallel interface. In doing this we will dispel some myths about serial data converters.



The figure shows an AD7890 8-Channel multiplexed 12-bit serial A/D converter (ADC) connected to the serial port of an ADSP-2105 digital signal processor (DSP). Also shown is the timing sequence that the DSP uses to communicate with the ADC. The 12 bits that constitute the conversion result are transmitted as a serial data stream over a single line. The data stream also includes three additional bits that identify the input channel that the AD7890's multiplexer is currently selecting. To distinguish the bits of the serial data stream from one another, a clock signal (SCLK) must be provided, usually by the DSP; However, sometimes the ADC supplies this clock as an output. The DSP usually (but not always) supplies an additional framing pulse that is active either for one cycle at the beginning of the communication or, as shown (TFS/RFS), for the duration of the transmission.

In this example, the DSP's serial port is used to program an internal 5-bit register in the ADC. The register's bits control such functions as selecting the channel to be converted, putting the device in power-down mode, and starting a conversion. It should be evident that the serial interface, in this case, must be bi-directional.

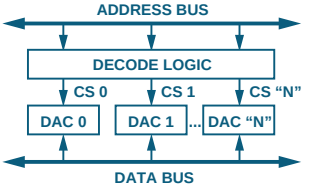
A parallel ADC, on the other hand, connects directly (or possibly through buffers) to the data bus of the processor it is interfaced with. The figure shows the AD7892 interfaced to an ADSP-2101. When a conversion is complete, the AD7892 interrupts the DSP, which responds by doing a single read of the ADC's decoded memory address.



The key difference between serial and parallel data converters lies in the number of interface lines required. From a space saving point of view, serial converters offer a clear advantage because of reduced device pin-count. This makes it possible to package a 12-bit serial ADC or DAC in an 8-pin DIP or SO package. More significantly, board space is saved because serial interface connections require fewer PCB tracks.

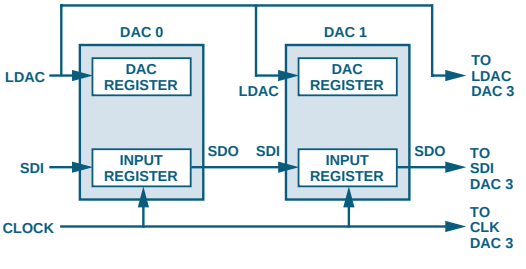
Q. My digital-to-analog converters have to be physically remote from the central processor and from one another. What is the best way to approach this?

A. Initially, you must decide whether to use serial or parallel DACs. With parallel DACs, you could map each one into a memory mapped I/O location, as shown in the figure. You would then program each DAC by simply doing a Write command to the appropriate I/O location. However, this configuration has a significant disadvantage. It requires a parallel data bus, along with some control signals, to all of the remote locations. Clearly, a serial interface, that can have as few as two wires, is much more economical.



Serial converters cannot in general be mapped into a processor's memory. But a number of serial DACs could be connected to the serial I/O port of the processor. Then, other ports on the processor could be used to generate Chip Select signals to enable the DACs individually. The Chip Select signals will require a line from each device to the interface. But there may be a limit to the number of lines on the processor that can be configured to transmit Chip Select signals.

One way of getting around this problem is to use serial DACs that can be daisy-chained together. The figure shows how to connect multiple DACs to a single I/O port. Each DAC has a Serial Data Out (SDO) pin that connects to the Serial Data In (SDI) pin of the next DAC in the chain. LDAC and SCLK are fed in parallel to all the DACs in the chain. Because the data clocked into SDI eventually appears at SDO (N clock cycles later), a single I/O port can address multiple DACs. However, the port must output a long data stream (N bits per DAC times the number of devices in the chain). The great advantage of this configuration is that device decoding is not needed. All devices are effectively at the same I/O location. The main drawback of daisy chaining is accessibility (or latency). To change the state of even a single DAC, the processor must still output a complete data stream from the I/O port.

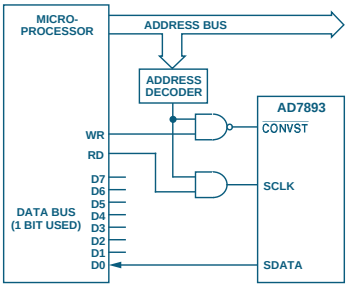


Q. If serial data converters save so much space and wire, why aren't they used in every space-sensitive application?

A. A major disadvantage of serial interfacing is the tradeoff of speed for space. For example, to program a parallel DAC, just place the data on the data bus and clock it into the DAC with a single pulse. However, when writing to a serial DAC, the bits must be clocked in sequentially (N clock pulses for an N-bit converter) and followed by a Load pulse. The processor's I/O port spends a relatively large amount of time communicating with a serial converter. Consequently, serial converters with throughput rates above 500 ksps are uncommon.

Q. My 8-bit processor doesn't have a serial port. Is there a way to interface a serial 12-bit ADC like the AD7893 to the processor's parallel bus?

A. It can of course be done using an external shift register, which is loaded serially (and asynchronously), then clocked into the processor's parallel port. However, if the sense of the question is "without external logic", the serial ADC can be interfaced as if it were a 1-bit parallel ADC. Connect the converter's SDATA pin to one of the processor's data bus lines (it is connected to D0 in the diagram). Using some decode logic, the converter can be mapped into one of the processor's memory locations so that the result of the conversion can be read with 12 successive Read commands. Then additional software commands integrate the LSBs of the 12 bytes that were read into a single 12-bit parallel word.



This technique, which is sometimes referred to as "bit banging", is very inefficient from a software perspective. But it may be acceptable in applications in which the processor runs much faster than the converter.

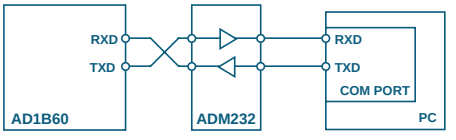
Q. In the last example, a gated version of the processor's write signal was used to start conversions on the AD7893. Are there problems with that approach?

A. I am glad you spotted that. In this example, a conversion can be initiated by doing a dummy write to the AD7893's mapped memory location. No data is exchanged, but the processor provides the write pulse needed to begin the conversion. From a hardware perspective, this configuration is very simple because it avoids the need to generate a conversion signal.

However, the technique is not recommended in ac data-acquisition applications, in which signals must be sampled periodically. Even if the processor is programmed to do periodic writes to the ADC, phase jitter on the Write pulse will seriously degrade the attainable signal-to-noise ratio (SNR). The gating process may make the Write signal jitter even worse. A sampling clock phase jitter level of as little as 1 ns, for example, would degrade the SNR of an ideal 100-kHz sine wave to about 60 dB (less than 10 effective bits of resolution). There is also an additional danger that overshoot and noise on the sampling signal will further degrade the integrity of the analog to digital conversion.

Q. When should I choose a converter with an asynchronous serial interface?

A. An asynchronous link allows devices to exchange unclocked data with each other. The devices must initially be programmed to use the identical data formats. This involves setting a particular data rate (usually expressed in baud, or bits per second). A convention, that defines how to initiate and end transmissions, is also necessary. We do this using identifiable data sequences called *start* and *stop* bits. The transmission may also include *parity* bits that facilitate error detection.



The figure shows how the AD1B60 Digitizing Signal Conditioner interfaces to a PC's asynchronous COM Port. This is a 3-wire bidirectional interface (the ground lines have been omitted for clarity). Notice that the receive and transmit lines exchange roles at the other end of the line.

An asynchronous data link is useful in applications in which devices communicate only sporadically. Since start and stop bits are included in every transmission, a device can initiate communication at any time by simply outputting its data. The number of connections between devices is reduced because clocking and control signals are no longer necessary.

Q. The data sheet of an ADC I am considering recommends using a non-continuous clock on the serial interface. Why?

A. The specification probably requires that the clock be kept inactive while the conversion is in progress. Some ADCs require this because a continuous data clock can feed through to the analog section of the device and adversely affect the integrity of the conversion. A continuous clock signal can be discontinued during conversion if the I/O port has a framing pulse; it is used as a gating signal that enables the serial clock to the converter only during data transfer.

Q. What makes a device SPI or MICROWIRE compatible?

A. SPI (Serial Peripheral Interface) and MICROWIRE are serial interface standards developed by Motorola and National Semiconductor, respectively. Most synchronous serial converters can be easily interfaced to these ports; but in some cases additional "glue" logic may be necessary.

Q. O.K. I decided to put prejudice aside and use a serial ADC in my current design. I have just wired it up as the data sheet specifies. When my micro reads the conversion result, the ADC always seems to output FFF_{HEX}. What's happening?

A. Perhaps you are having a communications problem. We need to look at the connections between the ADC and the processor—and at how the timing and control signals have been set up. We also need to look at the Interrupt structure. The next installment will return to this issue, discussing the problems encountered when designing serial interfaces.

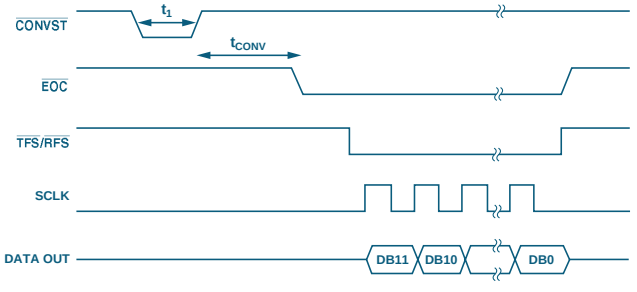
Ask The Applications Engineer—20

INTERFACING TO SERIAL CONVERTERS—II

by Eamon Nash

Q. At the end of our discussion in the last issue, I was having a problem establishing communication between my ADC and my microcontroller. If you recall, the microcontroller always seemed to be reading a conversion result of FFF_{HEX} regardless of the voltage on the analog input. What could be causing this?

A. There are a number of possible timing-related error sources. You could start trouble-shooting this problem by connecting all of the timing signals either to a logic analyzer or to a multi-channel oscilloscope (at least three channels are needed to look at all signals simultaneously). What you would see on the screen would look similar to the timing diagram in the figure below. First make sure that a Start Conversion command (CONVST) is being generated (coming either from the micro or from an independent oscillator). A frequent mistake is to apply a CONVST signal with the wrong polarity. The conversion is still performed, but not when you expect it to be. It is also important to remember that there is usually a minimum pulse width requirement on the CONVST signal (typically about 50 ns). The standard Write or Read pulse from fast microprocessors may not satisfy this requirement. If too short, the pulse width can be extended by inserting software Wait states.



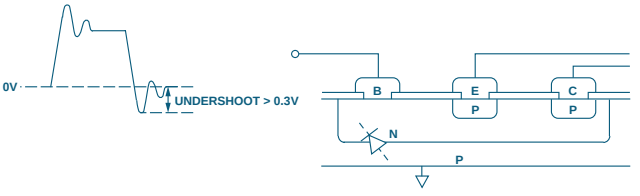
Make certain that the micro is waiting for the conversion to be completed before the Read cycle begins. Your software should either be taking note of the time required to convert or be waiting for an End of Conversion (EOC) indicator from the ADC to generate an interrupt in the micro. Make sure that the polarity of the EOC signal is correct, otherwise the ADC will cause an interrupt while the conversion is in progress. If the micro is not responding to the interrupt, you should examine the configuration of the interrupt in your software.

It is also important to consider the state of the serial clock line (SCLK) while it is not addressing the converter. As I mentioned in our previous discussion, some DACs and ADCs do not operate correctly with continuous serial clocks. In addition to this, some devices require that the SCLK signal always idles in one particular state.

Q. O.K. I've found and corrected some bugs in my software and things seem to be improving. The data from the converter are changing as I vary the input voltage but the conversion results seem to have no recognizable format.

A. Once again there are a number of possible error sources. The ADC will be outputting its conversion result either in straight binary or in twos complement format (BCD data converters are no longer widely used). Check that your micro is configured to accept the appropriate format. If the micro can't be configured to accept twos complement directly, you can convert the data to straight binary by exclusive-or'ing the number with 100 . . . 00 binary.

Normally the leading edge of the serial clock (either rising or falling) will enable the data out of the ADC and onto the data bus. The trailing edge then clocks the data into the micro. Make sure that both micro and ADC are operating under the same convention and that all Setup and Hold times are being met. A conversion result that is exactly half or double what one would expect is a tell-tale sign that the data (especially the MSB) is being clocked on the wrong edge. The same problem would manifest itself in a serial DAC as an output voltage that is half or double the expected value.

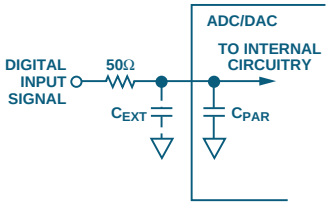


The digital signals driving the converter should be clean. In addition to causing possible long-term damage to the device, overshoot or undershoot can cause conversion and communication errors. The figure shows a signal with a large overshoot spike driving the clock input of a single-supply converter. In this case, the clock input drives the base of an PNP transistor. As is usual practice, the P-type substrate of the device is internally connected to the most negative potential available—in this case, ground. An excursion of more than 0.3 volts below ground on the SCLK line is sufficient to begin turning on a parasitic diode between the N-type base and the P-type substrate. If this happens frequently, over the long term, it may lead to damage to the device.

In the short term, though not causing damage, energizing the normally inert substrate affects other transistors in the device and can lead to multiple clock pulses being detected for each pulse applied. The resulting jitter is a serious matter in serial converters—but is less of a problem in parallel converters, because the Read and Write cycles generally depend upon the first applied pulse; subsequent pulses are ignored. However, the noise performance on both serial and parallel converters can suffer if signals of this kind are present during conversion.

The figure shows how overshoot can be easily reduced. A small resistor is placed in series on the digital line that is causing the problem. This resistance will combine with C_{par}, the parasitic capacitance of the digital input, to form a low-pass filter which should eliminate any ringing on the received signal. Typically a 50-Ω resistor is recommended, but some experimentation may be necessary. It may also be necessary to add an external capacitance from the input to ground if the internal capacitance

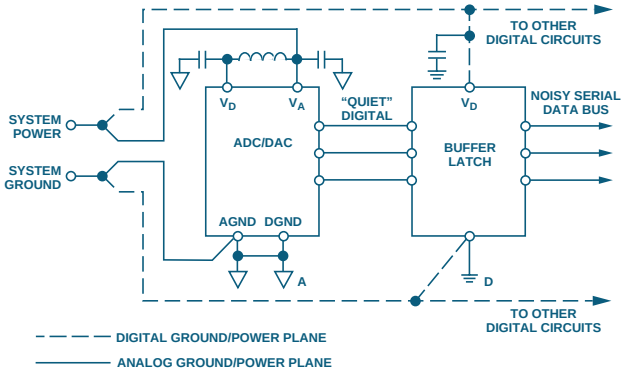
of the digital input is insufficient. Here again, experimentation is necessary—but a good starting point would be about 10 pF.



Q. You mentioned that clock overshoot can degrade the noise performance of a converter. Is there anything else I can do from an interfacing point of view to get a good signal to noise ratio?

A. Because your system is operating in a mixed-signal environment (i.e., analog and digital), the grounding scheme is critical. You probably know that—because digital circuitry is noisy—analog and digital grounds should be kept separate, joined at only one point. This connection is usually made at the power supply. In fact, if the analog and digital devices are powered from a common supply, as might be the case in a +5 V or +3.3 V single-supply system, there is no choice but to connect the grounds back at the supply. But the data sheet for the converter probably has an instruction to connect the pins AGND and DGND at the device! So how can one avoid creating a ground loop that can result if the grounds are connected in two places?

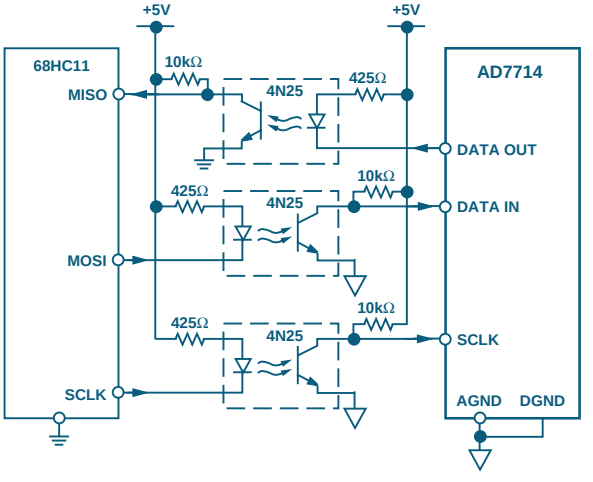
The figure below shows how to resolve this apparent dilemma. The key is that the AGND and DGND labels on the converter's pins refer to the parts of the converter to which those pins are connected. The device as a whole should be treated as *analog*. So after the AGND and DGND pins have been connected together, there should be a single connection to the system's analog ground. True, this will cause the converter's digital currents to flow in the analog ground plane, but this is generally a lesser evil than exposing the converter's DGND pin to a noisy digital ground plane. This example also shows a digital buffer, referred to digital ground, to isolate the converter's serial data pins from a noisy serial bus. If the converter is making a point-to-point connection to a micro, this buffer may be unnecessary.



The figure also shows how to deal with the increasingly common challenge of powering a mixed-signal system with a single power supply. As in the grounding case, we run separate power lines (preferably power planes) to the analog and digital portions of the circuit. We treat the digital power pin of the converter as analog. But some isolation from the analog power pin, in the form of an inductor, is appropriate. Remember that both power pins of the converter should have separate decoupling capacitors. The data sheet will recommend appropriate capacitors, but a good rule of thumb is 0.1 µF. If space permits, a single 10-µF capacitor per device should also be included.

Q. I want to design an isolated serial interface between an ADC and a microcontroller using opto-isolators. What should I be aware of when using these devices?

A. Opto-isolators (also known as opto-couplers) can be used to create a simple and inexpensive high-voltage isolation barrier. The presence of a galvanic isolation barrier between converter and micro also means that analog and digital system grounds no longer need to be connected. As shown in the figure, an isolated serial interface between the AD7714 precision ADC and the popular 68HC11 microcontroller can be implemented with as few as three optoisolators.



The designer should be aware, though, that the use of optoisolators having relatively slow rise and fall times with CMOS converters can cause problems, even when the serial communication is running at a slow speed.

CMOS logic inputs are designed to be driven by a definite logic zero or logic one. In these states, they source and sink a minimal amount of current. However, when the input voltage is in transition between logic zero and logic one (0.8 V to 2.0 V), the gate will consume an increased amount of current. If the opto-isolators used have relatively slow rise and fall times, the excessive amount of time spent in the dead-band will cause self-heating in the gate. This self-heating tends to shift the threshold voltage of the logic gate upwards, which can lead to a single clock edge being interpreted by the converter as multiple clock pulses. To prevent this threshold jitter, the lines coming from the optoisolators should be buffered using Schmitt trigger circuits, to deliver fast, sharp edges to the converter.

Ask The Applications Engineer—21

by Steve Guinta

CAPACITANCE AND CAPACITORS

I. Understanding the Parasitic Effects In Capacitors:

Q. I need to understand how to select the right capacitor for my application, but I'm not clear on the advantages and disadvantages of the many different types.

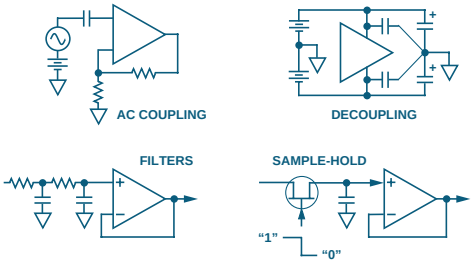
A. Selecting the right capacitor type for a particular application really isn't that difficult. Generally, you'll find that most capacitors fall into one of four application categories:

•*AC coupling*, including bypassing (passing ac signals while blocking dc)

•*decoupling* (filtering ac or high frequencies superimposed on dc or low frequencies in power, reference, and signal circuitry)

•active/passive RC *filters or frequency-selective networks*

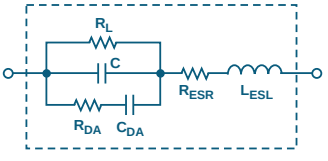
•*analog integrators* and *sample-and-hold* circuits (acquiring and storing charge)



Even though there are more than a dozen or so popular capacitor types—including poly, film, ceramic, electrolytic, etc.—you'll find that, in general, only one or two types will be best suited for a particular application, because the salient imperfections, or "parasitic effects" on system performance associated with other types of capacitors will cause them to be eliminated.

Q. What are these "parasitic effects" you're talking about?

A. Unlike an "ideal" capacitor, a "real" capacitor is typified by additional "parasitic" or "non-ideal" components or behavior, in the form of resistive and inductive elements, nonlinearity, and dielectric memory. The resulting characteristics due to these components are generally specified on the capacitor manufacturer's data sheet . Understanding the effects of these parasitics in each application will help you select the right capacitor type.

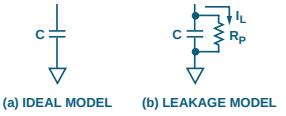


Model of a "Real" Capacitor

Q. OK, so what are the most important parameters describing non-ideal capacitor behavior?

A. The four most common effects are *leakage* (parallel resistance), *equivalent series resistance* (ESR), *equivalent series inductance* (ESL), and *dielectric absorption* (memory).

Capacitor Leakage, R_P: Leakage is an important parameter in ac coupling applications, in storage applications, such as analog integrators and sample-holds, and when capacitors are used in high-impedance circuits.



In an ideal capacitor, the charge, Q, varies only in response to current flowing externally. In a real capacitor, however, the leakage resistance allows the charge to trickle off at a rate determined by the R-C time constant.

Electrolytic-type capacitors (tantalum and aluminum), distinguished for their high capacitance, have very high leakage current (typically of the order of about 5-20 nA per µF) due to poor isolation resistance, and are not suited for storage or coupling applications.

The best choices for coupling and/or storage applications are Teflon (polytetrafluorethylene) and the other "poly" types (polyproplene, polystyrene, etc).

Equivalent Series Resistance (ESR), R_S: The equivalent series resistance (ESR) of a capacitor is the resistance of the capacitor leads in series with the equivalent resistance of the capacitor plates. ESR causes the capacitor to dissipate power (and hence produce loss) when high ac currents are flowing. This can have serious consequences at RF and in supply decoupling capacitors carrying high ripple currents, but is unlikely to have much effect in precision high-impedance, low-level analog circuitry.

Capacitors with the lowest ESR include both the mica and film types.

Equivalent Series Inductance (ESL), L_S: The equivalent series inductance (ESL) of a capacitor models the inductance of the capacitor leads in series with the equivalent inductance of the capacitor plates. Like ESR, ESL can also be a serious problem at high (RF) frequencies, even though the precision circuitry itself may be operating at DC or low frequencies. The reason is that the transistors used in precision analog circuits may have gain extending up to transition frequencies (F_t) of hundreds of MHz, or even several GHz, and can amplify resonances involving low values of inductance. This makes it essential that the power supply terminals of such circuits be decoupled properly at high frequency.

Electrolytic, paper, or plastic film capacitors are a poor choice for decoupling at high frequencies; they basically consist of two sheets of metal foil separated by sheets of plastic or paper dielectric and formed into a roll. This kind of structure has considerable self inductance and acts more like an inductor than a capacitor at frequencies exceeding just a few MHz.

A more appropriate choice for HF decoupling is a monolithic, ceramic-type capacitor, which has very low series inductance. It consists of a multilayer sandwich of metal films and ceramic dielectric, and the films are joined in parallel to bus-bars, rather than rolled in series.

A minor tradeoff is that monolithic ceramic capacitors can be microphonic (i.e., sensitive to vibration), and some types may

even be self-resonant, with comparatively high Q, because of the low series resistance accompanying their low inductance. Disc ceramic capacitors, on the other hand, are sometime quite inductive, although less expensive.

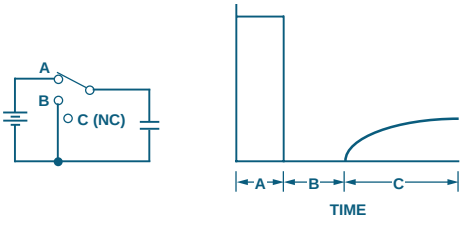
Q. I've seen the term "dissipation factor" used in capacitor selection charts. What is it?

A. Good question. Since leakage, ESR, and ESL are almost always difficult to spec separately, many manufacturers will lump leakage, ESR and ESL into a single specification known as *dissipation factor*, or DF, which basically describes the inefficiency of the capacitor. DF is defined as the ratio of energy dissipated per cycle to energy stored per cycle. In practice, this is equal to the power factor for the dielectric, or the cosine of the phase angle. If the dissipation at high frequencies is principally modeled as series resistance, at a critical frequency of interest, the ratio of equivalent series resistance, ESR, to total capacitive reactance is a good estimate of DF,

$$DF \approx \omega R_S C\,.$$

Dissipation factor also turns out to be the equivalent to the reciprocal of the capacitor's figure of merit, or Q, which is also sometimes included on the manufacturer's data sheet.

Dielectric Absorption, RDA, CDA: Monolithic ceramic capacitors are excellent for HF decoupling, but they have considerable *dielectric absorption*, which makes them unsuitable for use as the hold capacitor of a sample-hold amplifier (SHA). Dielectric absorption is a hysteresis-like internal charge distribution that causes a capacitor which is quickly discharged and then open-circuited to appear to recover some of its charge. Since the amount of charge recovered is a function of its previous charge, this is, in effect, a charge memory and will cause errors in any SHA where such a capacitor is used as the hold capacitor.



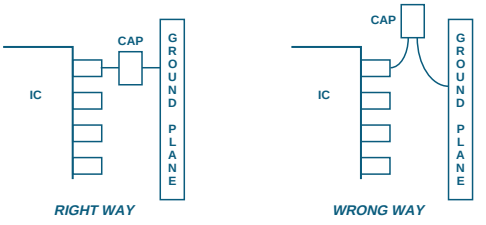
Capacitors that are recommended for this type of application include the "poly" type capacitors we spoke about earlier, i.e., polystyrene, polypropylene, or Teflon. These capacitor types have very low dielectric absorption (typically <0.01%).

The characteristics of capacitors in general are summarized in the capacitor comparison chart (page 21).

A note about high-frequency decoupling in general: The best way to insure that an analog circuit is adequately decoupled at both high and low frequencies is to use an electrolytic-type capacitor, such as a tantalum bead, in parallel with a monolithic ceramic one. The combination will have high capacitance at low frequency, and will remain capacitive up to quite high frequencies. It's generally not necessary to have a tantalum capacitor on each individual IC, except in critical cases; if there is less than 10 cm of reasonably wide PC track between each IC and the tantalum capacitor, it's possible to share one tantalum capacitor among several ICs.

Another thing to remember about high frequency decoupling is the actual physical placement of the capacitor. Even short lengths of wire have considerable inductance, so mount the HF decoupling capacitors as close as possible to the IC, and ensure that leads consist of short, wide PC tracks.

Ideally, HF decoupling capacitors should be surface-mount parts to eliminate lead inductance, but wire-ended capacitors are ok, providing the device leads are no longer than 1.5 mm.



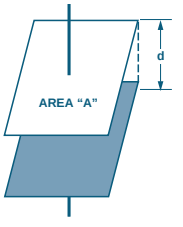
- USE LOW INDUCTANCE CAPACITORS (MONOLITHIC CERAMIC)
- MOUNT CAPACITOR CLOSE TO IC
- USE SURFACE MOUNT TYPE
- USE SHORT, WIDE PC TRACKS

II. Stray Capacitance:

A. Now that we've talked about the parasitic effects of capacitors as components, let's talk about another form of parasitic known as "stray" capacitance.

Q. What's that?

A. Well, just like a parallel-plate capacitor, stray capacitors are formed whenever two conductors are in close proximity to each other (especially if they're running in parallel), and are not shorted together or screened by a conductor serving as a Faraday shield.

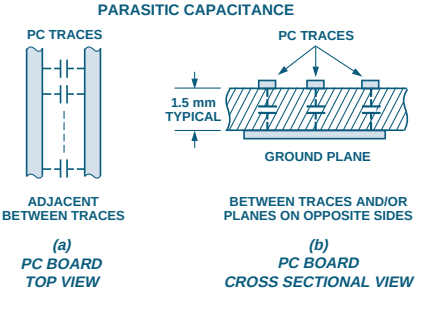


$$C = 0.0085 \times E_R \times \frac{A}{d}$$

C = CAPACITANCE IN pF
E_R = DIELECTRIC CONSTANT RELATIVE TO AIR
A = AREA OF PARALLEL CONDUCTORS IN mm^2
d = DISTANCE BETWEEN CONDUCTORS IN mm

Capacitor Model

Stray or "parasitic" capacitance commonly occurs between parallel traces on a PC board or between traces/planes on opposite sides of a PC board. The occurrence and effects of stray capacitance—especially at very high frequencies—are unfortunately often overlooked during circuit modelling and can lead to serious performance problems when the system circuit board is constructed and assembled; examples include greater noise, reduced frequency response, even instability.



For instance, if the capacitance formula is applied to the case of traces on opposite sides of a board, then for general purpose PCB material (E_R = 4.7, d = 1.5 mm), the capacitance between conductors on opposite sides of the board is just under 3 pF/cm^2. At a frequency of 250 MHz, 3 pF corresponds to a reactance of 212.2 ohms!

Q. So how can I eliminate stray capacitance?

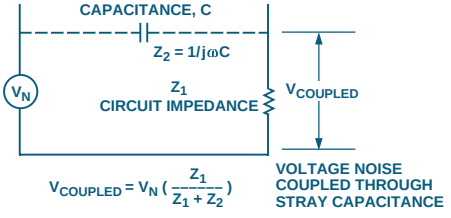
A. You can never actually "eliminate" stray capacitance; the best you can do is take steps to minimize its effects in the circuit.

Q. How do I do that?

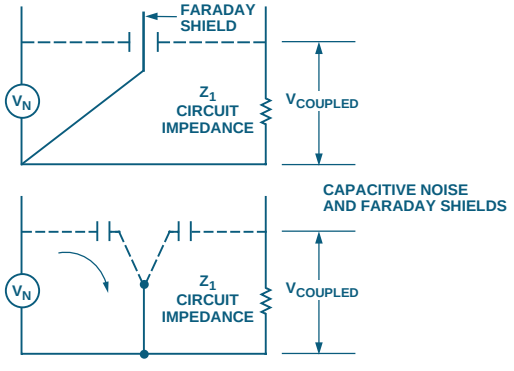
A. Well, one way to minimize the effects of stray coupling is to use a Faraday shield, which is simply a grounded conductor between the coupling source and the affected circuit.

Q. How does it work?

A. Look at the Figure; it is an equivalent circuit showing how a high-frequency noise source, V_N, is coupled into a system impedance, Z, through a stray capacitance, C. If we have little or no control over V_n or the location of Z_1, the next best solution is to interpose a Faraday shield:



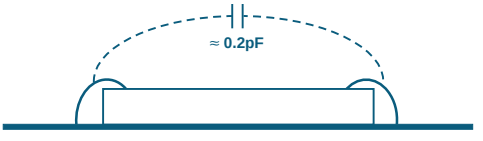
As shown, below, the Faraday shield interrupts the coupling electric field. Notice how the shield causes the noise and coupling currents to return to their source without flowing through Z_1.



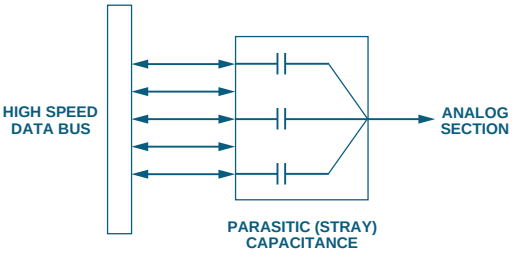
Another example of capacitive coupling is in side-brazed ceramic IC packages. These DIP packages have a small, square, conducting Kovar lid soldered onto a metallized rim on the ceramic package top. Package manufacturers offer only two options: the metallized rim may be connected to one of the corner pins of the package, or it may be left unconnected. Most logic circuits have a ground pin at one of the package corners, and therefore the lid is grounded. But many analog circuits do not have a ground pin at a package corner, and the lid is left floating. Such circuits turn out to be far more vulnerable to electric field noise than the same chip in a plastic DIP package, where the chip is unshielded.

Whatever the environmental noise level, it is good practice for the user to ground the lid of any side brazed ceramic IC where the lid is not grounded by the manufacturer. This can be done with a wire soldered to the lid (this will not damage the device, as the chip is thermally and electrically isolated from the lid). If soldering to the lid is unacceptable, a grounded phosphor-bronze clip may be used to make the ground connection, or conductive paint can be used to connect the lid to the ground pin. *Never attempt to ground such a lid without verifying that it is, in fact, unconnected*; there do exist device types with the lid connected to a supply rail rather than to ground!

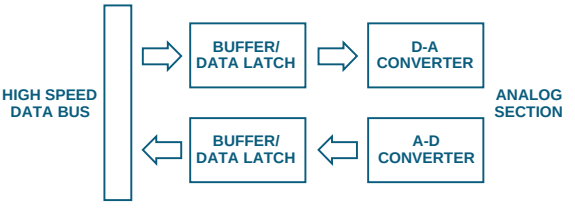
One case where a Faraday shield is impracticable is between the bond wires of an integrated circuit chip. This has important consequences. The stray capacitance between two chip bond wires and their associated leadframes is of the order of 0.2 pF; observed values generally lie between 0.05 and 0.6 pF.



Consider a high-resolution converter (ADC or DAC), which is connected to a high-speed data bus. Each line of the data bus, (which will be switching at around 2 to 5 V/ns),will be able to influence the converter's analog port via this stray capacitance; the consequent coupling of digital edges will degrade the performance of the converter.



This problem may be avoided by isolating the data bus, interposing a latched buffer as an interface. Although this solution involves an additional component that occupies board area, consumes power, and adds cost, it can significantly improve the converter's signal-to-noise.



CAPACITOR COMPARISON CHART

TYPE	TYPICAL DIELECTRIC ABSORPTION	ADVANTAGES	DISADVANTAGES
NPO ceramic	<0.1%	Small case size Inexpensive Good stability Wide range of values Many vendors Low inductance	DA generally low, but may not be specified Limited to small values (10 nF)
Polystyrene	0.001% to 0.02%	Inexpensive Low DA available Wide range of values Good stability	Damaged by temperature > +85°C Large case size High inductance
Polypropylene	0.001% to 0.02%	Inexpensive Low DA available Wide range of values	Damaged by temperature > +105°C Large case size High inductance
Teflon	0.003% to 0.02%	Low DA available Good stability Operational above +125°C Wide range of values	Relatively expensive Large size High inductance
MOS	0.01%	Good DA Small Operational above +125°C Low inductance	Limited availability Available only in small capacitance values
Polycarbonate	0.1%	Good stability Low cost Wide temperature range	Large size DA limits to 8-bit applications High inductance
Polyester	0.3% to 0.5%	Moderate stability Low cost Wide temperature range Low inductance (stacked film)	Large size DA limits to 8-bit applications High inductance
Monolithic ceramic (High K)	>0.2%	Low inductance Wide range of values	Poor stability Poor DA High voltage coefficient
Mica	>0.003%	Low loss at HF Low inductance Very stable Available in 1% values or better	Quite large Low values (<10 nF) Expensive
Aluminum electrolytic	High	Large values High currents High voltages Small size	High leakage Usually polarized Poor stability Poor accuracy Inductive
Tantalum electrolytic	High	Small size Large values Medium inductance	Quite high leakage Usually polarized Expensive Poor stability Poor accuracy

Ask The Applications Engineer—22

by Erik Barnes

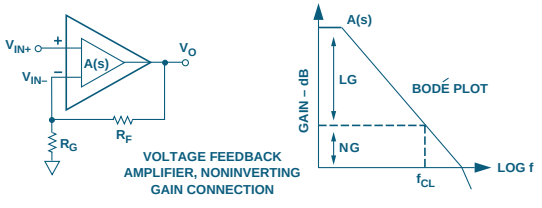
CURRENT FEEDBACK AMPLIFIERS—I

Q. I'm not sure I understand how current-feedback amplifiers work as compared with regular op amps. I've heard that their bandwidth is constant regardless of gain. How does that work? Are they the same as transimpedance amplifiers?

A. Before looking at any circuits, let's define voltage feedback, current feedback, and transimpedance amplifier. *Voltage feedback*, as the name implies, refers to a closed-loop configuration in which the error signal is in the form of a voltage. Traditional op amps use voltage feedback, that is, their inputs will respond to voltage changes and produce a corresponding output voltage. *Current feedback* refers to any closed-loop configuration in which the error signal used for feedback is in the form of a current. A current feedback op amp responds to an error current at one of its input terminals, rather than an error voltage, and produces a corresponding output voltage. Notice that both open-loop architectures achieve the same closed-loop result: zero differential input voltage, and zero input current. The ideal voltage feedback amplifier has high-impedance inputs, resulting in zero input current, and uses voltage feedback to maintain zero input voltage. Conversely, the current feedback op amp has a low impedance input, resulting in zero input *voltage*, and uses current feedback to maintain zero input *current*.

The transfer function of a *transimpedance amplifier* is expressed as a voltage output with respect to a current input. As the function implies, the open-loop "gain", v_O/i_{IN}, is expressed in ohms. Hence a current-feedback op amp can be referred to as a *transimpedance amplifier*. It's interesting to note that the closed-loop relationship of a voltage-feedback op amp *circuit* can also be configured as a transimpedance, by driving its dynamically low-impedance summing node with current (e.g., from a photodiode), and thus generating a voltage output equal to that input current multiplied by the feedback resistance. Even more interesting, since ideally any op amp application can be implemented with either voltage or current feedback, this same I-V converter can be implemented with a current feedback op amp. When using the term *transimpedance amplifier*, understand the difference between the specific current-feedback op amp architecture, and any closed-loop I-V converter circuit that acts like transimpedance.

Let's take a look at the simplified model of a voltage feedback amplifier. The noninverting gain configuration amplifies the difference voltage, $(V_{IN+} - V_{IN-})$, by the open loop gain $A(s)$ and feeds a portion of the output back to the inverting input through the voltage divider consisting of R_F and R_G. To derive the closed-loop transfer function of this circuit, V_o/V_{IN+}, assume



that no current flows into the op amp (infinite input impedance); both inputs will be at about the same potential (negative feedback and high open-loop gain)).

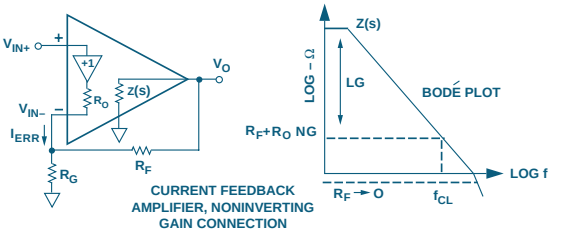
With $V_o = (V_{IN+} - V_{IN-})A(s)$

and $V_{IN-} = \frac{R_G}{R_G + R_F} V_o$

substitute and simplify to get:

$$\frac{V_o}{V_{IN}} = \left(1 + \frac{R_F}{R_G}\right)\frac{1}{1 + \frac{1}{LG}} \text{ where } LG = \frac{A(s)}{1 + \frac{R_F}{R_G}}$$

The closed-loop bandwidth is the frequency at which the loop gain, LG, magnitude drops to unity (0 dB). The term, $1 + R_F/R_G$, is called the *noise gain* of the circuit; for the noninverting case, it is also the signal gain. Graphically, the closed-loop bandwidth is found at the intersection of the open-loop gain, $A(s)$, and the noise gain, NG, in the Bodé plot. High noise gains will reduce the loop gain, and thereby the closed-loop bandwidth. If $A(s)$ rolls off at 20 dB/decade, the gain-bandwidth product of the amplifier will be constant. Thus, an increase in closed-loop gain of 20 dB will reduce the closed-loop bandwidth by one decade.



Consider now a simplified model for a current-feedback amplifier. The noninverting input is the high-impedance input of a unity gain buffer, and the inverting input is its low-impedance output terminal. The buffer allows an error current to flow in or out of the inverting input, and the unity gain forces the inverting input to track the noninverting input. The error current is mirrored to a high impedance node, where it is converted to a voltage and buffered at the output. The high-impedance node is a frequency-dependent impedance, $Z(s)$, analogous to the open-loop gain of a voltage feedback amplifier; it has a high dc value and rolls off at 20 dB/decade.

The closed-loop transfer function is found by summing the currents at the V_{IN-} node, while the buffer maintains $V_{IN+} = V_{IN-}$. If we assume, for the moment, that the buffer has zero output resistance, then $R_o = 0\Omega$

$$\frac{V_o - V_{IN-}}{R_F} + \frac{-V_{IN-}}{R_G} + I_{err} = 0 \text{ and } I_{err} = V_0 / Z(s)$$

Substituting, and solving for V_o/V_{IN+}

$$\frac{V_o}{V_{IN+}} = \left(1 + \frac{R_F}{R_G}\right)\frac{1}{1 + \frac{1}{LG}}, \text{ where } LG = \frac{Z(s)}{R_F}$$

The closed-loop transfer function for the current feedback amplifier is the same as for the voltage feedback amplifier, but the loop gain ($1/LG$) expression now depends only on R_F, the

feedback transresistance—and not $(1 + R_F/R_G)$. Thus, the closed-loop bandwidth of a current feedback amplifier will vary with the value of R_F, but not with the noise gain, $1 + R_F/R_G$. The intersection of R_F and $Z(s)$ determines the loop gain, and thus the closed-loop bandwidth of the circuit (see Bodé plot). Clearly the gain-bandwidth product is not constant—an advantage of current feedback.

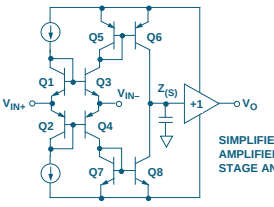
In practice, the input buffer's non-ideal output resistance will be typically about 20 to 40 Ω, which will modify the feedback transresistance. The two input voltages will not be exactly equal. Making the substitution into the previous equations with $V_{IN-} = V_{IN+} - I_{err}R_o$, and solving for V_o/V_{IN+} yields:

$$\frac{V_o}{V_{IN}} = \left(1+\frac{R_F}{R_G}\right)\frac{1}{1+\frac{1}{LG}}, \text{ where } LG = \frac{Z(s)}{R_F + R_o\left(1+\frac{R_F}{R_G}\right)}$$

The additional term in the feedback transresistance means that the loop gain will actually depend somewhat on the closed-loop gain of the circuit. At low gains, R_F dominates, but at higher gains, the second term will increase and reduce the loop gain, thus reducing the closed-loop bandwidth.

It should be clear that shorting the output back to the inverting input with R_G open (as in a voltage follower) will force the loop gain to get very large. With a voltage feedback amplifier, maximum feedback occurs when feeding back the entire output voltage, but the current feedback's limit is a short-circuit current. The lower the resistance, the higher the current will be. Graphically, $R_F = 0$ will give a higher-frequency intersection of $Z(s)$ and the feedback transresistance—in the region of higher-order poles. As with a voltage feedback amplifier, higher-order poles of $Z(s)$ will cause greater phase shift at higher frequencies, resulting in instability with phase shifts > 180 degrees. Because the optimum value of R_F will vary with closed-loop gain, the Bode plot is useful in determining the bandwidth and phase margin for various gains. A higher closed-loop bandwidth can be obtained at the expense of a lower phase margin, resulting in peaking in the frequency domain, and overshoot and ringing in the time domain. Current-feedback device data sheets will list specific optimum values of R_F for various gain settings.

Current feedback amplifiers have excellent slew-rate capabilities. While it is possible to design a voltage-feedback amplifier with high slew rate, the current-feedback architecture is inherently faster. A traditional voltage-feedback amplifier, lightly loaded, has a slew rate limited by the current available to charge and discharge the internal compensation capacitance. When the input is subjected to a large transient, the input stage will saturate and only its tail current is available to charge or discharge the compensation node. With a current-feedback amplifier, the low-impedance input allows higher transient currents to flow into the amplifier as needed. The internal current mirrors convey this input current to the compensation node, allowing fast charging and discharging—theoretically, in proportion to input step size. A faster slew rate will result in a quicker rise time, lower slew-induced distortion and nonlinearity, and a wider large-signal frequency response. The actual slew rate will be limited by saturation of the current mirrors, which can occur at 10 to 15 mA, and the slew-rate limit of the input and output buffers.



SIMPLIFIED CURRENT FEEDBACK AMPLIFIER, ILLUSTRATING INPUT STAGE AND CURRENT MIRRORS.

Q. What about dc accuracy?

A. The dc gain accuracy of a current feedback amplifier can be calculated from its transfer function, just as with a voltage feedback amplifier; it is essentially the ratio of the internal transresistance to the feedback transresistance. Using a typical transresistance of 1 MΩ, a feedback resistor of 1 kΩ, and an R_o of 40 ohms, the gain error at unity gain is about 0.1%. At higher gains, it degrades significantly. Current-feedback amplifiers are rarely used for high gains, particularly when absolute gain accuracy is required.

For many applications, though, the settling characteristics are of more importance than gain accuracy. Although current feedback amplifiers have very fast rise times, many data sheets will only show settling times to 0.1%, because of thermal settling tails— a major contributor to lack of settling precision. Consider the complementary input buffer above, in which the V_{IN-} terminal is offset from the V_{IN+} terminal by the difference in V_{BE} between Q1 and Q3. When the input is at zero, the two V_{BE}s should be matched, and the offset will be small from V_{IN+} to V_{IN-}. A positive step input applied to V_{IN+} will cause a reduction in the V_{CE} of Q3, decreasing its power dissipation, thus increasing its V_{BE}. Diode-connected Q1 does not exhibit a V_{CE} change, so its V_{BE} will not change. Now a different offset exists between the two inputs, reducing the accuracy. The same effect can occur in the current mirror, where a step change at the high-impedance node changes the V_{CE}, and thus the V_{BE}, of Q6, but not of Q5. The change in V_{BE} causes a current error referred back to V_{IN-}, which—multiplied by R_F—will result in an output offset error. Power dissipation of each transistor occurs in an area that is too small to achieve thermal coupling between devices. Thermal errors in the input stage can be reduced in applications that use the amplifier in the inverting configuration, eliminating the common-mode input voltage.

Q. In what conditions are thermal tails a problem?

A. It depends on the frequencies and waveforms involved. Thermal tails do not occur instantaneously; the thermal coefficient of the transistors (which is process dependent) will determine the time it takes for the temperature change to occur and alter parameters—and then recover. Amplifiers fabricated on the Analog Devices high-speed complementary bipolar (CB) process, for example, don't exhibit significant thermal tails for input frequencies above a few kHz, because the input signal is changing too fast. Communications systems are generally more concerned with spectral performance, so additional gain errors that might be introduced by thermal tails are not important. Step waveforms, such as those found in imaging applications, can be adversely affected by thermal tails when dc levels change. For these applications, current-feedback amplifiers may not offer adequate settling accuracy.

Part II will consider common application circuits using current-feedback amplifiers and view their operation in more detail.

Ask The Applications Engineer—23

by Erik Barnes

CURRENT FEEDBACK AMPLIFIERS—II

Part I (*Analog Dialogue* 30-3) covers basic operation of the current-feedback (CF) op-amp. This second part addresses frequently asked questions about common applications.

Q. I now have better understanding of how a current feedback op-amp works, but I'm still confused when it comes to applying one in a circuit. Does the low inverting input impedance mean I can't use the inverting gain configuration?

A. Remember that the inverting mode of operation works *because* of the low-impedance node created at the inverting input. The summing junction of a voltage-feedback (VF) amplifier is characterized by a low input impedance after the feedback loop has settled. A current feedback op amp will, in fact, operate very well in the inverting configuration because of its inherently low inverting-input impedance, holding the summing node at "ground," even before the feedback loop has settled. CF types don't have the voltage spikes that occur at the summing node of voltage feedback op amps in high-speed applications. You may also recall that advantages of the inverting configuration include maximizing input slew rate and reducing thermal settling errors.

Q. So this means I can use a current feedback op-amp as a current-to-voltage converter, right?

A. Yes, they can be configured as I-to-V converters. But there are limitations: the amplifier's bandwidth varies directly with the value of feedback resistance, and the inverting input current noise tends to be quite high. When amplifying low level currents, higher feedback resistance means higher signal-to-(resistor-) noise ratio, because signal gain will increase proportionally, while resistor noise goes as $\sqrt{R}$. Doubling the feedback resistance doubles the signal gain and increases resistor noise by a only factor of 1.4; unfortunately the contribution from current noise is doubled, and, with a current feedback op amp, the signal bandwidth is halved. Thus the higher current noise of CF op amps may preclude their use in many photodiode-type applications. When noise is less critical, select the feedback resistor based on bandwidth requirements; use a second stage to add gain.

Q. I did notice the current noise is rather high in current feedback amplifiers. So will this limit the applications in which I can use them?

A. Yes, the inverting input current noise tends to be higher in CF op amps, around 20 to 30 pA/$\sqrt{\text{Hz}}$. However, the input voltage noise tends to be quite low when compared with similar voltage feedback parts, typically less than 2 nV/$\sqrt{\text{Hz}}$, and the feedback resistance will also be low, usually under 1 kΩ. At a gain of 1, the dominant source of noise will be the inverting-input noise current flowing through the feedback resistor. An input noise current of 20 pA/$\sqrt{\text{Hz}}$ and an R_F of 750 Ω yields 15 nV/$\sqrt{\text{Hz}}$ as the dominant noise source at the output. But as the gain of the circuit is increased (by reducing input resistance), the output noise due to input current noise will not increase, and the amplifier's input *voltage* noise will become the dominant factor. At a gain, of say, 10, the contribution from the input noise current is only 1.5 nV/$\sqrt{\text{Hz}}$ when referred to the input; added to the input voltage noise of the amplifier in RSS fashion, this gives an input-referred noise voltage of only 2.5 nV/$\sqrt{\text{Hz}}$ (neglecting resistor noise). Used thus, the CF op amp becomes attractive for a low noise application.

Q. What about using the classic four-resistor differential configuration? Aren't the two inputs unbalanced and therefore not suitable for this type of circuit?

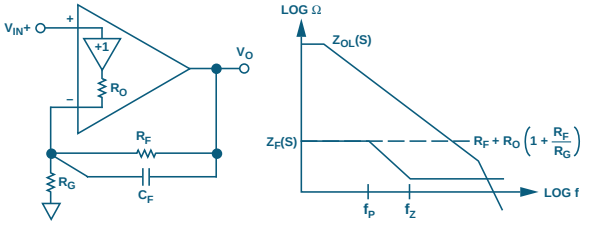
A. I'm glad you asked; this is a common misconception of CF op-amps. True, the inputs are not matched, but the transfer function for the ideal difference amplifier will still work out the same. What about the unbalanced inputs? At lower frequencies, the four-resistor differential amplifier's CMR is limited by the matching of the external resistor ratios, with 0.1% matching yielding about 66 dB. At higher frequencies, what matters is the matching of time constants formed by the input impedances. High-speed voltage-feedback op amps usually have pretty well matched input capacitances, achieving CMR of about 60 dB at 1 MHz. Because the CF amplifier's input stage is unbalanced, the capacitances may not be well matched. This means that small external resistors (100 to 200 Ω) must be used on the noninverting input of some amplifiers to minimize the mismatch in time constants. If careful attention is given to resistor selection, a CF op-amp can yield high frequency CMR comparable to a VF op amp. If higher performance is needed, the best choice would be a monolithic high speed *difference amplifier*, such as the AD830. Requiring no resistor matching, it has a CMR > 75 dB at 1 MHz and about 53 dB at 10 MHz.

Q. What about trimming the amplifier's bandwidth with a feedback capacitor? Will the low impedance at the inverting input make the current feedback op amp less sensitive to shunt capacitance at this node? How about capacitive loads?

A. First consider a capacitor in the feedback path. With a voltage feedback op amp, a pole is created in the noise gain, but a pole and a zero occur in the feedback transresistance of a current feedback op amp, as shown in the figure below. Remember that the phase margin at the intersection of the feedback transresistance and the open loop transimpedance will determine closed-loop stability. Feedback transresistance for a capacitance, C_F, in parallel with R_F, is given by

$$Z_F(s)=\left[R_F+R_O\left(1+\frac{R_F}{R_G}\right)\right]\frac{1+\dfrac{sC_FR_FR_GR_O}{R_FR_G+R_FR_O+R_GR_O}}{1+sC_FR_F}$$

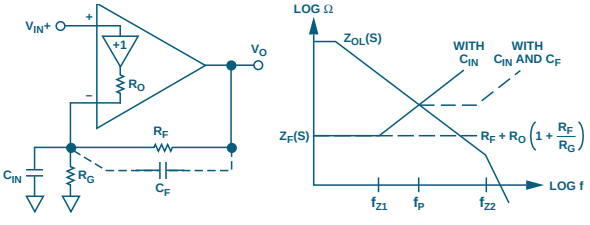
The pole occurs at $1/2\pi R_FC_F$, and the zero occurs higher in frequency at $1/[2\pi(R_F||R_G||R_O)C_F]$. If the intersection of Z_F and Z_{OL} occurs too high in frequency, instability may result from excessive open loop phase shift. If $R_F \rightarrow \infty$, as with an integrator circuit, the pole occurs at a low frequency and very little resistance exists at higher frequencies to limit the loop gain. A CF integrator can be stabilized by a resistor in series with the integrating capacitor to limit loop gain at higher frequencies. Filter topologies that use reactive feedback, such as multiple feedback types, are not suitable for CF op amps; but Sallen-Key filters, where the op amp is used as a fixed-gain block, are feasible. In general, it is not desirable to add capacitance across R_F of a CF op amp.



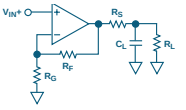
Another issue to consider is the effect of shunt capacitance at the inverting input. Recall that with a voltage feedback amplifier, such capacitance creates a zero in the noise gain, increasing the rate of closure between the noise gain and open loop gain, generating excessive phase shift that can lead to instability if not compensated for. The same effect occurs with a current feedback op amp, but the problem may be less pronounced. Writing the expression for the feedback transresistance with the addition of C_{IN}:

$$Z_F(s)=\left[R_F+R_O\left(1+\frac{R_F}{R_G}\right)\right]\left[1+\frac{sC_{IN}R_FR_GR_O}{R_FR_G+R_FR_O+R_GR_O}\right]$$

A zero occurs at $1/[2\pi(R_F||R_G||R_O)C_{IN}]$, shown in the next figure (f_{Z1}). This zero will cause the same trouble as with a VF amplifier, but the corner frequency of the zero tends to be higher in frequency because of the inherently low input impedance at the inverting input. Consider a wideband voltage feedback op amp with R_F = 750 Ω, R_G = 750 Ω, and C_{IN} = 10 pF. The zero occurs at $1/[2\pi(R_F\|R_G)C_{IN}]$, roughly 40 MHz, while a current feedback op-amp in the same configuration with an R_O of 40 Ω will push the zero out to about 400 MHz. Assuming a unity gain bandwidth of 500 MHz for both amplifiers, the VF amplifier will require a feedback capacitor for compensation, reducing the effect of C_{IN}, but also reducing the signal bandwidth. The CF device will certainly see some additional phase shift from the zero, but not as much because the break point is a decade higher in frequency. Signal bandwidth will be greater, and compensation may only be necessary if in-band flatness or optimum pulse response is required. The response can be tweaked by adding a small capacitor in parallel with R_F to reduce the rate of closure between Z_F and Z_{OL}. To ensure at least 45° of phase margin, the feedback capacitor should be chosen to place a pole in the feedback transresistance where the intersection of Z_F and Z_{OL} occurs, shown here (f_P). Don't forget the effects of the higher frequency zero due to the feedback capacitor (f_{Z2}).



Load capacitance presents the same problem with a current feedback amplifier as it does with a voltage feedback amplifier: increased phase shift of the error signal, resulting in degradation of phase margin and possible instability. There are several well-documented circuit techniques for dealing with capacitive loads, but the most popular for high speed amplifiers is a resistor in series with the output of the amplifier (as shown below). With the resistor outside the feedback loop, but in series with the load capacitance, the amplifier doesn't directly drive a purely capacitive load. A CF op amp also gives the option of increasing R_F to reduce the loop gain. Regardless of the approach taken, there will always be a penalty in bandwidth, slew rate, and settling time. It's best to experimentally optimize a particular amplifier circuit, depending on the desired characteristics, e.g., fastest rise time, fastest settling to a specified accuracy, minimum overshoot, or passband flatness.



Q. Why don't any of your current feedback amplifiers offer true single-supply operation, allowing signal swings to one or both rails?

A. This is one area where the VF topology is still favored for several reasons. Amplifiers designed to deliver good current drive and to swing close to the rails usually use common-emitter output stages, rather than the usual emitter followers. Common emitters allow the output to swing to the supply rail minus the output transistors' V_{CE} saturation voltage. With a given fabrication process, this type of output stage does not offer as much speed as emitter followers, due in part to the increased circuit complexity and inherently higher output impedance. Because CF op amps are specifically developed for the highest speed and output current, they feature emitter follower output stages.

With higher speed processes, such as ADI's XFCB (extra-fast complementary bipolar), it has been possible to design a common-emitter output stage with 160-MHz bandwidth and 160-V/µs slew rate, powered from a single 5-volt supply (AD8041). The amplifier uses voltage feedback, but even if, somehow, current feedback had been used, speed would still be limited by the output stage. Other XFCB amplifiers, with emitter-follower output stages (VF or CF), are much faster than the AD8041. In addition, single-supply input stages use PNP differential pairs to allow the common-mode input range to extend down to the lower supply rail (usually ground). To design such an input stage for CF is a major challenge, not yet met at this writing.

Nevertheless, CF op amps can be used in single-supply applications. Analog Devices offers many amplifiers that are specified for +5- or even +3-volt operation. What must be kept in mind is that the parts operate well off a single supply *if the application remains within the allowable input and output voltage ranges*. This calls for level shifting or ac coupling and biasing to the proper range, but this is already a requirement in most single-supply systems. If the system must operate to one or both rails, or if the maximum amount of headroom is demanded in ac-coupled applications, a current feedback op amp may simply not be the best choice. Another factor is the rail-to-rail output swing specifications when driving heavy loads. Many so-called rail-to-rail parts don't even come close to the rails when driving back- terminated 50- or 75-Ω cables, because of the increase in V_{CESAT} as output current increases. If you really need true rail-to-rail performance, you don't want or need a current feedback op amp; if you need highest speed and output current, this is where CF op amps excel.

INDEX


ANALOG DEVICES
Analog Dialogue
James Bryant
Chris Hyde
1991
Op-amp issues - Noise (continued) -
How grainy is current flow?
Seminars and support - Supply decoupling
Non-idealities of resistors
1990
James Bryant
Used correctly, high-speed converters provide many useful circuit functions
Op-amp issues
Noise
1992
James Bryant
Walt Kester
1993
PC-board tracks and ground planes
Voltage references
Grounding (Again) -
Time references - More on trimming
Oli Josephson
Using sigma-delta converters (part 1)
Using sigma-delta converters (part 2) and questions on noise in converters
Confused about amplifier distortion specs?
High-frequency signal conditioning -
A Reader notes
Walt Kester
James Bryant