Analog Dialogue

Volume 51, Number 1, 2017

Your Engineering Resource for Innovative Design

Visit *analogdialogue.com*

ANALOG DEVICES
AHEAD OF WHAT'S POSSIBLE™

In This Issue

Jim Surber, Editor

Jim became editor of *Analog Dialogue* in May of 2015, when the preceding editor, Scott Wayne, decided to retire. Jim has been with Analog Devices for 40+ years, starting with a company called Computer Labs, a maker of high speed data converters, which was acquired by ADI in 1978.

Jim has worked in production test, applications engineering, product line marketing, and, most recently, marketing communications. He enjoys writing and has published several technical articles in various magazines down through the years. Residing in Greensboro, North Carolina, Jim enjoys spending time with his family on frequent mountain outings.

You can reach him at *jim.surber@analog.com* or at (336) 605-4365.

AnalogDialogue

Analog Dialogue is a technical magazine created and published by Analog Devices. It provides in-depth design related information on products, applications, technology, software, and system solutions for analog, digital, and mixed-signal processing. Published continuously for 49 years—starting in 1967—it is produced as a monthly online edition and as a printable quarterly journal featuring article collections. For history buffs, the *Analog Dialogue* archive includes all issues, starting with Volume 1, Number 1, and three special anniversary editions. To access articles, the archive, the journal, design resources, and to subscribe, visit the *Analog Dialogue* homepage, *analogdialogue.com*.

High Definition, Low Delay, SDR-Based Video Transmission in UAV Applications

By **Wei Zhou**

Share on

Abstract

Integrated RF agile transceivers are not only widely employed in software-defined radio (SDR)[1] architectures in cellular telephone base stations, such as multiservice distributed access system (MDAS) and small cell, but also for wireless HD video transmission for industrial, commercial, and military applications, such as unmanned aerial vehicles (UAVs). This article will examine a wideband wireless video signal chain implementation using the AD9361/AD9364[2,3] integrated transceiver ICs, the amount of data transmitted, the corresponding RF occupied signal bandwidth, the transmission distance, and the transmitter's power. It will also describe the implementation of the PHY layer of OFDM and present hopping frequency time test results to avoid RF interference. Finally, we will discuss the advantages and disadvantages between Wi-Fi and the RF agile transceiver in wideband wireless applications.

The Signal Chain

Figure 1 illustrates the simplified wireless video transmission scheme using the AD9361/AD9364 and a BBIC. The camera captures the image and transmits video data to a baseband processor via Ethernet, HDMI®, USB, or another interface. Image coding/decoding can be handled by hardware or the FPGA. The RF front includes the switcher, LNA, and PA to the programmable integrated transceiver.

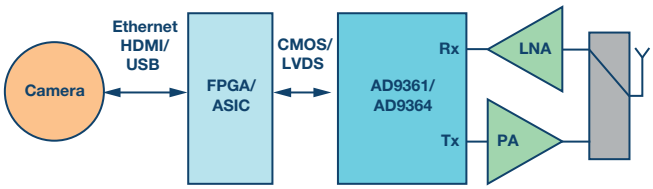


Figure 1. Wireless video transmission diagram.

How Much Data Need To Be Transmitted

Table 1 shows the significant size difference between the uncompressed and compressed data rates. By using high efficiency video coding (HEVC), also known as H.265 and MPEG-H Part 2, we can decrease the data rate and save bandwidth. H.264 is currently one of the most commonly used formats for the recording, compression, and distribution of video content. It presents a huge step forward in video compression technology and is one of several potential successors to the widely used AVC (H.264 or MPEG-4 Part 10).

$$Compressed\ Data\ Rate = \frac{Uncompressed\ Data\ Rate}{Compression\ Ratio}$$

Table 1 summarizes the uncompressed and compressed data rates in different video formats. Assumptions include a video bit depth of 24 bits and a frame rate of 60 fps. In the 1080p example, the data rate is 14.93 Mbps after compression, which then can be easily handled by the baseband processor and the wireless PHY layer.

Table 1. Compressed Data Rate

Format	Horizontal Lines	Vertical Lines	Pixels	Uncompressed Data Rate (Mbps)	Compressed Data Rate (Mbps) Compressed Ratio = 200
VGA	640	480	307,200	442	2.2
720p	1280	720	921,600	1328	6.64
1080p	1920	1080	2,073,600	2986	14.93
2k	2048	1152	2,359,296	3400	17.0
4k	4096	2160	8,847,360	12,740	63.7

Signal Bandwidth

The AD9361/AD9364 support channel bandwidths from <200 kHz to 56 MHz by changing the sample rate, digital filters, and decimation. The AD9361/AD9364 are zero-IF transceivers with I and Q channels to transmit the complex data. The complex data includes real and imaginary parts, corresponding to I and Q respectively, which locate at the same frequency bandwidth to double the spectrum efficiency when compared to a single part. The compressed video data can be mapped to the I and Q channels to create constellation points, which are known as symbols. Figure 2 shows a 16 QAM example where each symbol represents four bits.

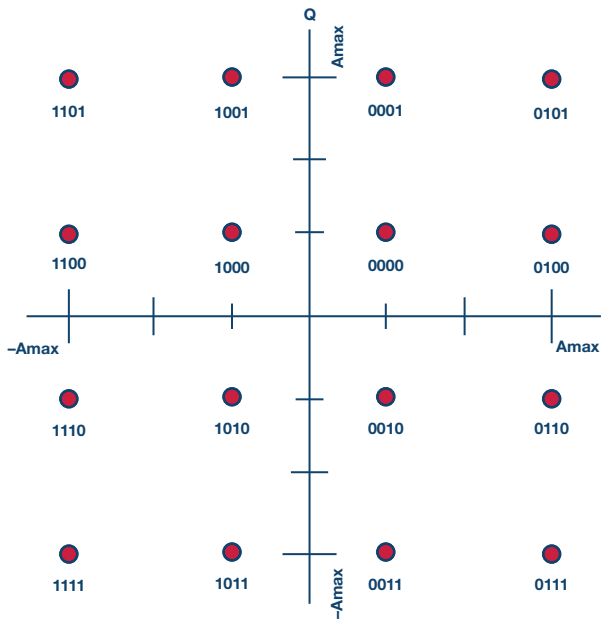


Figure 2. 16 QAM constellation.[4]

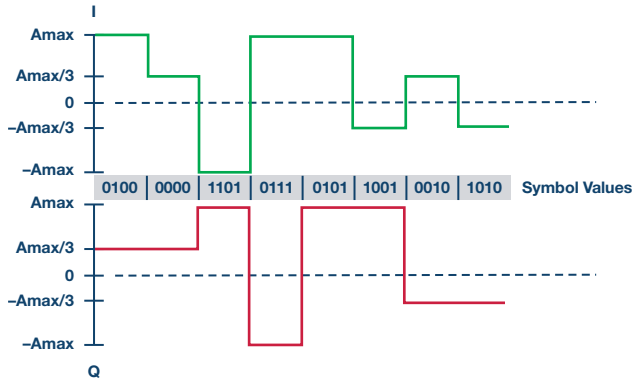


Figure 3. I and Q digital waveform from the constellation.[4]

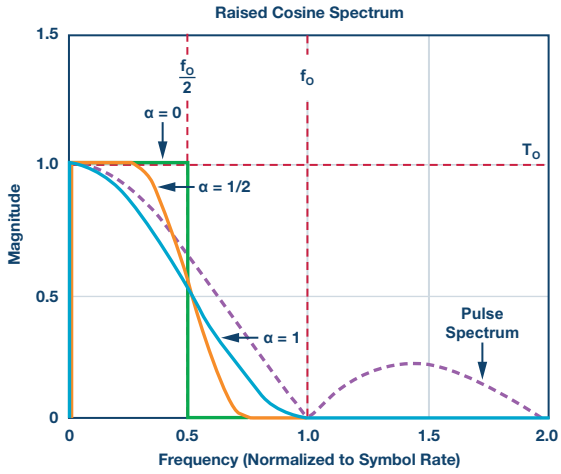


Figure 4. Pulse shaping filter response.[4]

For a single-carrier system, the I and Q digital waveform needs to pass through a pulse-shaping filter before the DAC in order to shape the transmitted signal within a limited bandwidth. An FIR filter can be used for pulse shaping, and the filter response is illustrated in Figure 4. In order to maintain the fidelity of the information, there is a minimum signal bandwidth corresponding to the symbol rate. And the symbol rate is proportional to the compressed video data rate, as shown in the equation below. For the OFDM system, the complex data should be modulated to the subcarriers using the IFFT, which also transmits the signal in a limited bandwidth.

$$Symbol\ Rate = \frac{Bit\ Rate}{The\ Number\ of\ Bits\ Transmitted\ with\ Each\ Symbol}$$

The number of bits transmitted with each symbol depends on the modulation order.

The occupied signal bandwidth is given by,

$$RF\ Occupied\ Signal\ BW = Symbol\ Rate \times (1 + \alpha)$$

In which α is the filter bandwidth parameter.

From the previous formulas we can deduce this equation,

$$RF\ Occupied\ Signal\ BW = \frac{Compressed\ Data\ Rate}{The\ Number\ of\ Bits\ Transmitted\ with\ Each\ Symbol} \times (1 + \alpha)$$

So we can calculate the RF occupied signal bandwidth as summarized in Table 2.

Table 2. Occupied RF Signal Bandwidth with Kinds of Modulation Order (α = 0.25)

Format	Compressed Data Rate (Mbps)	QPSK (Signal BW, MHz)	16 QAM (Signal BW, MHz)	64 QAM (Signal BW, MHz)
VGA	2.2	1.375	0.6875	0.4583
720p	6.6	4.1250	2.0625	1.3750
1080p	14.9	9.3125	4.6563	3.1042
2k	17.0	10.6250	5.3125	3.5417
4k	63.7	39.8125	19.9063	13.2708

The AD9361/AD9364, with up to 56 MHz signal bandwidth, support all the Table 2 video format transmissions and even higher frame rates. Higher order modulation occupies smaller bandwidth and the symbol represents more information/bits, but a higher SNR is needed to demodulate.

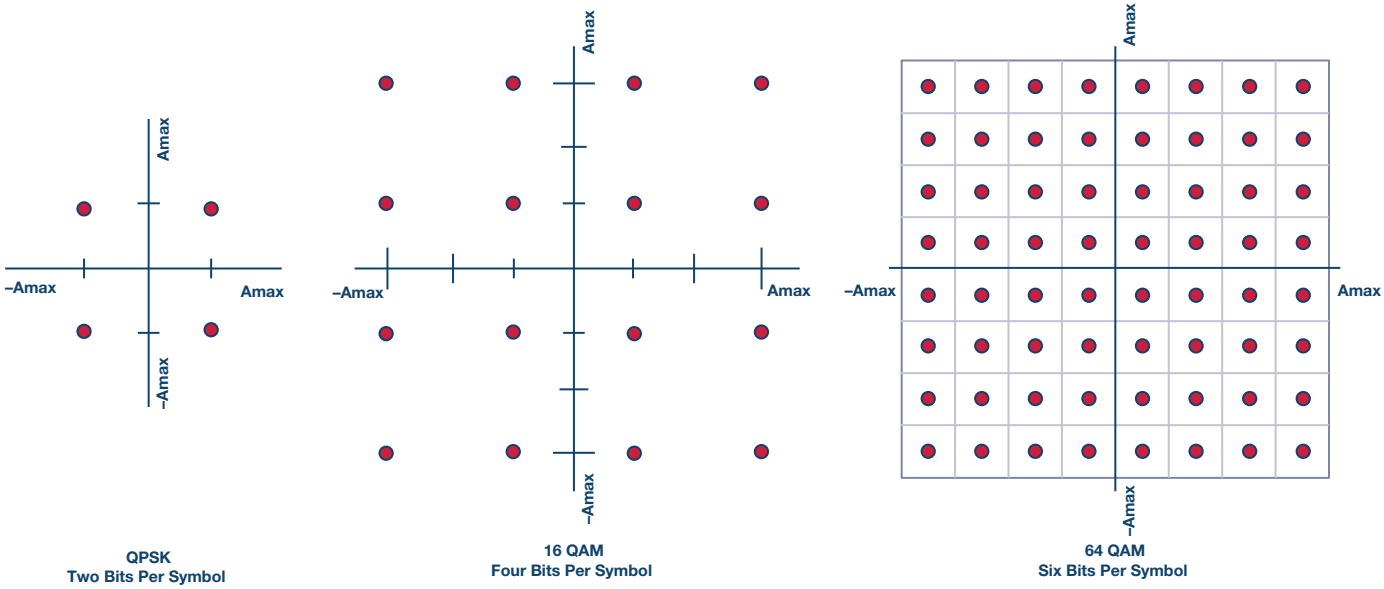


Figure 5. Modulation order.

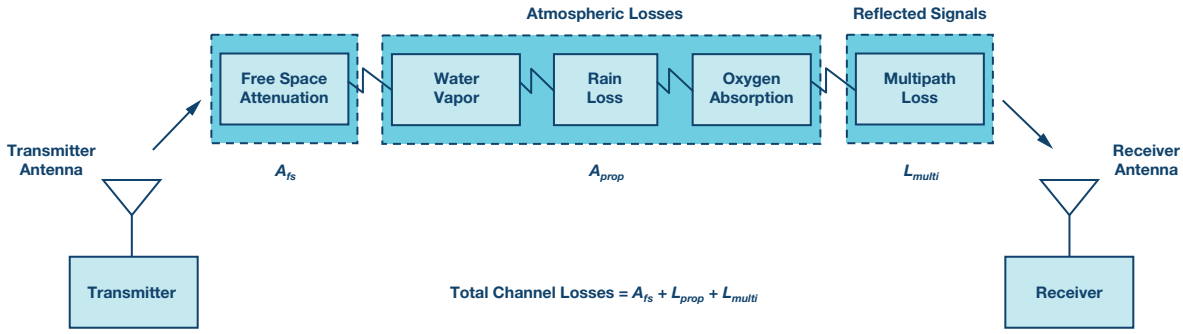


Figure 6. Wireless communication channel loss model.[5]

The Transmission Distance and the Transmitter Power

In applications such as UAVs, the maximum transmission distance is a critical parameter. However, it is equally important that communication not be cutoff even at a limited distance. Oxygen, water, and other obstacles (except for free space attenuation) can attenuate the signal.

Figure 6 shows the wireless communication channel-loss model.

Receiver sensitivity is normally taken as the minimum input signal (S_{min}) required to demodulate or recover the information from the transmitter. After getting the receiver sensitivity, the maximum transmission distance can be calculated with some assumptions, as shown here:

$$S_{min} = 10\log(kT_0B) + NF + \left(\frac{S}{N}\right)_{min} = -174\ \text{dBm} + 10\log B + NF + \left(\frac{S}{N}\right)_{min}$$

$(S/N)_{min}$ is the minimum signal-to-noise ratio needed to process a signal

NF is the noise figure of the receiver

k is Boltzmann's constant = 1.38×10^{-23} joule/k

T_0 is the absolute temperature of the receiver input (Kelvin) = 290 K

B is the receiver bandwidth (Hz)

The parameter $(S/N)_{min}$ depends on the modulation/demodulation order. With the same SNR, lower order modulation gets a lower symbol error, and with the same symbol error, higher order modulation needs higher SNR to demodulate. So when the transmitter is far away from the receiver, the signal is weaker and the SNR is not able to support the high order demodulation. In order to keep the transmitter online and maintain a video format with the same video data rate, the baseband should use lower order modulation at the expense of increasing bandwidth. This helps ensure the received images are not blurred. Fortunately, software-defined radio with digital modulation and demodulation offers the capability to change the modulation. The previous analysis is based on the assumption that the transmitter RF power is constant. While greater RF transmitting power with the same antenna gain will reach a more distant receiver with the same receiver sensitivity, the maximum transmitting power should comply with FCC/CE radiation standards.

In addition, the carrier frequency will have an influence on the transmission distance. As a wave propagates through space, there is a loss due to dispersion. The free space loss is determined by

$$A_{fs} = 20\log\left(\frac{4\pi R}{\lambda}\right) = 20\log\left(\frac{4\pi Rf}{C}\right)$$

In which the R is the distance, λ is the wave length, f is the frequency, and C is the speed of light. Therefore, the larger frequency will have more loss over the same free space distance. For example, the carrier frequency at 5.8 GHz will be attenuated by more than 7.66 dB as compared to 2.4 GHz over the same transmission distance.

RF Frequency and Switching

The AD9361/AD9364 have a programmable frequency range from 70 MHz to 6 GHz. This will satisfy most NLOS frequency applications, including various types of licensed and unlicensed frequencies, such as 1.4 GHz, 2.4 GHz, and 5.8 GHz.

The 2.4 GHz frequency is widely used for Wi-Fi, Bluetooth®, and IoT short-range communication, making it increasingly crowded. Using it for wireless video transmission and control signals increases the chances for signal interference and instability. This creates undesirable and often dangerous situations for UAVs. Using frequency switching to maintain a clean frequency will keep the data and control connection more reliable. When the transmitter senses a crowded frequency, it automatically switches to another band. For example, two UAVs using the frequency and operating in close proximity will interfere with each other's communications. Automatically switching the LO frequency and reselecting the band will help maintain a stable wireless link. Adaptively selecting the carrier frequency or channel during the power-up period is one of the excellent features in high end UAV.

Frequency Hopping

Fast frequency hopping, which is widely used in electronic countermeasures (ECM), also helps avoid interference. Normally if we want to hop the frequency, the PLL needs to relock after the procedure. This includes writing the frequency registers and going through VCO calibration time and PLL lock time so that the interval of the hopping frequency is approximate to hundreds of microseconds. Figure 7 shows an example of hopping transmitter LO frequency from 816.69 MHz to 802.03 MHz. The AD9361 is used in normal frequency changing mode and the transmitter RF output frequency jumps from 814.69 MHz to 800.03 MHz with a 10 MHz reference frequency. The hopping frequency time is tested by using the E5052B as shown in Figure 7. The VCO calibration and PLL lock time is about 500 µs according to the Figure 7b. The signal source analyzer E5052B can be used to capture the PLL transient response. Figure 7a shows the wideband mode of transient measurement, while Figure 7b and 7d provide significantly fine resolution in frequency and phase transient measurement with frequency hopping.[6] Figure 7c shows the output power response.

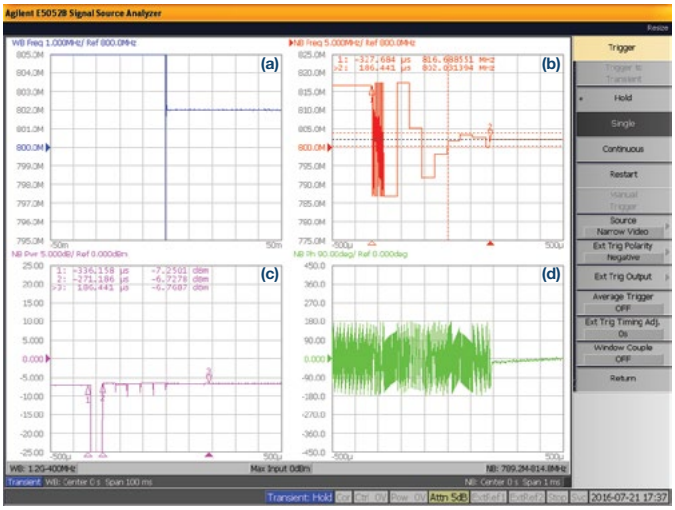


Figure 7. Hopping frequency from 804.5 MHz to 802 MHz with 500 µs.

500 µs is a very long interval for the hopping application. However, the AD9361/AD9364 include a fast lock mode that makes it possible to achieve faster than normal frequency changes by storing sets of synthesizer programming information (called profiles) in the device's registers or the baseband processor's memory space. Figure 8 shows the test result by using the fast lock mode to implement the hopping frequency from 882 MHz to 802 MHz. The time is down to less than 20 µs, according to the Figure 8d phase response. The phase curve is drawn by referring to the phase of 802 MHz. The SPI writing time and the VCO calibration time are both eliminated in that mode due to the frequency information and calibration results being saved in profiles. As we can see, Figure 8b shows the fast frequency hopping capability of the AD9361/AD9364.

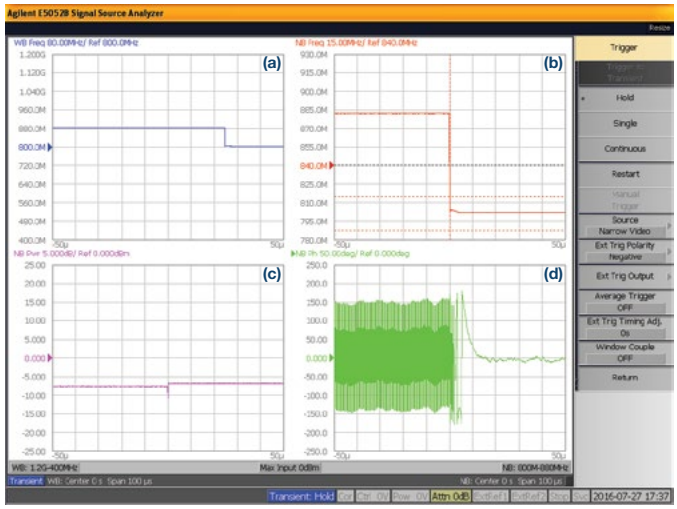


Figure 8. Hopping frequency from 882 MHz to 802 MHz within 20 µs in fast lock mode.

Implementation of the PHY Layer—OFDM

Orthogonal frequency division multiplexing (OFDM) is a form of signal modulation that divides a high data rate modulating stream onto many slowly modulated narrow-band close-spaced subcarriers. This makes it less sensitive to selective frequency fading. The disadvantages are a high peak to average power ratio and sensitivity to carrier offset and drift. The OFDM is widely applied in the wideband wireless communication PHY layer. The critical technology of the OFDM includes IFFT/FFT, frequency synchronization, sampling time synchronization, and symbol/frame synchronization. The IFFT/FFT should be implemented via FPGA in the fastest way possible. It is also very important to select the interval of the subcarriers. The interval should be large enough to withstand the mobility communication with Doppler frequency shift and small enough to carry more symbols within the limited frequency bandwidth to increase the spectrum efficiency. COFDM refers to a combination of encoding technology and OFDM modulation. COFDM with its high resistance of signal attenuation and forward error correcting (FEC) advantages can send video signals from any moving object. The encoding will increase the signal bandwidth but it is usually worth the trade-off.

By combining the model-based design and automatic code generation tools from MathWorks with the powerful Xilinx® Zynq SoCs and Analog Devices integrated RF transceivers, SDR system design, verification, testing, and implementation can be more effective than ever, leading to higher performance radio systems and reducing the time to market.[7]

What Are the Advantages over Wi-Fi?

Drones equipped with Wi-Fi are very easy to connect to a cell phone, laptop, and to other mobile devices, which make them very convenient to use. But for wireless video transmission in UAV applications, the FPGA

plus AD9361 solution offers many advantages over Wi-Fi. First of all, in the PHY layer, agile frequency switching and fast hopping of the AD9361/AD9364 help avoid interference. Most integrated Wi-Fi chips also operate on the crowded 2.4 GHz frequency band with no frequency band reselection mechanism to make the wireless connection more stable.

Secondly, with the FPGA plus AD9361 solution, transmission protocol can be defined and developed flexibly by designers. Wi-Fi protocol is standard and based on a two-way handshake with every packet of data. With Wi-Fi, each data packet has to confirm that a packet was received, and that all 512 bytes in the packet were received intact. If one byte is lost, the whole 512 byte packet must be transmitted again.[8] While this protocol ensures data reliability, it is complex and time consuming to re-establish the wireless data link. The TCP/IP protocol will cause high latency that results in nonreal-time video and control, which can lead to a UAV crash. The SDR solution (FPGA plus AD9361) uses a one-way stream of data, which means the drone in the sky transmits the video signal like a TV broadcast. There is no time for resending packets when real-time video is the goal.

Additionally, Wi-Fi does not offer the proper level of security for many applications. By utilizing the encryption algorithm and user-defined protocol, the FPGA plus AD9361/AD9364 solution is far less susceptible to security threats.

Furthermore, the one way broadcast data stream delivers transmission distance capabilities two to three times that of Wi-Fi approaches.[8] The flexibility from the software-defined radio capability enables digital modulation/demodulation adjustment to satisfy the distance requirements and adjust to changing SNR in complex space radiation environments.

Conclusions

This article illustrated the critical parameters of using an FPGA plus AD9361/AD9364 solution for high definition wireless video transmission implementation. With agile frequency band switching and fast-frequency hopping, it is possible to establish a more stable and reliable wireless link to resist the increasingly complex radiation in space and decrease the probability of a crash. In the protocol layer, the solution is more flexible, using a one-way transmission to reduce wireless establishment time and create a lower latency connection. In industrial and commercial applications such as agriculture, power-line inspection, and surveillance, stable, secure, and reliable transmissions are vital to success.

References

[1] Software-Defined Radio Solutions from Analog Devices. Analog Devices, Inc.

[2] AD9361 data sheet. Analog Devices, Inc.

[3] AD9364 data sheet. Analog Devices, Inc.

[4] Ken Gentile. Application Note AN-922, *Digital Pulse-Shaping Filter Basics*. Analog Devices, Inc.

[5] Scott R. Bullock. *Transceiver and System Design for Digital Communications*, 4th edition. SciTech Publishing, Edison, NJ, 2014.

[6] E5052B Signal Source Analyzer, Advanced Phase Noise and Transient Measurement Techniques. Agilent, 2007.

[7] Di Pu, Andrei Cozma, and Tom Hill. "Four Quick Steps to Production: Using Model-Based Design for Software-Defined Radio." *Analog Dialogue*, Volume 49, 2015.

[8] John Locke. "Comparing the DJI Phantom 4's Lightbridge vs. Yuneec Typhoon H's Wi-Fi." Drone Compares.

Wei Zhou [Wei.Zhou@analog.com] is an applications engineer for Analog Devices, Inc., supporting the design and development of RF transceiver products and applications, especially in the wireless video transmission and wireless communication fields. He has been working in ADI's Central Applications Center located in Beijing, China, for five years supporting various products including DDS, PLL, high speed DAC/ADC, and clocks. Prior to joining ADI, he received his B.S. degree from Wuhan University, Wuhan, China, in 2006, and his M.S. degree from Institute of Electronics, Chinese Academy of Science (CAS), Beijing, China, in 2009. He worked as an RF and microwave circuits and system design engineer for an aerospace technology company from 2009 to 2011.

Wei Zhou

Improving Precision Data Acquisition Signal Chain Density Using SiP Technology

By **Ryan Curran**

Share on

A common desire in the precision data acquisition market space is to improve the density of the signal chain while maintaining performance. As more applications are moving to an ADC-per-channel approach, or trying to fit more channels into the same form factor, channel density is a great concern for many data acquisition signal chain designers. On top of this, there has been a trend to make precision circuitry easier to use and more easily achieve data sheet performance. This provides an opportunity to build subsystems that address these concerns by implementing signal chains using system-in-package (SiP) technology.

The first family of devices produced by Analog Devices as a result of this subsystem strategy is the new ADAQ798x data acquisition products. The ADAQ798x is a 16-bit analog-to-digital converter subsystem that integrates four common signal processing and conditioning blocks into a SiP design that supports a variety of applications. The device also contains the most critical passive components, eliminating many of the design challenges associated with traditional signal chains that utilize successive approximation register (SAR) analog-to-digital converters (ADCs). These passive components are crucial to achieving the specified device performance.

When looking across applications and markets that utilize SAR ADCs, such as industrial, instrumentation, communications, and healthcare, it was seen that certain sections of data acquisition signal chains are very common across these applications, while some portions of the signal chains can differ extensively. It was seen that these signal chains use a varying array of input sources and sensors. As a result, there are going to be various forms of signal conditioning implemented before presenting the signal to the ADC. With the sources varying, this means the system full scales can be different and require different reference values to maximize dynamic range. Some applications are multichannel and, therefore, implement a front-end multiplexer. Differing supply schemes would be implemented based upon the key performance criteria of the application. However, regardless of the application, there are components that are common to many of these applications. The ADAQ7980 and ADAQ7988 are part of an all Analog Devices active component solution that contains a high accuracy, low power, 16-bit SAR ADC; a low power, high bandwidth, high input impedance ADC driver; a low power, stable reference buffer; and an efficient power management block. These signal chain components have been integrated into a data acquisition subsystem using SiP technology.

Housed within a small footprint, 5 mm × 4 mm LGA package, this new style component will simplify the design process for data acquisition systems. The ADAQ798x's level of system integration solves many design challenges, yet the device still provides the flexibility of a configurable ADC driver block to allow for gain and/or common-mode adjustments. A set of four device supplies provides optimal system performance, but single supply operation is possible with minimal impact on the device's operating specifications. The ADAQ798x family provides a significant level of integration while still being flexible enough to adapt to a wide assortment of applications.

When developing this product, ADI analyzed common design mistakes to determine how to help solve these challenges. It was seen that many of these signal chain level design mistakes were primarily centered around two areas of the SAR ADC—the reference input and the analog input. Many of these signal chain errors are associated with the circuitry peripheral to the ADC that is critical to the overall analog-to-digital conversion performance. With respect to the reference, common mistakes include improper layout and sizing of the reference bypass capacitor, insufficient drive strength of the reference source, and too large of a noise spectral density generated by the reference source. These unsuitable design conditions at the reference input of a SAR ADC can lead to the ADC making incorrect bit decisions. As for the ADC analog input, common design issues that have been observed include incorrect ADC driver selection, incorrect filter bandwidth between the ADC driver and the ADC, and incorrect filter capacitor dielectric material choices. Any combination of these system-level design issues can lead to severe degradation of ADC conversion performance. The choices made in developing the ADAQ798x devices were intended to address these concerns.

As just discussed, to achieve data sheet performance from a SAR ADC-based conversion system, there are some design considerations that must be taken into account. The SAR ADC reference source and analog input source characteristics are critical to successfully designing the conversion signal chain. Typically, a SAR ADC requires a low impedance reference source and a large, properly laid out decoupling capacitor. This bypass capacitor is used to replenish charge consumed by the ADC during the SAR bit trials and can be thought of as a component of the SAR array that is external to the ADC. The ADC also needs an analog input source with sufficient noise performance and bandwidth to properly settle the ADC input to the desired resolution. Figure 1 shows the block diagram of the ADAQ798x devices.

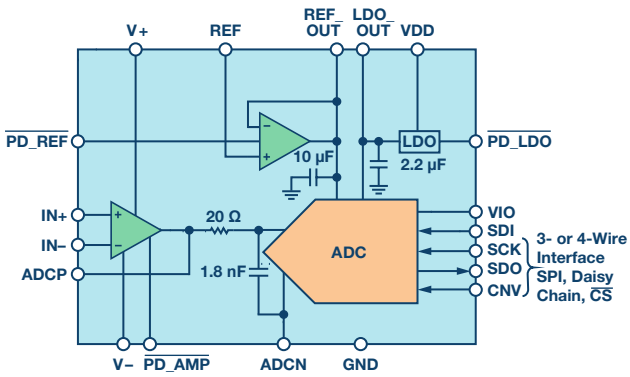


Figure 1. ADAQ798x block diagram.

As seen in Figure 1, ADAQ798x houses a reference buffer and a corresponding 10 µF decoupling capacitor. The decoupling capacitor is ideally laid out in proximity to the reference input of the ADC. The goal of this component placement is to reduce all parasitic impedances between the decoupling capacitor and the SAR capacitor array. This path should be as low of an impedance as possible to allow the capacitor to quickly add charge onto the SAR array to be redistributed as part of the conversion process. As well, the trace resistance between the reference buffer and the decoupling capacitor has been controlled. A trace dimension was chosen to provide a resistance that will keep the reference buffer stable, while not causing a voltage drop large enough to create a conversion gain error. The amplifier used to buffer the reference signal is configured in unity gain. This presents a high impedance input to the external reference source instead of presenting the traditional switched capacitor load of the SAR ADC reference input. The ADAQ798x user can now implement a lower power or unbuffered reference to drive the ADAQ798x reference input (REFIN) pin. By presenting a high impedance, this also gives the user more flexibility in choosing the physical PCB location of the reference source. By using this SiP component, the reference source layout has become much less critical because of the inclusion of a well controlled reference buffer within the ADAQ798x. By only including a buffer and not the reference source itself, the user has the freedom to choose a wide range of reference values and ultimately maximize the system dynamic range with this reference selection because the reference sets the converter full-scale voltage.

The ADAQ798x also features an ADC driver and corresponding low-pass filter between the driver and the ADC input. The filter bandwidth selection is critical to achieving desired performance levels. The bandwidth is selected as a trade-off between settling time and filtering wideband noise from the high speed ADC driver. Any disturbances at the ADC input node must be settled to a sufficient resolution within the acquisition time of the ADC. When a SAR ADC is performing its conversion process, the ADC input is disconnected from its external input sources. During the conversion, the voltage potential at the ADC input could change. However, at the end of conversion, the voltage on the SAR capacitor array is essentially the same as it was when the conversion started. When the ADC returns to acquisition (track) mode, the SAR capacitor array load is now present at the ADC input. This capacitance is placed in parallel with the capacitor from the external low-pass filter. With differing voltages on these capacitors, a charge redistribution will occur to balance out the voltage on all of these capacitors. This will result in a voltage step at the ADC input that needs to be settled during the acquisition period. The worst-case step occurs when a full-scale transition is presented to the ADC. This scenario can arise in systems with a multiplexed input. This voltage step is attenuated by the ratio of the external capacitor and the internal SAR capacitance. The ADAQ798x products feature a low-pass filter capacitor of 1800 pF. Assuming a 5 V reference voltage, the maximum ADC input voltage step is calculated as follows:

$$V_{STEP} = \frac{5\text{ V} \times C_{SAR}}{C_{EXT} + C_{SAR}} = \frac{5\text{ V} \times 27\text{ pF}}{1800\text{ pF} + 27\text{ pF}} = 73.9\text{ mV}$$

This voltage step must be settled within the minimum acquisition time of 290 ns. The number of time constants required to settle this voltage step can be calculated as the natural logarithm of the ratio of the step size to the settling error. The settling error is chosen to be ½ LSB. Therefore, the number of time constants is found by:

$$\#\ of\ Time\ Constants = \ln\left(\frac{V_{STEP}}{V_{half\_LSB}}\right) = \ln\left(\frac{73.9\text{ mV}}{\frac{5\text{ V}}{2^{16+1}}}\right) = 7.57$$

With the number of time constants known, then the tau (τ) of the RC low-pass filter can be determined:

$$\tau = \frac{Minimum\ Acquisition\ Time}{\#\ of\ Time\ Constants} = \frac{290\text{ ns}}{7.57} = 38.3\text{ ns}$$

This tau can be used to determine the required filter bandwidth with the following equation:

$$RC\ Bandwidth = \frac{1}{2 \times \pi \times \tau} = \frac{1}{2 \times \pi \times 38.3\text{ ns}} = 4.15\text{ MHz}$$

To provide some margin and utilize standard value components, the ADAQ798x products feature a filter composed of a 20 Ω resistor and 1800 pF capacitor. This provides a filter bandwidth of 4.42 MHz, thus allowing the ADAQ798x filter to settle the largest anticipated voltage steps within the acquisition time of the ADC. The filter bandwidth calculated also represents the trade-off point between noise filtering and settling. Using a filter bandwidth near the minimum needed to ensure settling will maximize the noise filtering benefit of the passive low-pass filter.

While the voltage step from the SAR ADC returning to acquisition mode is the limiting factor in the filter settling calculation, it should be noted that the filter can also settle the actual voltage change from a multiplexer full-scale step well within the minimum conversion period of 1 µs. To settle a full-scale step to a ½ LSB resolution requires 11.78 time constants. This is calculated from the natural logarithm of N+1 number of quantization levels. In this case, 2^{17} or 131072 codes. 11.78 time constants at 38.3 ns per time constant is approximately 450 ns, which is of no concern as compared to the 1 µs conversion period. This assumes the multiplexer channel is switched directly after a conversion is initiated.

The ADC driver bandwidth is also extremely important to ensure proper performance of the conversion signal chain. In unity gain, the limiting factor in settling is the voltage step that needs to be settled in 290 ns that is associated with the converter returning to acquisition mode. Therefore, in this case, the small signal bandwidth is the most important amplifier bandwidth specification. To settle a multiplexer full-scale step within the minimum conversion period of 1 µs, an ADC driver large signal bandwidth must be maintained that allows for 11.78 time constants in the 1 µs time period.

The ADC driver should not contribute excessive noise to the conversion signal chain. The total subsystem noise performance is calculated as the root-sum-square combination of the ADC noise, the ADC driver noise, and the reference buffer noise. Due to the limited bandwidth of the reference circuit as a result of the large bypass capacitor, the reference buffer noise is negligible in the RSS calculation. A target for the ADC driver noise in a unity-gain configuration has been chosen to be no greater than ⅓ of the ADC noise. The ADC driver is specified to have a noise spectral density of 5.2 nV/√Hz. To calculate the overall system noise, the ADC driver's noise spectral density must be converted to µV rms with the following equation:

$$v_{n,rms} = \underset{Gain}{Noise} \times e_{n,rms} \times \sqrt{\frac{\pi}{2} \times \underset{Bandwidth}{RC\ Filter}} = (1) \times \frac{5.2\text{nV}}{\sqrt{\text{Hz}}} \times \sqrt{\frac{\pi}{2} \times 4.42\text{ MHz}}$$

$$v_{n,rms} = 13.7\ \mu\text{V rms}$$

The ADC features a typical dynamic range specification of 92 dB with a 5 V reference. The noise floor of the ADC can be calculated as follows:

$$ADC\ Noise\ Floor = V_{full\text{-}scale,rms} \times 10^{-DR/20} = \frac{5}{2\sqrt{2}} \times 10^{-92/20} = 44.4\ \mu\text{V rms}$$

With an ADC driver noise floor of 13.7 µV rms, this falls below the target of ⅓ of the ADC noise. The overall system dynamic range is reduced from 92 db to 91.6 dB due to the noise contribution of the ADC driver in a unity-gain configuration. Because of the ADC driver's limited impact on system noise, there is no need to change the low-pass filter bandwidth for lower sample rate applications that provide more settling time due to longer acquisition periods. The best improvement in unity gain one could hope for by reducing the filter bandwidth would be gaining back the 0.4 dB of dynamic range loss. However, increasing the filter resistance to reduce bandwidth can have detrimental effects on THD performance, while the ADC driver may have difficulty driving larger capacitive loads. If additional filtering is required, the ADC driver can be configured to provide a filtering benefit.

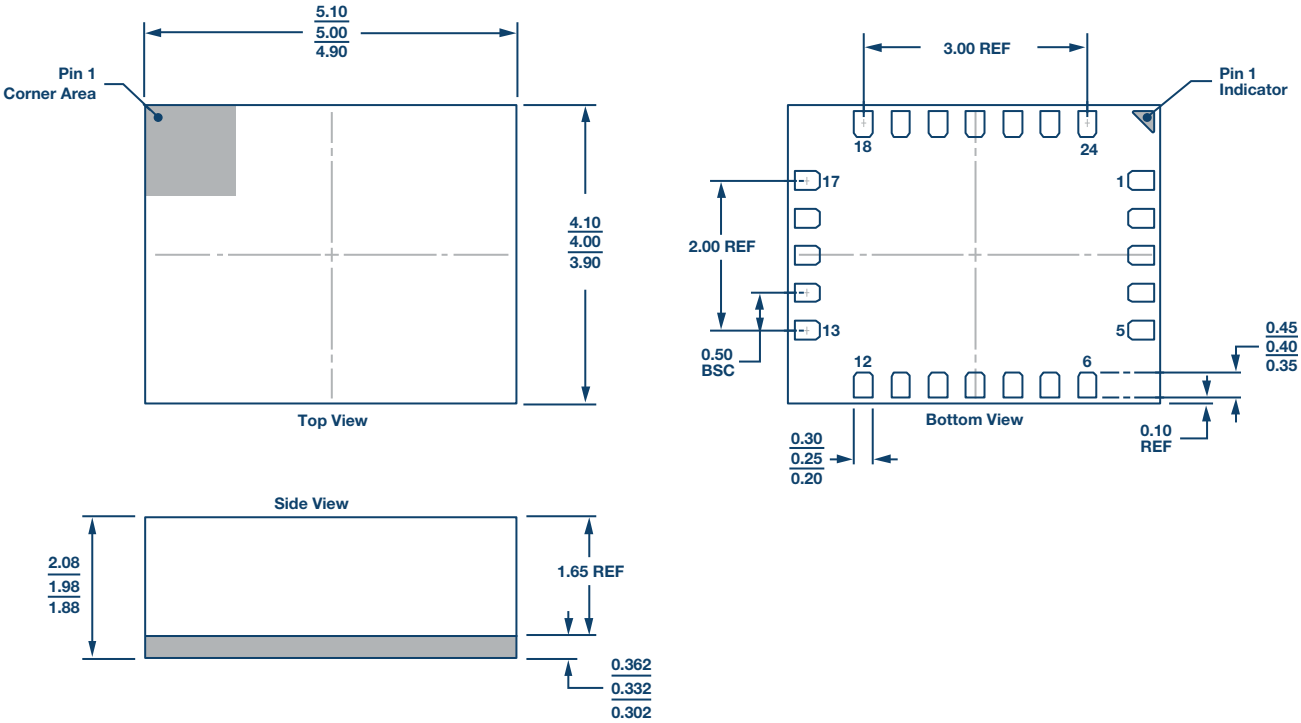


Figure 2. ADAQ798x package outline drawing.

Also included within the ADAQ798x products is a low noise, 2.5 V CMOS LDO linear regulator. Some SAR ADC products require a specific supply of 2.5 V with a small tolerance range. For users that do not have a 2.5 V supply rail available, one would have to be generated specifically for the ADC. With this component, the supplies have been greatly simplified as a result of the inclusion of the LDO. The on-board LDO is used to supply the converter and the LDO input now acts as the ADC supply. This provides a much wider range of usable supply voltages. It also provides a level of simplicity. The positive amplifier supply can be used as the LDO input to create a single-supply system. Additionally, the supply voltage choices can be made to optimize for performance or power consumption. The device features full power-down capability. The flexibility of the supply configurations allows the ADAQ798x user to make the trade-offs most appropriate for their application.

The ADAQ798x package dimensions are 5 mm × 4 mm × 2 mm. The four layer laminate is 0.35 mm thick, while the mold cap is 1.65 mm thick. This over-mold encapsulation features full mold compound and underfill just like any typical encapsulated integrated circuit. The laminate presents an LGA footprint to the user and features 24 I/O pads. Figure 2 shows the package outline drawing of the ADAQ798x. Figure 3 is a model of the ADAQ798x assembly without any encapsulation or mold compound. Figure 3 displays that the subsystem is a mix of Analog Devices active components and commonly available open market passive components. The laminate traces have been designed to control impedances and eliminate any crosstalk effects. The culmination of all these design and assembly techniques has led to the development of a product that can save up to 50% PCB area when compared to similar individual component designs.

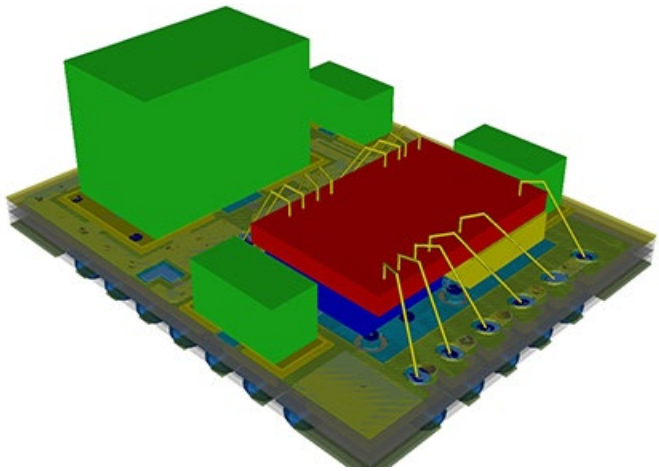
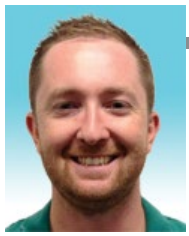


Figure 3. ADAQ798x 3D assembly model.

On top of the area savings, the ADAQ798x gives signal chain designers a better chance of achieving desired performance and reduces the risk of system redesign. Ultimately, this should lead to a shorter time to market and reduced development cost. The system bill of materials is also simplified by choosing the ADAQ798x and now more of the system is covered by one data sheet. This SiP component is robust. It is designed and extensively qualified to withstand harsh industrial environments. It delivers excellent quality ratings and is specified over a –55°C to +125°C temperature range. Overall, the ADAQ798x delivers an exceptional balance of integration vs. flexibility without compromising signal chain performance.

Ryan Curran [ryan.curran@analog.com] is a product applications engineer within the Precision Converters Business Unit at Analog Devices. He joined ADI in 2005 and has focused on SAR ADCs since joining the company. Ryan received his B.S.E.E. from the University of Maine in Orono, ME, and is currently pursuing an M.B.A. from the Isenberg School of Management at UMass Amherst.



Ryan Curran

Also by this Author:

RF-to-Bits Solution Offers Precise Phase and Magnitude Data for Material Analysis

Volume 48, Number 4

Rarely Asked Questions—Issue 137 Psychotic ADC?

By **David Buchanan**

Share on

Question:

I was initially testing one of your ADCs successfully, and suddenly I started getting some rather bizarre FFT results. What's going on?

Answer:

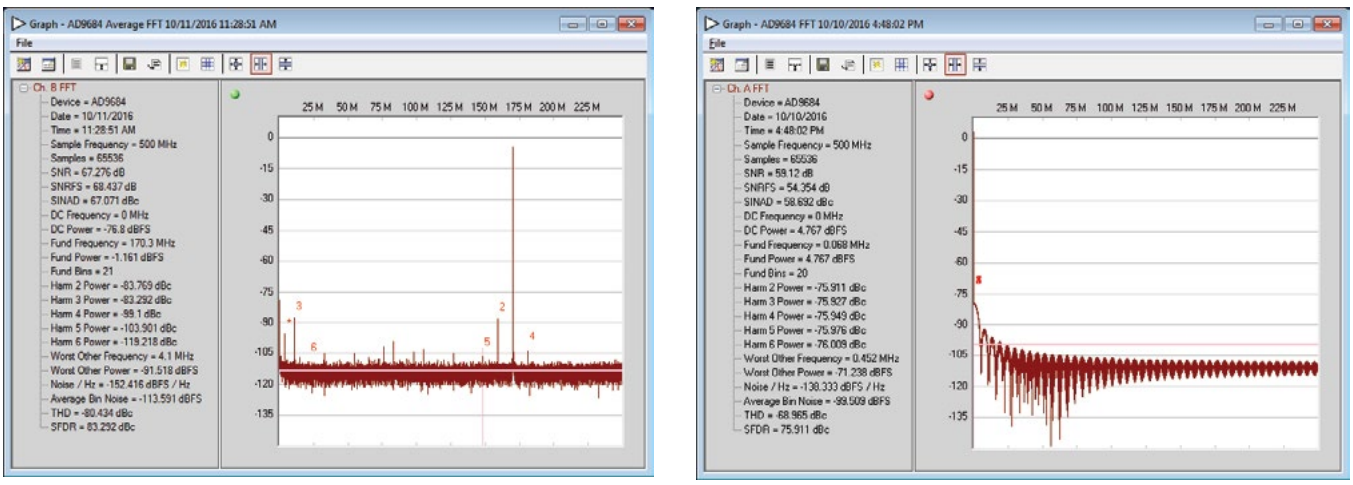
I received this inquiry recently and I was able to resolve it rather quickly. The designer's problem is illustrated in the FFT results in Figure 1.

What the customer reported was that not only did the FFT results seem crazy, they were also inconsistent. This behavior was also consistent with my initial guess at what was happening: the clock source was turned off or not connected, and the converter's input sample clock receiver was oscillating on its own. This can also happen if the cable connecting the clock is intermittent or a component in the signal path is flaky. As I said, this one didn't take too long to solve because I've seen similar results many times. Figure 2 has a flavor of some other FFTs you might see in this operating condition.

In almost all applications, you want the sample clock input to be a single frequency. Any variation due to phase or thermal noise, frequency instability, or unwanted frequency content will cause the expected relationships between the sample clock and analog input signal to break down when you look in the frequency domain. See application note AN-756 for some common examples of how subtle phase noise or modulation on the clock can distort the input signal as it is sampled.

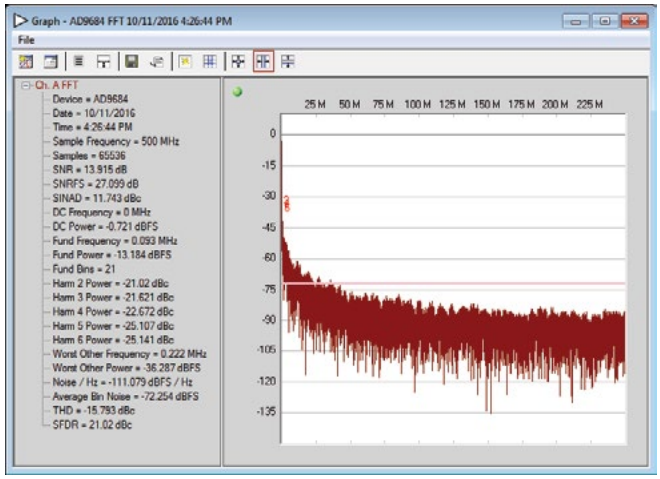
What is the culprit in this case? The sample clock inputs for high speed ADCs are typically differential inputs that share the same common-mode bias, and the receiver has a very high gain. So, with no differential signal applied, the inputs are biased at the same voltage and any noise that is not common mode can cause the sample clock receiver to oscillate. In this condition, the oscillation will not be a pure frequency (if it were, it might be a nice feature). Instead, the frequency will vary randomly. With the sample clock frequency varying randomly, the analog input's energy will be spread across the Nyquist bandwidth in the frequency domain.

In most cases you will just want to recognize this and restore the intended clock reference to continue with your testing. But if you want to verify that this is the issue, observe the ADC's data clock out (DCO) (note—this is not

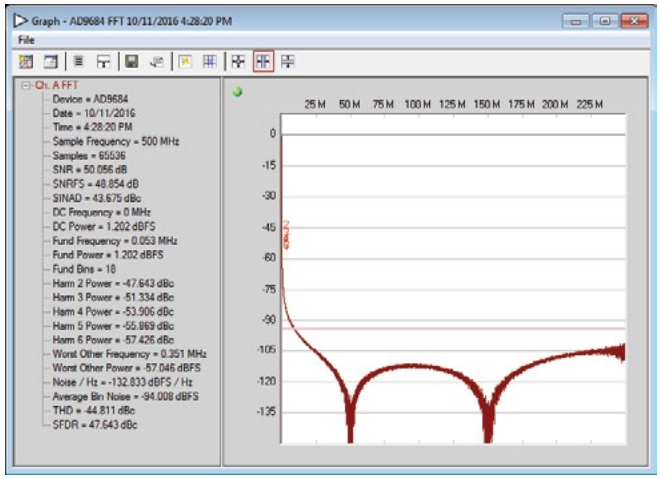


(a) (b)

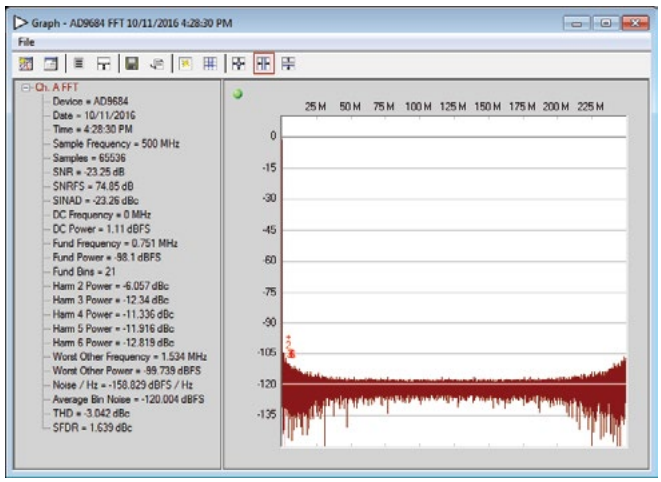
Figure 1. Good and bad FFT results for the AD9684 ADC sampling at 500 MSPS, 170.3 MHz A_{IN} at –1 dBFS.



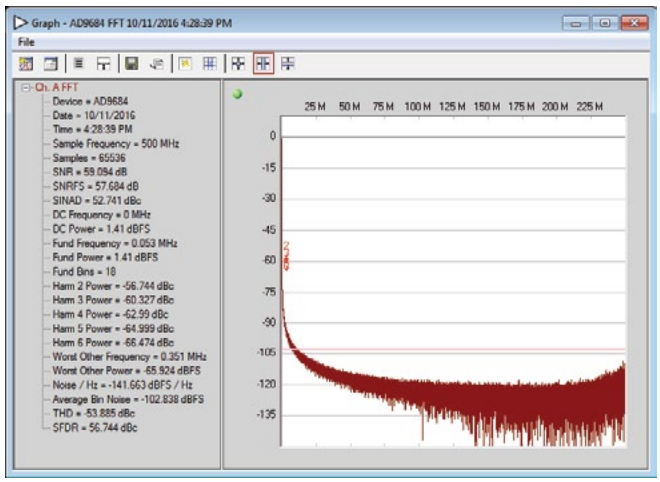
(a)



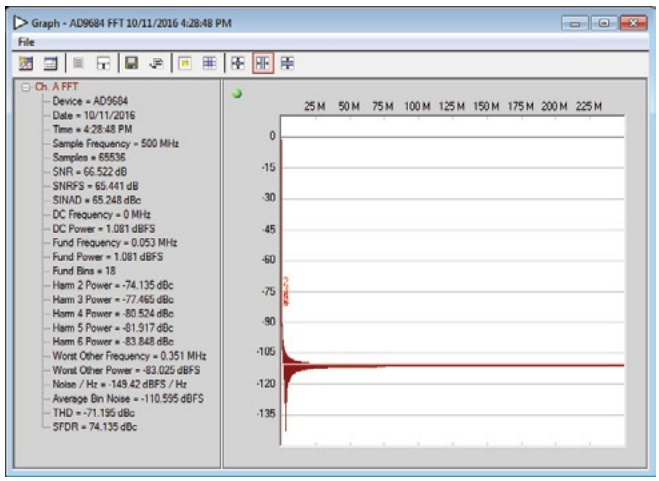
(b)



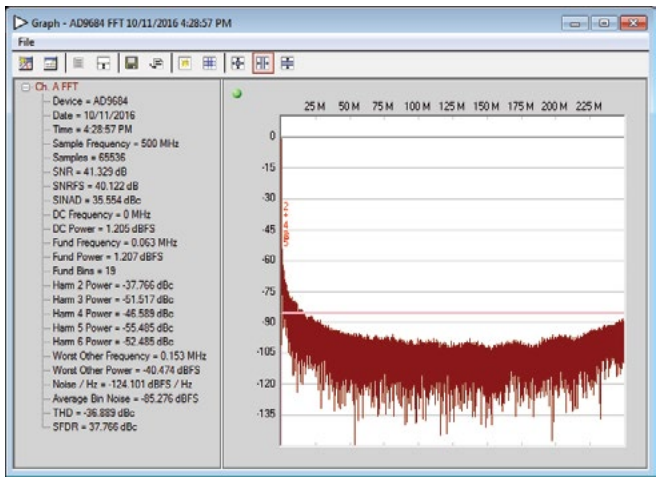
(c)



(d)



(e)

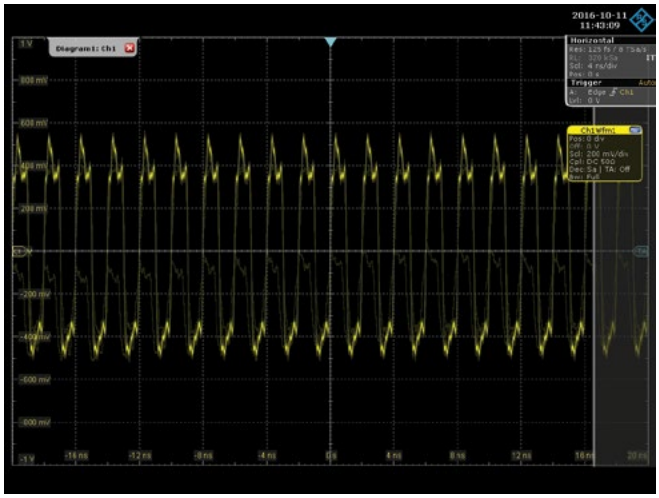


(f)

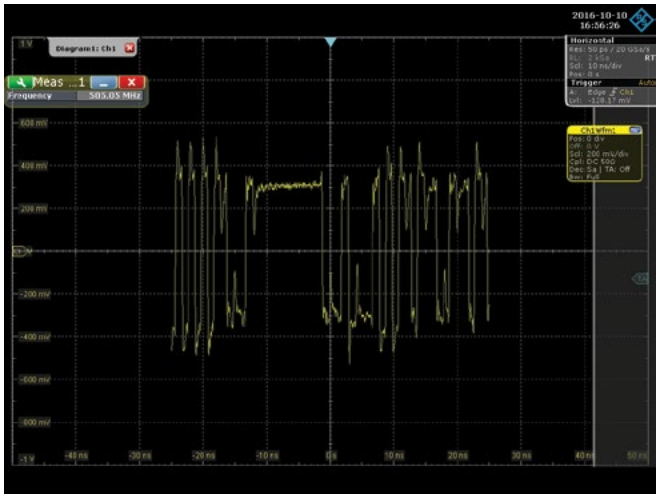
Figure 2. Sample FFT results for unstable clock oscillation.

applicable to JESD204B outputs). This is usually a delayed replica of the ADC's sample clock, or a divided down version of the sample clock, if you are employing any digital features that decimate the data rate. For the good and bad FFTs in Figure 1, the data clock outputs are shown in Figure 3.

As you can see, the period is varying as we expected. I certainly understand why you wouldn't recognize this on the first encounter (or even first few). On face value it looks like the test bed is functioning, but the results are suddenly confusing. Was the ADC damaged? Is the data capture confused? Is the software corrupted? No, there is just a missing signal source.



Output Clock for Good FFT Result



Output Clock for Bad FFT Results

Figure 3. ADC data clock outputs for the two FFT cases shown in Figure 1.

David Buchanan [david.buchanan@analog.com] received a B.S.E.E. from the University of Virginia in 1987. Employed in marketing and applications engineering roles by Analog Devices, Adaptec, and STMicroelectronics, he has experience with a variety of high performance analog semiconductor products. He is currently a senior applications engineer with ADI's high speed converters product line in Greensboro, North Carolina.

David Buchanan

Complex RF Mixers, Zero-IF Architecture, and Advanced Algorithms: The Black Magic in Next-Generation SDR Transceivers

By **Frank Kearney** and **Dave Frizelle**

Introduction

There is an interesting interaction between complex mixers, zero-IF architecture, and advanced algorithm development. The objective of this article is to establish the basic fundamentals of each: the principles of operation and the value they deliver in terms of system design, and then to discuss the interdependability of the three.

RF engineering is often regarded as the black art of electronics. It can be a strange mix of mathematics, mechanics, and, in some instances, just trial and error. It unsettles many a good engineer and many others settle for understanding the outcome rather than the detail. Much of the existing literature jumps straight into the theoretical and mathematical explanation without establishing the underlying concepts.

Demystifying the Complex RF Mixer

Figure 1 provides an overview of the complex mixer in an upconverter (transmitter) configuration. Two parallel paths with independent mixers are fed from a common local oscillator whose phase is offset 90° to one of the mixers. The independent outputs are then summed in a summing amplifier to produce the desired RF output.

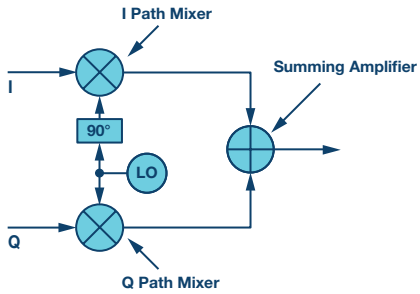


Figure 1. Basic architecture of a complex transmitter.

The configuration has a very useful application. Let's assume, as shown in Figure 2, that we feed a tone signal only on the I input, and the Q input is undriven. Given that the tone at the I input has a frequency of x MHz, the mixer in the I path produces an output at the LO frequency ±x. As there is no signal applied to the Q input, the mixer in its path produces an empty spectrum, and the output from the I mixer passes straight to the RF output.

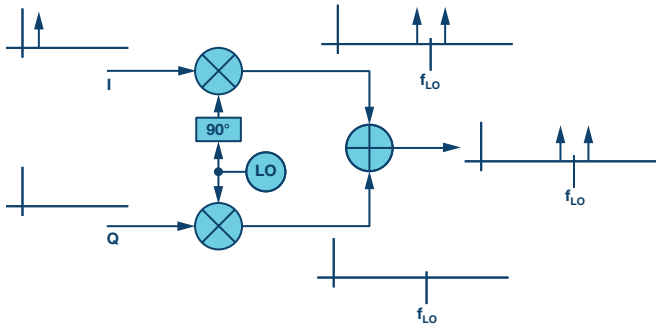


Figure 2. I path analysis.

Alternatively, let's assume that a signal tone at frequency x is applied solely to the Q input. The Q mixer in turn produces an output with tones at the LO frequency ±x. With nothing applied to the I input, its mixer output is muted and the output from the Q mixer goes straight to the RF output.

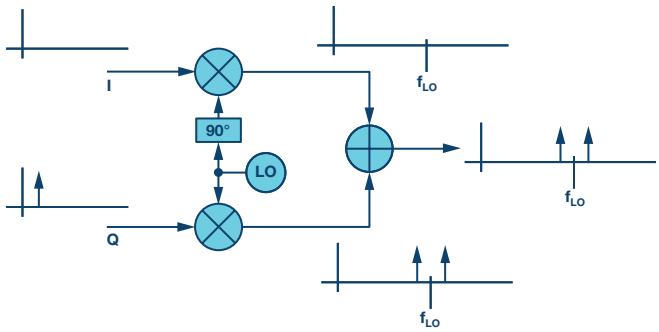


Figure 3. Q path analysis.

At first glance it may seem that the outputs from Figure 2 and Figure 3 are identical. However, there is one critical difference, namely phase. Let's assume, as shown in Figure 4, that we apply the same tone to both I and Q inputs, and that there is a 90° phase shift between the input channels.

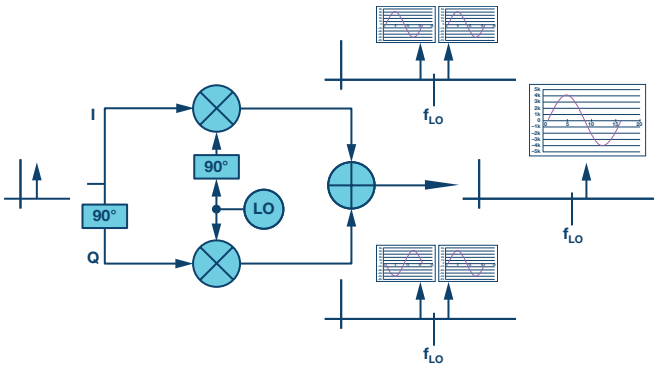


Figure 4. Simultaneous I and Q signal path analysis.

If we look closely at the output of the mixers, we observe that signals at the LO frequency plus the input frequency are in phase, whereas signals produced at the LO frequency minus the input frequency are out of phase. This results in the tones on the upper side of LO adding while the tones on the lower side cancel. Without any filtering we have removed one of the tones (or sidebands) and created an output that sits entirely on one side of the LO frequency.

The example shown in Figure 4 has the I signal leading the Q signal by 90°. If the configuration was to change such that the Q signal led the I signal by 90°, then we could expect a similar summing and cancellation, but in this instance all the signal would appear on the lower side of the LO.

Figure 5 shows the results of lab measurements of a complex transmitter. The left hand side shows the test case when I leads Q by 90°, resulting in the output tone placement on the upper side of the LO. The right hand side of Figure 5 shows the relationship swapped so that Q now leads I by 90° and the resultant output tone sits on the lower side of the LO.

In theory it should be possible to have all the energy on only one side of the LO. However, as the result from the lab experiments in Figure 5 shows, in practice full cancellation may not occur, leaving some energy on the other side of LO, known as the image. Also note that energy at the LO frequency is present, known as LO leakage or LOL. Other energy is also evident in the results—these are harmonics of the wanted signal and are not discussed in this article.

For perfect image cancellation, the outputs of the I and Q mixers must be of precisely the same amplitude, and be exactly 180° out of phase with respect to each other on the image side of the LO. If the phase and amplitude requirements are not met, then the summing/cancellation process, as shown in Figure 4, becomes less than perfect and energy at the image frequency will remain.

Implications

The use of a conventional, single-mixer architecture produces LO± products. Before transmission, one of the sidebands will need to be removed, usually through the addition of a band-pass filter. The filter roll off must be such that it removes the unwanted image signal without affecting the wanted signal.

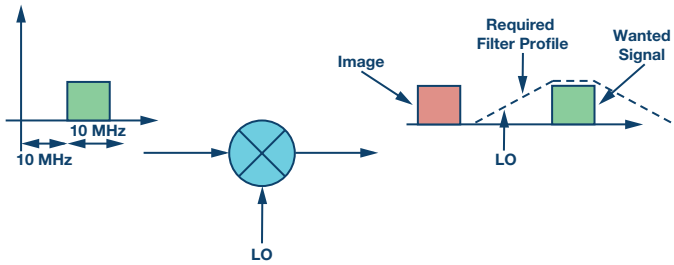


Figure 6. Single-mixer image filter requirements.

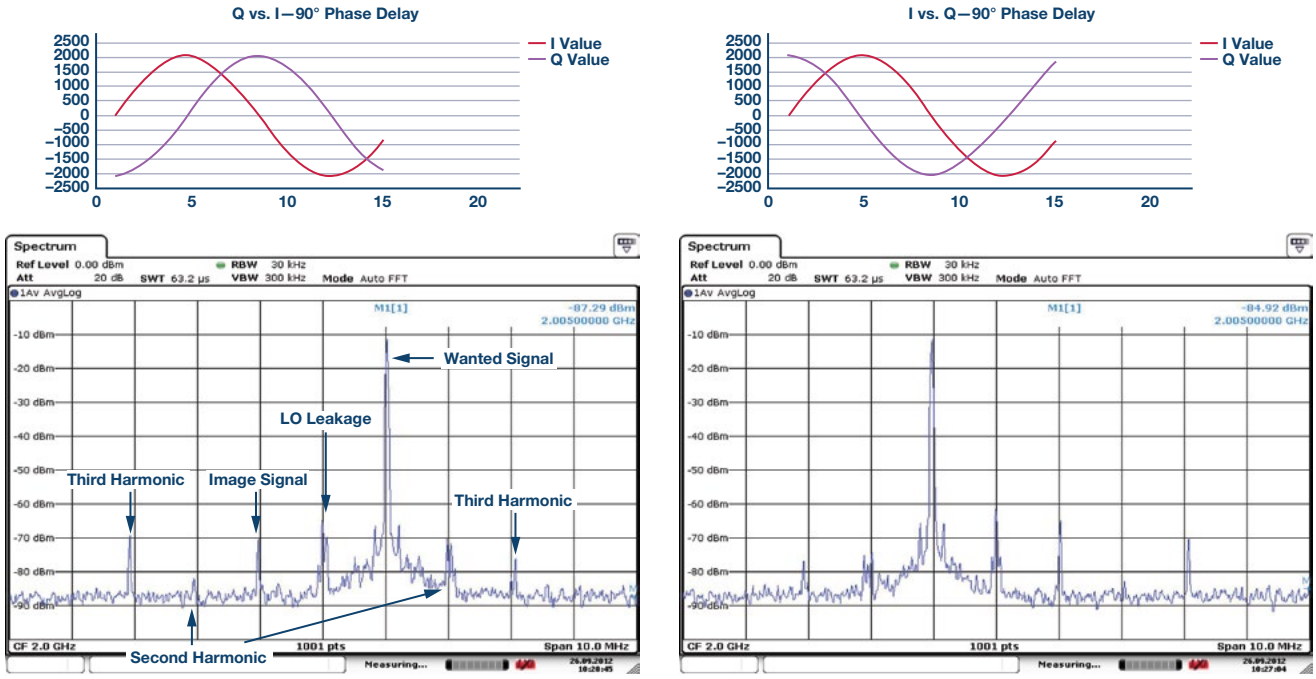


Figure 5. Tone placement dependent of the I and Q phase relationship.

The spacing between the image and the wanted signal directly affects the filter requirements. Where the spacing is large, a simple low cost filter with a gentle roll off can be used. If the spacing is narrow, then designs must implement a filter with a sharp response; typically employing multipole or SAW filters. Hence it would be correct to state that spacing must be maintained between the image and the wanted signal so that the image can be filtered without affecting the wanted signal, and that the spacing is inversely proportional to the complexity and cost of the filter. Furthermore, the filter must be tunable in frequency if the LO frequency is variable, which further increases the complexity of the filter.

The spacing between the image and the wanted signal will be determined by the signal that we apply to the mixer. The example in Figure 6 shows a 10 MHz bandwidth signal shifted 10 MHz away from dc. The resultant output from the mixer places the image 20 MHz from the wanted signal. In this configuration, to achieve a 10 MHz wanted signal spectrum at the output, we had to have a 20 MHz baseband signal path to the mixer. 10 MHz of the baseband bandwidth is unused, and the data interface rate to the mixer circuit is higher than necessary.

Returning to the complex mixer as shown in Figure 5, we know that its architecture eliminates the image without the need for external filtering. What's more, in a zero-IF architecture we can optimize efficiency so that the signal path processing bandwidth is equal to that of the wanted signal. Figure 7 shows a conceptual diagram of how this is achieved. As previously shown, if I leads Q by 90°, there will be an output on the upper side of LO only. If Q leads I by 90°, there will be an output on the lower side of LO only. Therefore, if two independent baseband signals are generated, where one is designed to produce an upper sideband output only and the other is designed to produce a lower sideband output only, they can be summed in baseband and applied to the complex transmitter. The result will be an output with different signals appearing above and below LO. In a practical application the combined baseband signal would be produced digitally. The summing nodes shown in Figure 7 are solely to illustrate the concept.

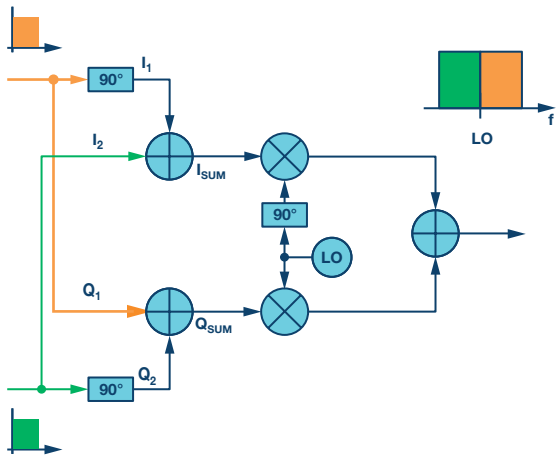


Figure 7. Zero-IF complex mixer architecture.

The Zero-IF Dividend

The use of the complex transmitter to generate a single sideband output provides substantial advantages in terms of the RF filtering required to remove the image. However, if the image cancellation performance is good enough to make the image negligible, we can exploit the architecture more by using it in zero-IF mode. Zero-IF allows us to take specially created baseband data and produce an RF output with independent signals appearing on either side of the LO. Figure 8 provides an illustration of how this might be done. We have two sets of I and Q data, where each is independent and encoded with symbol data that can be decoded at the receiver with respect to the phase of the reference carrier.

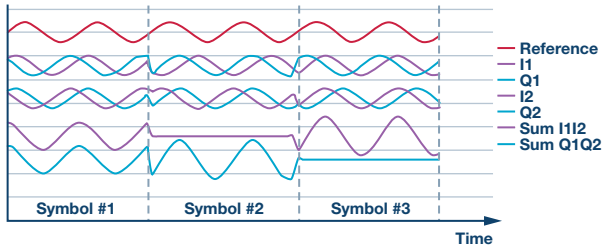


Figure 8. Taking a closer look at I/Q signaling in a zero-IF complex mixer configuration.

Initial observation shows that Q1 leads I1 by 90° and that the amplitude of both are matched. Likewise, I2 leads Q2 by 90° and their amplitudes are also matched. The independent signals are combined so that I1 + I2 = SumI1I2 and Q1 + Q2 = SumQ1Q2. The summed I and Q signals no longer exhibit phase and amplitude correlation—their amplitudes are not equal at all times and the phase relationship between them varies. The resultant output from the mixer places I1/Q1 data on one side of the carrier and I2/Q2 data on the other side of the carrier as previously explained and shown in Figure 7.

The use of zero-IF complements the advantages of the complex transmitter by positioning independent data blocks directly adjacent to each other on either side of LO. The data processing path bandwidth never exceeds that of the RF data bandwidth. So in theory, the use of a complex mixer used in a zero-IF architecture provides a solution that requires no RF filtering while also optimizing baseband power efficiency, delivering lower cost per unit of unusable signal bandwidth.

Up to this point, the focus of this article has been on the complex mixer used as a zero-IF transmitter. The same principles work in reverse and the complex mixer architecture can be used as a zero-IF receiver. The same advantages that have been described for the transmitter equally apply to the receiver. When using a single-mixer to receive a signal, the image frequency must first be filtered out using an RF filter. In the zero-IF mode of operation there is no image frequency to worry about, and signals above LO will be received independently of signals below LO.

A complex receiver is shown below. The input spectrum is applied to both I and Q mixers. One mixer is driven with LO, the other with LO + 90°. The outputs of the receiver are I and Q. In the case of a receiver, it is not as easy to prove empirically what the output will look like for a given input, but if a tone is input above LO, as shown, the I and Q outputs will be at the (tone – LO) frequency and there will be an expected phase shift between I and Q where I leads Q. Similarly, if the tone were input below LO, the I and Q outputs would again be at (LO – tone) frequency but this time Q will lead I. In this way the complex receiver can distinguish energy above LO from energy below LO.

The output of the complex receiver will be the sum of the I/Q information representing the spectrum that was received above LO and the I/Q information representing the spectrum that was received below LO. This concept was described earlier for the complex transmitter where a summed I and summed Q signal is applied to the complex transmitter. In the case of the complex receiver, the baseband processor receiving the summed I and summed Q information will easily be able to distinguish upper and lower frequencies using a complex FFT.

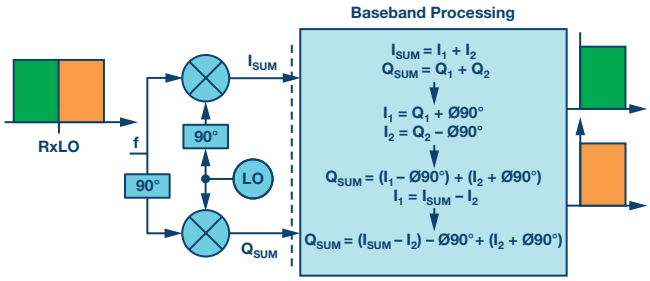


Figure 9. Zero-IF complex mixer receiver configuration.

When the summed I and summed Q signals are received, there are two knowns—the summed I signal and the summed Q signal—but there are four unknowns, namely I1, Q1, I2, and Q2. Because there are more unknowns than knowns, it would seem impossible to solve for I1, Q1, I2, and Q2. However, it is also known that I1 = Q1 + 90° and that I2 = Q2 – 90°, and with these two additional knowns it is now possible to solve for I1, Q1, I2, and Q2 using the received summed I and summed Q signals. In fact, we only need to solve for I1 and I2 because the Q signals are just copies of the I signals with a ±90° phase shift.

Limitations

In practice, the performance of the complex mixer has struggled to completely eliminate the image signal. This limitation could be considered as having two pronounced effects on radio architecture design.

Even with the performance limitation, complex IF does bring tangible benefits. Let us consider the low IF example in Figure 10. Accepting the performance limitation, we do still see an image. However, that image is heavily attenuated from that which we would expect to see from a single-mixer design (see Figure 6). Although the complex mixer continues to require a filter, the filter profile can be much more relaxed and its implementation simpler and lower cost.

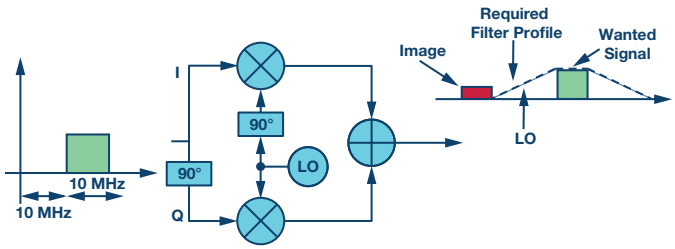


Figure 10. Practical implementation of the complex mixer. Note the attenuated image.

The filter complexity is inversely proportional to the distance between the image and the wanted signal. If we go to a zero-IF configuration, then that distance becomes zero and the image sits within the wanted signal band. The practical application of zero-IF theory has struggled, resulting in in-band image levels that degraded performance beyond an acceptable level (see Figure 11).

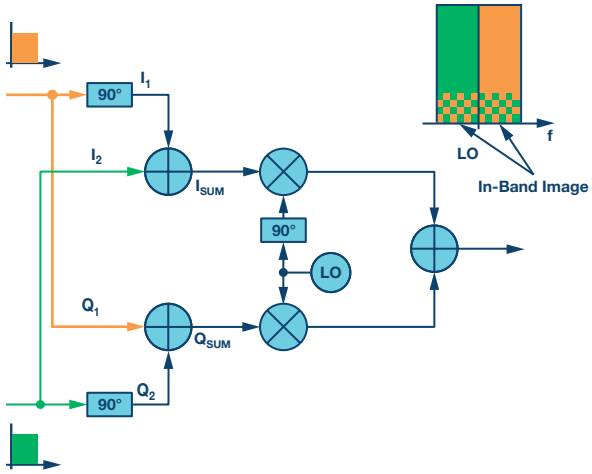


Figure 11. Zero-IF implementation restrictions.

The principles of the complex transmitter and receiver only hold true when the phase and amplitude requirements of the I and Q data paths are met. Mismatches in the signal paths will cause inaccurate cancellation of the image signals on both sides of the LO. Examples of such issues can be seen in Figure 10 and 11. In instances where zero-IF is not being used, filtering could be used to remove the image. However, if a zero-IF architecture is to be used, then the unwanted image falls directly within the spectrum of the wanted signal and a failure condition will occur if the image power is large enough. Therefore, the use of zero-IF and complex mixing can deliver an optimal system design solution but only when the design can eliminate the phase of amplitude mismatches along the signal paths.

Advanced Algorithm Enablement

The concept of the complex mixer architecture has existed for many years but the challenges of meeting the phase and amplitude requirements in a dynamic radio environment have restricted its use in a zero-IF mode. Analog Devices has overcome the challenge by a combination of smart silicon design and advanced algorithms. The design accepts that there will be signal path impairments; however, these are minimized by smart silicon design. The remaining imperfections are calibrated out by self-optimizing quadrature error correction (QEC) algorithms. Figure 12 provides a conceptual overview.

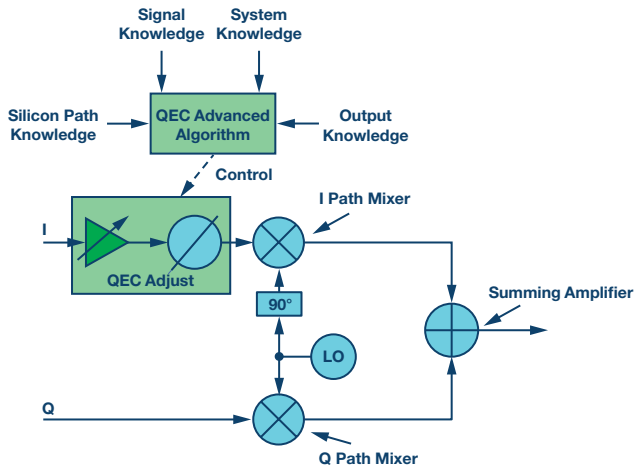


Figure 12. Advanced QEC algorithm and smart silicon design enabling zero-IF architecture.

On ADI transceiver devices such as the AD9371, the QEC algorithm sits within the on-chip ARM® processor. It has constant knowledge of the silicon signal path, the modulated RF output, and the input signal. It uses this knowledge to intelligently adapt the signal path profile in a controlled, predictive fashion, rather than a kneejerk reactive one. The algorithm performance is such that it can be best described as digitally assisting the performance of the analog signal path.

The dynamic QEC calibration algorithm is just one example, albeit a prominent one, of the advanced algorithms that reside and operate inside ADI transceivers. Others such as LO leakage cancellation coexist and lift the zero-IF architecture to an optimal level of performance. While these first generation of transceiver algorithms were primarily required for technology enablement, the second generation, such as digital predistortion (DPD), enhance the performance not just of the transceiver, but of the entire system.

All systems have imperfections that limit their performance. Whereas the first generation of algorithms have primarily focused on calibrating out on-chip limitations, the next generation uses the intelligence of algorithms to compensate for system performance and efficiency limitations external to the transceiver. Examples include PA distortion and efficiency (DPD and CFR), duplexer performance (TxNc), and passive intermodulation issues (PIM).

Conclusion

Complex mixers have existed for many years, but the image rejection performance that they provided did not allow them to be used in a zero-IF configuration. The combination of smart silicon design and advanced algorithms remove the performance barriers that had previously impeded the adaption of zero-IF architectures in high performance systems. With the performance limitations removed, the use of zero-IF architecture delivers savings in terms of filtering, power, system complexity, size, heat, and weight (the topic is extensively covered in an earlier article from Brad Brannon[1]).

In the case of complex mixers and zero-IF, we can consider the QEC and LOL algorithms as an enablement function. However, as the scope of the algorithmic development extends, it provides system designers with increased performance levels that allow them more flexibility in their radio designs. They may choose enhanced performance but they may also use the gains achieved from the algorithm to compensate for lower cost or size components in their radio designs.

References

[1] Brad Bannon. "Where Zero-IF Wins: 50% Smaller PCB Footprint at ⅓ the Cost." *Analog Dialogue*, Sept 2016.

Dave Frizelle [david.frizelle@analog.com] works as an applications manager in the Transceiver Product Group at Analog Devices Limerick supporting the integrated transceiver family of products. He has worked at ADI since graduation in 1998. His previous engineering roles include six years working in Japan and Korea supporting the development and design-in of ADI components into advanced consumer products.

Dave Frizzelle

Frank Kearney [frank.kearney@analog.com] works as algorithm development manager in the Communication Systems Engineering Team at Analog Devices Limerick. He has worked at ADI since graduation in 1988. He recently returned from China where he held the position of senior applications manager for the systems engineering team in Asia Pacific. He is currently a doctoral candidate at University College Dublin.

Frank Kearney

Understanding and Extending Safety Operation in a Sigma-Delta ADC

By **Miguel Usach Merino**

Share on

Abstract

New international standards and regulations have accelerated the need for safety systems in industrial equipment. The objective of functional safety is to protect people and assets from harm. This is achieved through the use of safety functions that target specific hazards. Safety functions consist of a chain of subsystems including sensor, logic, and output blocks and require system-level and integrated circuit level expertise to deliver an IC with the right set of features. This articles explores the AD7770 Σ-Δ ADC as an example of a high performance IC conceived and designed to provide an advanced set of features in both the analog and digital domains that simplifies the design of safety systems.

Introduction

Paraphrasing one of the Murphy's laws: "If the possibility exists of several things going wrong, the one that will go wrong is the one that will do the most damage."

A system that could produce a direct threat to human life, or an indirect threat, like a failure in machinery, must be designed to minimize the probability of failure and its consequential negative effects. To guarantee that the level of probabilistic random and deterministic failure is kept as low as possible, a specific design methodology must be followed. In the industry, this design methodology is called functional safety. This methodology requires a meticulous analysis of the system to identify any potentially hazardous situations and the application of best practices to bring the risk of malfunction down to tolerable levels in the component, subsystem, or system, such as unsafe states (that is, the voltage is too high, or diagnostic failure).

The idea behind functional safety is to keep the system in a safe state when an error is detected, like disconnecting the active outputs if the conversion results from an external sensor are out of bonds.

IEC-61508 is the standard reference for functional safety design in industrial equipment, and has been adapted/interpreted for different industries, like the ISO-26262 for automotive, or IEC-61131-6 for programmable controllers.

Designing to a functional safety standard can be quite tedious as a top-down meticulous analysis must be done, from the overall system description to the internal functional blocks of the components used. This analysis is necessary to guarantee enough level of protection to avoid any hazardous situation and to minimize the probability of the occurrence of undetected errors. A functional safety system should be designed in such a way that the system is capable of detecting any error and reacting fast enough to minimize the probability of the hazardous situation, as shown in Figure 1.

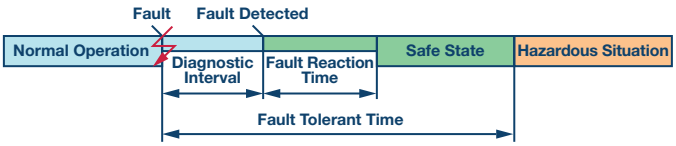


Figure 1. Reaction time in a functional safety system.

How to Design a Functional Safety System

The first step is a hazard analysis to identify the ways that somebody could get hurt. After the analysis of those situations, the system should be designed in such a way that hazardous situations can be avoided. If there is an unavoidable situation, add a functional system to detect the unsafe state and bring the system to a safe situation.

To illustrate the problem, let's assume the hypothetical system as shown in Figure 2. Depending on the tank temperature, a valve connected to a tank is opened a percentage to minimize the risk of explosion. A DAC controls the aperture of the valve through a motor. The system described is open-loop.

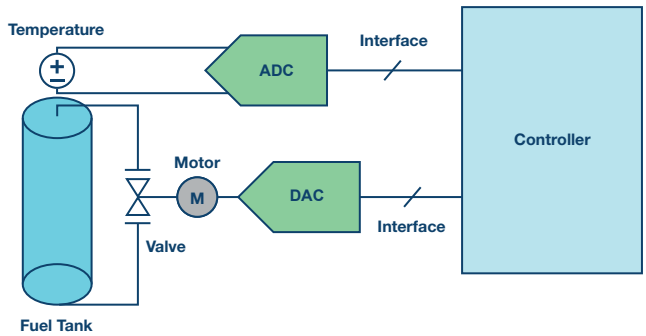


Figure 2. Open-loop valve control system signal chain.

The hazard analysis reveals two situations that could produce an unsure state:

- The temperature is incorrectly measured. Consequently, the aperture of the valve is incorrect.
- The DAC fails to open/close the valve correctly.

The next step is to evaluate the risk associated with each hazard as,

$$Risk = probability\ of\ occurrence\ of\ harm \times severity\ of\ the\ harm$$

Once the risk is identified, the next step is to design a functional safety system capable of reducing the risk to a tolerable level.

IEC-61508 defines four safety integrity levels (SIL) that define the level of risk reduction achieved by a safety function. There are two different target probabilities: the failure on demand, which applies to systems that are in stand-by until an event triggers (airbags are a good example), and probability of failure per hour, which applies to systems that are constantly operating, as could be the case in the previous example. Table 1 summarizes rough equivalences between SIL according to IEC61508, ISO 26262 (ASIL, automotive), and the avionics standard for expected failures on demand and per hour.

Table 1. Risk Levels Approximations for Different Standards

Probability of Failure on Demand	Probability of Failure Per Hour	Standard		
		IEC 61508 SIL Level	Automotive	Avionics
0.1 to 0.01	10^{-5} to 10^{-6}	1	A	D
0.01 to 0.001	10^{-6} to 10^{-7}	2	B	C
0.001 to 0. 0001	10^{-7} to 10^{-8}	3	C/D	B
0.0001 to 0.00001	10^{-8} to 10^{-9}	4		A

SILs are based on the required reduction and minimization of an undetected failure generating a malfunction on the system and potentially triggering an undesirable situation.

What Are Diagnostic Coverage Requirements?

The probability of undetected failure decreases with the increment of the diagnostic coverage. If the system can provide 99% diagnostic coverage, SIL3 can be achieved; for 90% diagnostic coverage, SIL2 can be claimed. If the coverage is only 60%, SIL1 can be achieved. In other words, the occurrence of undetected errors decreases with the level of redundancy.

The easier way to achieve SIL2 or SIL3 is by employing components already qualified for this grade of protection. This is not always possible, as these types of components target specific applications, which may not be identical to your circuit or system. Consequently, the assumptions applied to qualify the device may not apply, and the level of protection may not be the same.

Another approach for achieving high diagnostic coverage is by applying redundancy at the component level. In this case the error detection is not done directly, but indirectly by comparing two (or more) outputs that should be the same. However, this approach will increase the power consumption, the area, and, probably more importantly, the final cost of the system.

Increasing Error Detection and Redundancy at the Component Level

A common source of error is the data transmission in the external interface; if any single bit is corrupted during the transmission, the data can be misinterpreted by the receiver and can generate an undesirable situation. To calculate the total error that occurs in transmitting data, the BER (bit error rate) can be used. The BER indicates the number of bits corrupted due to noise, interferences (EMC), or any other physical reason.

$$BER = \frac{bits\ corrupted}{bits\ transmitted}$$

The BER can be physically measured in the system. Generally, this number is defined in many standards, as is the case in HDMI®, or an estimated value can be used. The minimum standard BER for modern data traffic is 10^{-7}. This number may be considered too pessimistic for many applications, but it can be used for reference purpose.

A BER of 10^{-7} means that 1 bit in every 10 million bits will be corrupted. For a SIL3 system, the target maximum probability of errors per hour is 10^{-7}. If our system transmits 32 bits of data between the ADC to the controller with an output data rate of 1 kSPS, then in one hour it will transmit:

$$bits\ per\ hour = 32 \times 1000 \times 3600 = 115{,}200{,}000\ bits$$

In this case, the error rate will increase up to $1.5e^{-5}$, and this is only the contribution from one interface; the total contribution of transmission errors should be kept to between 0.1% to 1% of the total error budget.

In this case, the error can be detected by adding a CRC algorithm. The number of bits corrupted that can be detected is defined by the Hamming distance of the CRC polynomial, such as $X^8 + X^2 + X + 1$, which has a Hamming distance of 4 and is capable of detecting up to three corrupted bits per frame transmitted. Table 2 summarizes the probability of error based on the number of bits transferred per hour for a CRC Hamming distance of 4 at different bits per hour, when transferring 32 bits of data plus 8 bits CRC.

Table 2. Probability of Error for a CRC Hamming Distance of 4

Number of Data Bits Per Hour	Probability of Undetected Error Per Hour
144,000,000	$2e^{-14}$
432,000,000	$6e^{-14}$
2,160,000,000	$3e^{-13}$

The level of diagnostic using the CRC can be augmented by reading back the register that was written and confirming that the data has been correctly transferred. This action will increase the level of diagnostic, but the level of error detection on the CRC polynomial used must be capable of detecting the expected number of bits corrupted based on the BER probability.

What Can Be Done to Minimize the Failure Probability?

A manufacturer who claims that a component has been designed for a functional safety system should be able to provide the FIT and, more importantly, failure modes, effects, and diagnostic analysis (FME(D)A). This data is used to analyze the IC in a specific application to calculate diagnostic coverage (DC), safe failure fraction (SFF), and dangerous failures rates for the system.

The FIT is a measurement of the reliability of a device. FIT for an IC can be calculated based on accelerated life testing or based on industry standards, such as IEC62380 and SN29500 where the average operating temperature in the application, package type, and number of transistors are considered to generate a FIT prediction. The FIT does not provide any information about the root cause of the failure, just a reliability prediction for the device. Generally speaking, unless each functional block can be checked, directly or indirectly, the final error probability will be too high to meet the SIL targets for any SIL2 or SIL3 safety functions.

The objective of the FME(D)A is to provide a comprehensive document covering the analysis of all the blocks implemented in the silicon, the consequences of a failure in the block directly or indirectly, and the different mechanism or methods that allow detection of the failure. As previously mentioned, those analyses are done based on a given signal chain/application, but the level of detail provided should be high enough to easily generate an FME(D)A analysis for a different system/application.

What Can Go Wrong in a Σ-Δ ADC?

A general analysis of a Σ-Δ ADC highlights multiple sources of errors due to the internal complexity of this device, such as:

- Reference disconnected/damage
- Input/output buffers/PGAs damage
- ADC core damage/saturation
- Incorrect internal regulator supply
- Incorrect external supply

These are only some problems that could generate a failure in a device block, but there are other sources of failure that may not be as obvious as the previously listed ones, such as:

- Internal bonding damage
- Bonding short circuit with adjacent pin
- Leakage current increment

For example, could the component detect if the V_{REF} leakage current increases, generating a drop on the internal reference voltage?
To check this type of malfunction, the ADC should be capable of selecting between different references for conversion and have the V_{REF} as an input for conversions.

How could you detect if the internal fuses have regrown or are otherwise corrupted, which could load an incorrect configuration at power-up? Those are examples of things that can go wrong even if the probability is really low. All of the potential failures, especially the ones that can be very rare, and the way that they can be detected (if any), must be well documented in the FME(D)A document. This document summarizes the failures and the assumptions made based on a specific application and/or configuration to maximize the detection and minimization of undetected errors.

ADI's modern ADI Σ-Δ ADCs, like the AD7770, AD7768, or AD7764, implement multiple diagnostic detectors to increase the fault tolerance protection, and to detect functional errors in both digital and analog blocks. Example of these blocks are:

- CRC checker for the fuses, registers, and interfaces
- Overvoltage/undervoltage detectors
- Reference and LDO voltage detectors
- Internal fixed voltage for PGA gain testing
- External clock detector
- Multiple reference voltage sources

In addition to these features, the AD7770 ADC integrates an auxiliary 12-bit SAR ADC that can be used to increase the diagnostic capability of the device, with uses such as:

- It implements an alternate architecture that can offer some benefits like providing a different level of immunity to EMC.
- It is powered through different supply pins, which can be used as a reference.
- It is fast enough to monitor the eight Σ-Δ channels for a single conversion of a Σ-Δ channel as a monitor but with a different accuracy.
- It provides conversion results using a different serial interface (SPI).
- It provides access to all internal voltage nodes for diagnostics like the external supplies, V_{REF}, V_{CM}, LDO output voltage, or internal voltage reference.

Figure 3 shows the internal block diagram of the AD7770 ADC. The blocks that include an internal monitor are highlighted in purple, the blocks highlighted in green can be actively monitored, and the blocks highlighted in blue contain both internal and active monitoring functionality.

Conclusion

Functional safety consists of reducing the mathematical probability of undetected errors by increasing system/block monitoring and diagnostic coverage. The easier way to increment the coverage is by adding redundancy, but this penalizes the system in multiple ways, especially cost. Recent ADI Σ-Δ ADCs, like the AD7124 or AD7768, implement many internal error detectors, which simplify the design of a functional safety system, keeping the overall complexity low compared with other solutions. The AD7770 is a good example of a precision Σ-Δ ADC design that is ahead what's possible due to its integrated monitoring and diagnostic capabilities, which include an internal redundant converter to maximize diagnostic coverage.

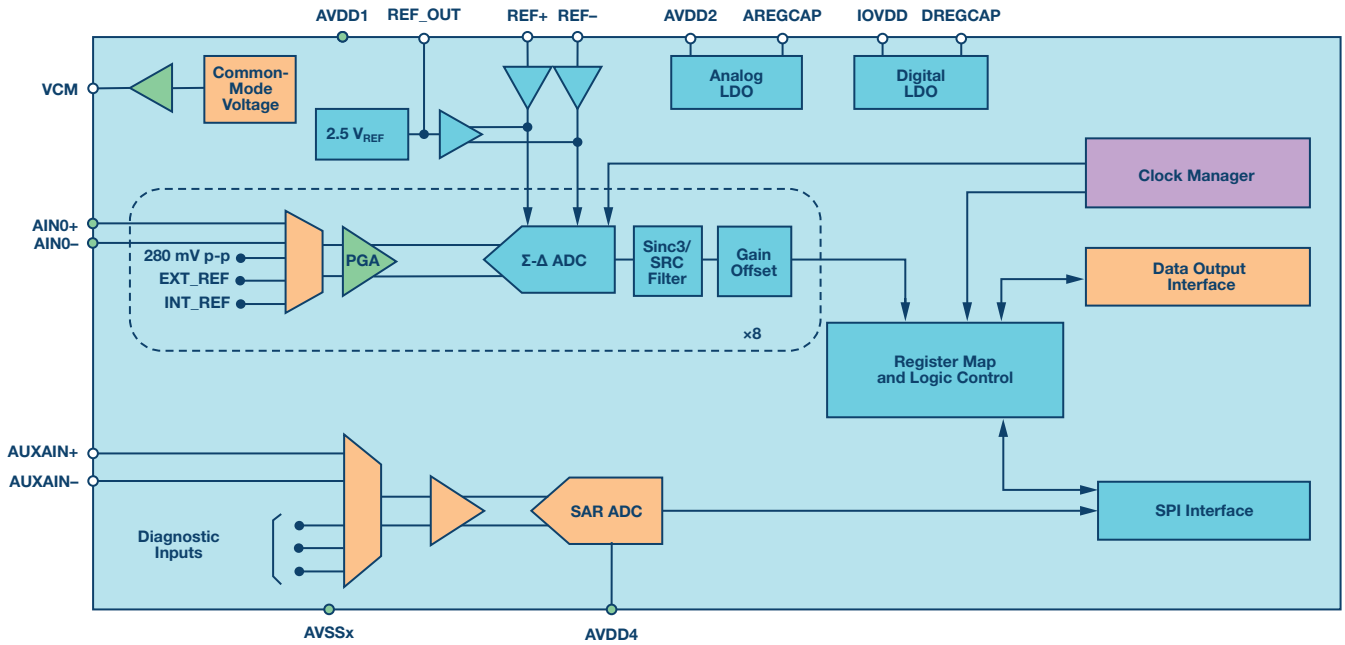


Figure 3. AD7770 ADC's diagnostic and monitoring blocks.

Miguel Usach Merino [miguel.usach@analog.com] received his degree in electronic engineering from Universitat de Valencia. Miguel joined ADI in 2008 and works as an applications engineer in the Linear and Precision Technology Group in Valencia, Spain.

Miguel Usach Merino

Also by this Author:

Integrated Capacitive PGAs in ADCs: Redefining Performance

Volume 50, Number 3

Rarely Asked Questions—Issue 138
This Noise Will Keep You Up at Night

By **Gustavo Castro**

Share on

Question:

What is the smallest voltage I can measure?

Answer:

My first project as an engineer was to measure the settling time of what was going to be a 6½ digit DMM. It didn't seem like a big deal, I just had to figure out the final settled value and work my way backward to the smallest detectable change. I got everything setup, shorted the inputs, and started increasing the aperture time[i]. As expected, the noise would come down … until it didn't. The baseline just kept moving. I had eliminated extrinsic noise sources, thermal EMF, and even drafts from the ac vents. These random fluctuations were coming from the intrinsic noise of the circuit. But after eliminating most of the broadband noise, there was this other noise that would not go away. Anyone who has tried this would have noticed the same limitation. To the contrary, we may find more noise than if we had stopped sooner! We know that we are in the 1/f noise region when that happens.

This so-called 1/f noise (or flicker noise) is the most pervasive limitation for precision measurements. The name comes from the fact that its power spectral density is inversely proportional to frequency, as expressed by:

$$Noise\_Power\left(f\right)=\left(\frac{k}{f^{\alpha}}\right)$$

Where k is a magnitude coefficient and α is an exponent that will take values greater than 0, but the canonical form is for α = 1. This noise eventually becomes smaller than broadband noise, producing a corner as shown in Figure 1. Evidence of this type of noise has been found outside of electronic circuits, including the rotation of the Earth, economic indicators, and biological systems, to name a few. While the fundamental cause keeps eluding the most brilliant scientists, we must understand how to mitigate it if we want to perform low level measurements.

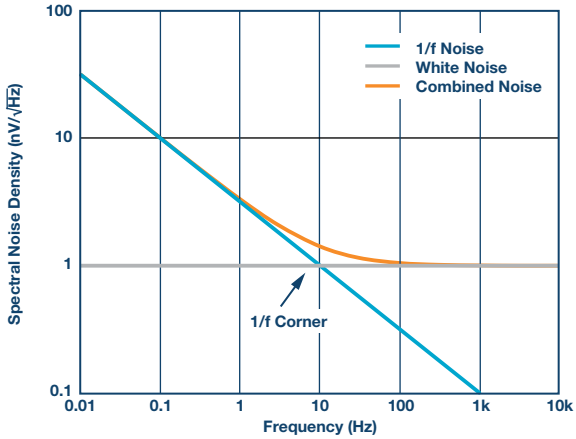


Figure 1. Typical noise spectral density plot for a low noise electronic component.

Let's start with off-the-shelf components. The highest sensitivity ADC you will find these days in an IC is AD7177-2, and that is 200 nV p-p at 5 SPS. But we can do better than that by adding some gain before the ADC. We need an amplifier that is both low noise and with a low 1/f corner. The easiest is to look up the 0.1 Hz to 10 Hz noise specification on the data sheet, which is equivalent to recording measurements for 10 seconds with 10 Hz bandwidth.

If you have been paying attention, you may have read about the AD797 op amp being used in the LIGO experiment to detect gravitational waves for the first time in human history. The AD797 has a noise specification of

50 nV p-p (8 nV rms) from 0.1 Hz to 10 Hz. The AD8428, the lowest noise instrumentation amplifier, is only 40 nV p-p (7 nV rms). Because these amplifiers are built in bipolar processes, their current noise can be significant if used with large source resistance (including gain resistors), and current noise also has a 1/f corner! And don't forget that resistors themselves can show current-dependent excess noise due to their construction. Metal foil and wirewound resistors tend to have the lowest noise indices.

A neat trick to avoid 1/f noise is to modulate the signal to a region where there is no 1/f noise and then demodulate it. This trick, known as chopper stabilization, has been used for decades to shift the 1/f noise to a different frequency band, where it can be easily filtered out. Zero-drift op amps like the ADA4528-1 and ADA4522-1 take advantage of this (and other tricks) to get about 100 nV p-p (16 nV rms) from 0.1 Hz to 10 Hz, mostly due to white noise. A simpler alternative is to parallel multiple amplifiers to reach lower noise levels, since this is equivalent to averaging uncorrelated noise sources.

The bottom line is that with off-the-shelf components, you can detect signals just below the 10 nV mark, and paralleling amplifiers will get you close to the 1 nV level. Anything below that will require special (and perhaps pricy) techniques. But no matter what you do, 1/f will always find a way to resurface.

So, what if we were to record several measurements for a really long time? Would 1/f noise make this an impossible task? Let us bring some perspective: if we had recorded the AD797 noise from the moment of the Big Bang until the time of reading this article,[ii] it would only be three times larger than if we had measured it for the last 10 seconds;[iii] so I wouldn't lose any sleep over that.

References

[i] The aperture time in a DMM is the interval during which the signal gets integrated or averaged.

[ii] Assuming 4.32e17 seconds have passed since the Big Bang.

[iii] Hypothetically speaking, since there is no evidence 1/f follows this curve for that long. Aging and other factors start to play in at longer measurement intervals.

Gerstenhaber, Moshe, Rayal Johnson, and Scott Hunt. "No Pain High Gain: Building a Low Noise Instrumentation Amplifier with Nanovolt Sensitivity." *Analog Dialogue*, Vol. 49, May 2015.

Horowitz, Paul and Winfield Hill. *The Art of Electronics*. Cambridge University Press, 1989.

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

Seifert, Frank. "Resistor Current Noise Measurements." Open access LIGO document, LIGO-T0900200.

"The Chirp Heard 'Round the Universe." Analog Devices, Inc.

van der Ziel, Aldert. "Unified Presentation of 1/f Noise In Electronic Devices: Fundamental 1/f Noise Sources." *Proceedings of the IEEE*, vol. 76, no. 3, March 1988.

Weissman, M. B. "1/f Noise and Other Slow Nonexponential Kinetics in Condensed Matter." *Reviews of Modern Physics*, 1988.

West, Bruce and Michael Shlesinger. "The Noise in Natural Phenomena." *American Scientist*, 78(1), 1990.

Gustavo Castro [gustavo.castro@analog.com] is a system applications engineer in the Linear and Precision Technology Group in Wilmington, MA. His main interests are analog and mixed-signal design for precision signal conditioning and electronic instrumentation. Prior to joining Analog Devices in 2011, he worked for 10 years designing high performance digital multimeters and precision dc sources at National Instruments. Gustavo received his B.S. degree in electronic systems from Tecnológico de Monterrey and his M.S. degree in microsystems and materials from Northeastern University. He holds three patents.

Gustavo Castro

Also by this Author:

Common Sense for Current Sensing

Volume 50, Number 3

Passive Intermodulation (PIM) Effects in Base Stations: Understanding the Challenges and Solutions

By **Frank Kearney** and **Steven Chen**

Share on

Introduction

It is well known that *active* components will generate nonlinearities in systems. Various techniques have been developed to improve the performance of such devices during both the design and operational phase. It is easy to neglect that the passive devices can also introduce nonlinear effects; although sometimes relatively small, these nonlinearities can, if not corrected, have serious effects on system performance.

PIM stands for "passive intermodulation." It represents the intermodulation products generated when two or more signals transit through a passive device with nonlinear properties. The interaction of mechanical components generally causes nonlinear elements. This is particularly evident at the junction of two different metals. Examples include: loose cable connections, dirty connectors, poor performance duplexers, or aged antennas.

Passive intermodulation is a significant issue within the cellular industry and it is extremely difficult to troubleshoot. In cell communication systems, PIM can create interference and will reduce receiver sensitivity or may even inhibit communication completely. This interference can affect the cell that created it, as well as other nearby receivers. For example, in LTE Frequency Band 2, the downlink is specified from 1930 MHz to 1990 MHz, while the uplink ranges from 1850 MHz to 1910 MHz. If two transmitter carriers, located at 1940 MHz and 1980 MHz, are transmitting from the base station system with PIM, their intermodulation will lead to a component at 1900 MHz, which will fall into the receive band. This will affect the receiver. Furthermore, the intermodulation item at 2020 MHz may affect other systems.

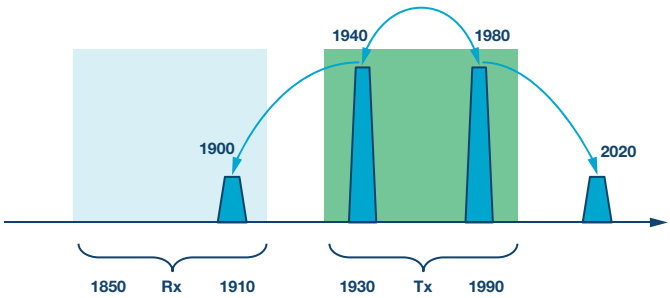


Figure 1. Passive intermodulation, falling back into the receiver band.

As the spectrum has become more crowded and antenna sharing schemes become more common, there is a corresponding increase in the possibility of PIM generation from the intermodulation of different carriers. The traditional way of using frequency planning to avoid PIM becomes almost impossible. Coupled with the challenges just mentioned, the adoption of new digital modulation schemes like CDMA/OFDM means that the peak power of the communication system also increases, adding to the severity of the PIM issue.

PIM has been highlighted as a serious problem for service providers and equipment suppliers. Detecting and, where possible, solving the problem delivers increased system reliability and reduced operation cost. In this article we attempt to review the sources and causes of the PIM, along with technologies proposed to detect and solve it.

PIM Classification

Our initial observations indicate that PIM has three distinctive types, each with differing characteristics and each requiring differing solutions. We choose to classify those type as *design PIM*, *assembly PIM*, and *rusty bolt PIM*.

Design PIM

Certain passive components, in combination with their transmission lines, are known to contribute to passive intermodulation. Therefore, when designing a system, development teams will choose passive elements with minimal or acceptable levels of PIM as specified by the component manufacturer. Circulators, duplexers, and switches are particularly prone to the effect. Designers may choose to accept higher levels of passive intermodulation by selecting lower cost, smaller size, or lower performance options.

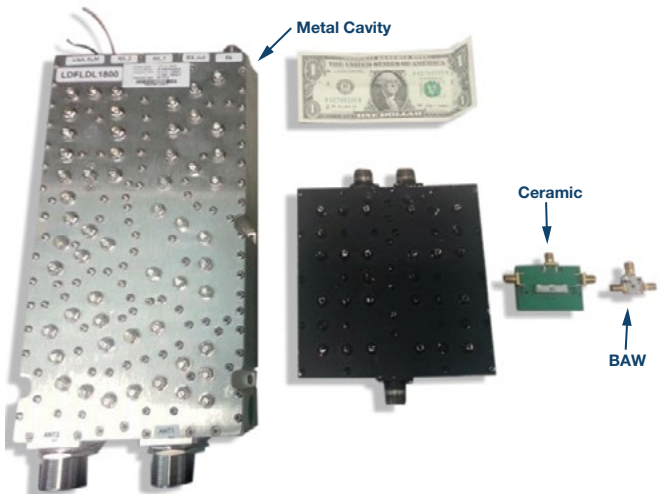


Figure 2. Component design trade-offs, size, power, rejection, and PIM performance.

If designers do choose to use the lower performance components, the resultant higher levels of intermodulation may fall back within the band of the receiver and result in desensitizing it. It is important to note that in these instances, the unwanted spectral emissions or loss of power

efficiencies may not be as concerning as the desensitizing effect of PIM on the receiver. This problem is of particular significance in small cell radio designs. ADI is currently at an advanced development stage, whereby the PIM contributed by the static passive elements such as the duplexer can be detected, modeled, and subtracted (canceled) from the received signal (see Figure 3).

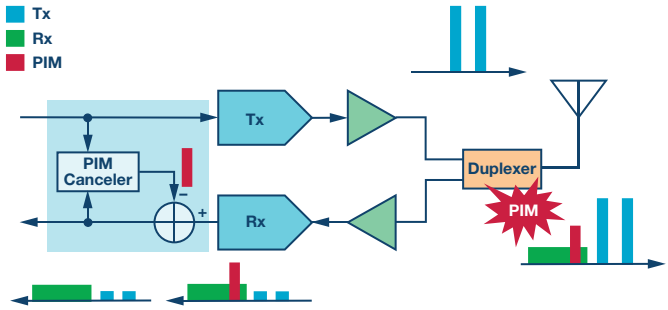


Figure 3. PIM generation and cancellation algorithm.

The algorithm works because it has knowledge of the carriers and can use correlation at the receiver to determine the intermodulation artifact before subtracting it from the received signal.

Limitations on the algorithm start to emerge when correlation can no longer be used to determine the intermodulation artifact. Figure 4 provides an example of this. In this instance, two separate transmitters share the same antenna. If we assume that the baseband processing for each path is independent from the other, then the algorithm is unlikely to have knowledge of both and, hence, is limited in the correlation/cancellation it can perform at the receiver.

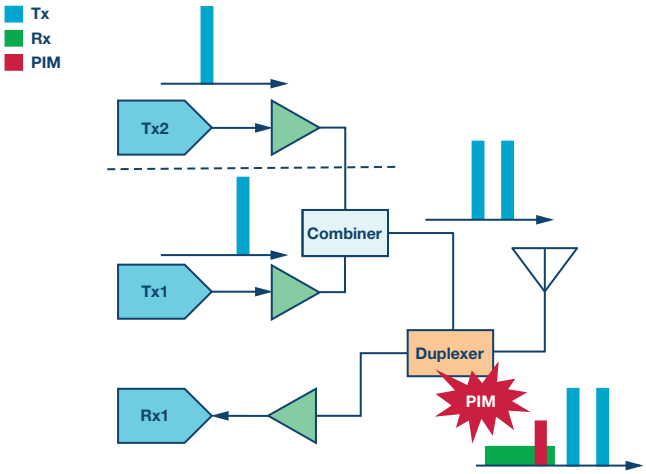


Figure 4. Multiple sources sharing the same antenna.

Complexity Adding to the PIM Challenge

As site access and costs challenge service providers, we are starting to see an increasing number of instances where separate transmitters share a single wideband antenna. The architectures can be a mix of band and formats: T_{DD} + F_{DD}; T_{DD}: F + A + D, F_{DD}: B3, etc. Figure 5 provides an overview of such a configuration. In this instance, the customer is implementing a complex but real configuration; one branch is T_{DD} dual band and the other is F_{DD} single band, employing a duplexer. The signals are combined and share a single antenna. The intermodulation between the Tx1 and Tx2 signals occurs passively in the path from combiner, in the transmission line to the antenna, and in the antenna itself. The resultant intermodulation artifact falls back within the band of the F_{DD} receiver, Rx2.

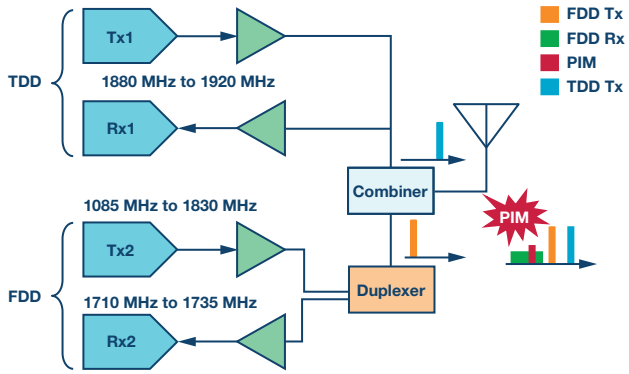


Figure 5. F_{DD}/T_{DD} single antenna implementation.

Figure 6 shows a practical analysis for a dual-band system. Note that in such instances we need to consider well beyond third-order passive modulation artifacts. In this case, the focus is on intermodulation artifacts from one band (intra) falling within the receive band of the other.

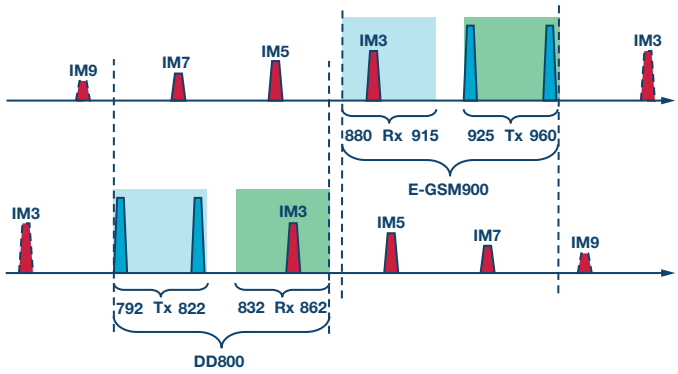


Figure 6. Multiband PIM issues.

Assembly PIM

The second categorization of PIM is what we might term *assembly PIM*. While the system may operate satisfactorily when installed, the performance will often degrade over time due to weather or poor initial installation. When this occurs, passive elements (that is, connectors, cables, cable assemblies, waveguide assemblies, and components) of the signal path will typically start to exhibit nonlinear behavior. In fact, some of the major occurrences of PIM will be caused by connectors, connections, and even the feeder for antennas themselves. The resultant effect can be similar to that of the design PIM, as discussed earlier. Hence the same PIM measurement theory can be used that is specifically looking for the presence of passive intermodulation products.

Typical contributors to assembly PIM are:

- Connector mating interface (typically Type N or DIN7/DIN16),
- Cable attachment (mechanical stability of the cable/connector junction),
- Materials (brass and copper are advised, ferromagnetic materials show nonlinear characteristics),
- Cleanliness (contamination from dirt or moisture),
- Cable considerations (quality and robustness of cable),
- Mechanical robustness (flexing due to wind and vibration),
- Electrothermally induced PIM (due to the variation of conductance as temperature varies in response to the time varying power dissipated by RF signals with nonconstant envelopes).

Environments where there are wide temperature variations, salt air/polluted air, or excessive vibrations tend to exacerbate PIM. Although the same PIM measurement techniques can be used as for design PIM, the presence of assembly PIM can be considered as an indication of system degradation both in terms of performance and reliability. If unresolved, the weaknesses that are causing PIM may continue to escalate until complete transmission path failure occurs. The approach of using PIM cancellation for assembly PIM might be considered as masking an issue, rather than resolving it.

In such circumstances it would be expected that users will not want to cancel the PIM but be notified of its existence with the aim to rectify its root cause. Elimination comes from first determining where on the system the PIM is been introduced and then repairing or replacing that specific element.

Whereas we might consider design PIM as being quantifiable and stable, assembly PIM, as previously described, is not stable. It may exist under a very narrow set of conditions and its amplitude variation can be in excess of 100 dB. One single shot offline sweep may fail to capture such instances; ideally the transmission line diagnostic needs to be captured in tandem with the PIM event.

PIM Beyond the Antenna (Rusty Bolt PIM)

PIM is not limited to the wired transmission path but can also happen beyond the antenna. The effect is also known as rusty bolt PIM. In such a circumstance, the passive intermodulation occurs after the signals have left the transmitter antenna with the resultant intermodulation reflecting back into the receiver. The term rusty bolt comes from the fact that in many instances the intermodulation source can be a rusty metallic object such as a mesh fence, a barn, or even a drain pipe.

Reflections caused by metal objects are to be expected. In these instances, however, the metal objects do not only reflect the received signals but also produce and radiate intermodulation artifacts. The intermodulation occurs just as it did within the wired signal path—that being at the junction of two different metals or junctions of dissimilar materials. The electromagnetic waves create surface currents that mix and reradiate (see Figure 7). The reradiated signals are usually very low in amplitude. However, if the radiating element (rusty fence, barn, or downpipe) is close to the receiver of a base station and if its intermodulation product falls within the receive band, the result will be receiver desensitization.

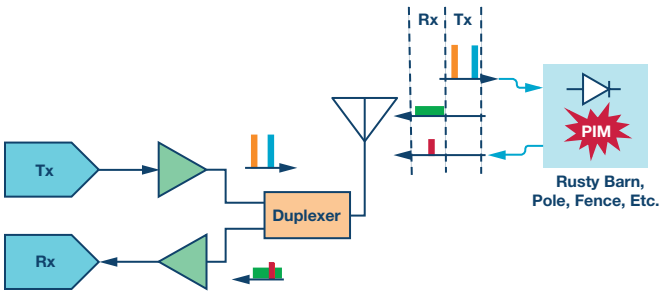


Figure 7. Beyond the antenna, or rusty bolt PIM.

In some instances, detection of the PIM source can be achieved by antenna positioning: the PIM level is monitored while the antenna position is changed. In other instances, time delay estimation can also be used to locate the source. If PIM levels are static, then standard algorithmic cancellation techniques can be used to compensate for the PIM. However, in many instances, vibration, wind, and mechanical movement can modulate the PIM contribution and make the cancellation challenge even more difficult.

PIM Detection: Locating the Source of PIM

Line Sweeping

Various *line sweeping* techniques can be implemented. Line sweeping measures the signal losses and reflections within a transmission system over the band of interest. It cannot be assumed that line sweeping will always accurately indicate the possible causes of PIM. Line sweeping can be considered more as a diagnostic tool that helps identify issues on the transmission line path. Early stage assembly issues can manifest themselves as PIM; if left unresolved, those assembly issues can escalate into more serious transmission line failures. Line sweeping usually breaks down into two basic tests: return loss and insertion loss. Both are very frequency dependent and both can vary greatly within a specified band. Return loss measures the power transfer efficiency of the antenna system. It is essential that minimal power is reflected back toward the transmitter. Any reflected power can distort the transmitted signal and, when powerful enough, cause damage to the transmitter. A return loss figure of 20 dB indicates that 1% of the transmitted signal is being reflected back to the transmitter and 99% is reaching the antenna—this is generally considered good performance. A return loss of 10 dB indicates that 10% of the signal is being reflected and should be considered poor. If the return loss measured 0 dB, 100% of the power would be reflected, and this would likely be the result of an open or short circuit.

Time Domain Reflections

Advanced time domain reflectometry (TDR) techniques could be used to first provide a reference map of an optimal system and second be used to determine where exactly along the transmission path impairments start to occur. Such a technique may allow operators to locate sources of PIM and make targeted and efficient repairs. Transmission line mapping could also alert operators to early signs of failure before they start to have a significant impact on performance. TDR measures the reflections that result from a signal traveling through the transmission line. The TDR instrument sends a pulse through the medium and compares the reflections from the unknown transmission environment to those produced by a standard impedance. A simplified TDR measurement block setup is shown in Figure 8.

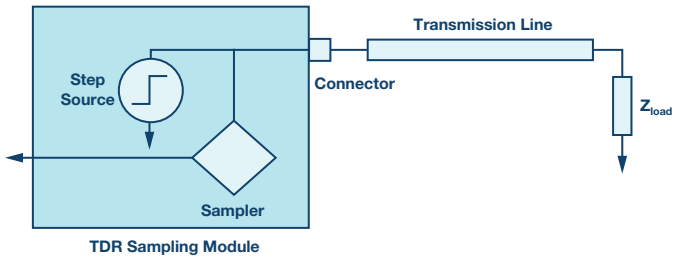


Figure 8. TDR setup block diagram.

Figure 9 provides an example of TDR transmission line mapping.

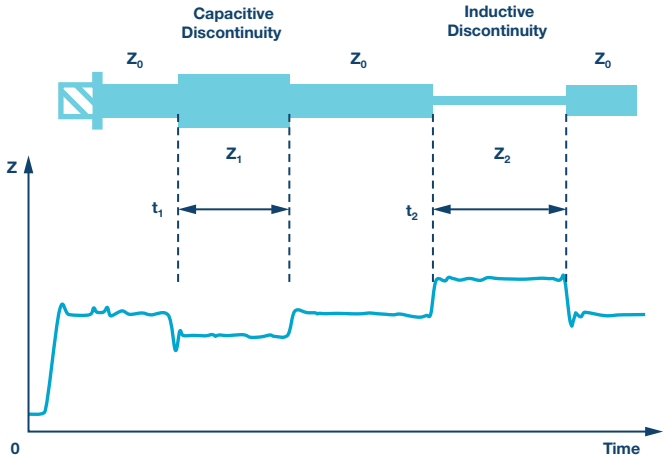


Figure 9. TDR mapping of transmission line.

Frequency Domain Reflections

While TDR and FDR both rely on the principle of sending stimuli down the transmission line and analyzing the reflections, the implementation of the two techniques is very different. The FDR technique uses an RF signal sweep instead of dc pulses as used by TDR. FDR is also far more sensitive than TDR and can locate faults or degradation in system performance with higher precision. The frequency domain reflectometry principle involves a vector addition of the source signal with reflected signals from faults and other reflective characteristics within the transmission line. While TDR uses very short dc pulses that inherently cover a very large bandwidth as stimulus, FDR swept RF signals can actually be run at the specific frequencies of interest (usually within the range where the system is expected to operate).

Figure 10. FDR principles, swept frequency return loss vs. distance.

Distance to PIM

It is important to note that while line sweeping may indicate impedance mismatches and, hence, the source of transmission line PIM, PIM, and transmission line impedance mismatches can be mutually exclusive. PIM nonlinearity can occur at points where line sweeping results do not indicate any transmission line issues. Hence a more sophisticated implementation is required whenever users are to be provided with a solution that not only indicates the presence of PIM, but also allows them to identify precisely where along the transmission line path the issue occurs.

Comprehensive PIM line testing operates in a similar mode to that described for design PIM cancellation, except in the instance where the algorithm examines the time delay estimation of the intermodulation product. It should be noted that the priority in these instances is not the cancellation of the PIM artifact, but the pinpointing of where along the transmission path the intermodulation is occurring. The concept is also known as distance to PIM (DTP). For example, in a two tone test:

Tone 1:

$$e^{j(w_1 (t + t_0) + \theta_1)}$$

Tone 2:

$$e^{j(w_2 (t + t_0) + \theta_2)}$$

w_1 and w_2 are the frequency; 0_1 and 0_2 are initial phase; t_0 is the initial time.

The IMD (lower side, for example) will be:

$$e^{j((2w_1 - w_2) (t + t_0) + (2\theta_1 - \theta_2))}$$

Many existing solutions require the user to break the transmission path and insert a PIM standard (a PIM standard is a device known to generate a fixed amount of PIM, which is used to calibrate the test equipment). The use of the PIM standard provides the user with a reference IMD that has a known phase at a specific position/distance along the transmitter path. Figure 11(a) provides an overview. The IMD phase 0_{32}, as shown in Figure 11, is used as a reference to position zero.

Once the initial calibration is performed, the system is then reconstructed and a system PIM measurement is taken, as shown in Figure 11(b). The phase difference between Θ_{32} and Θ'_{32} can be used to calculate the distance to PIM.

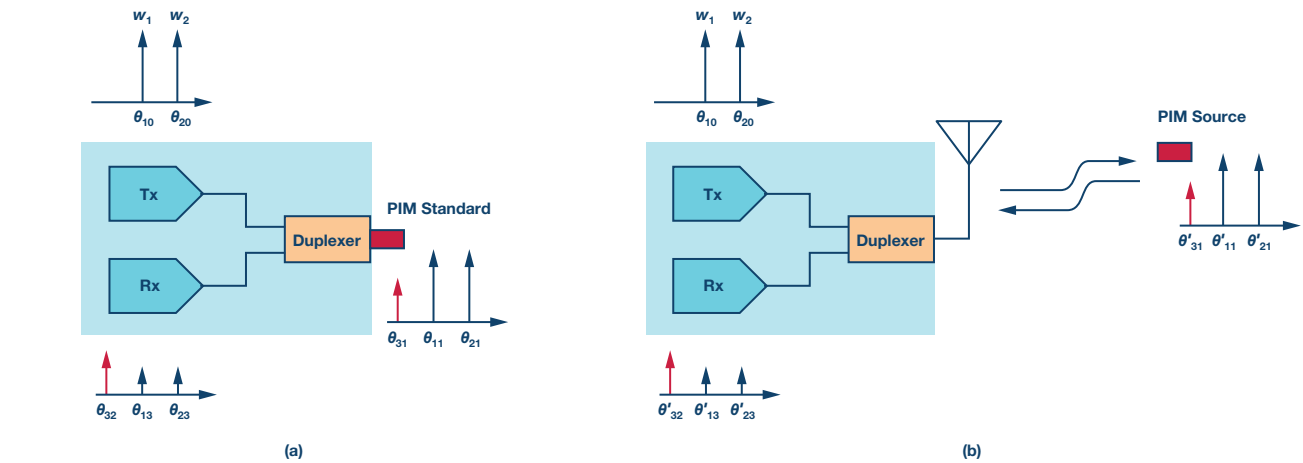


Figure 11. Distance to PIM.

$$(2w_1 - w_2) \times \frac{(2D)}{S} = \theta'_{32} - \theta_{32}$$

where D is the distance to PIM and S is the wave propagation speed (dependent on the transmission media).

Assembly and *rusty bolt* PIM can be slow and incremental processes; the base station may work efficiently after the initial installation, but over time these types of PIM phenomena may start to become more pronounced. As the level of PIM may be subject to environmental issues such as vibration or wind, the nature and characteristics of the PIM may be dynamic and fluctuating. Masking or canceling the PIM may not only be difficult, it could also be seen as masking a more serious issue that may, if unresolved, cause total system failure. In such circumstances operators will want to avoid the cost of total system tear-down but instead efficiently locate the PIM contributor and replace it.

Distance to PIM (DTP) technology also offers base station operators the possibility to track the degradation of their systems over time and highlight in advance what could emerge as problem issues. The knowledge allows the replacement of the weak points during scheduled maintenance, thus avoiding costly system downtime and dedicated repair efforts

Conclusion

Passive intermodulation is nothing new. It's a phenomenon that has existed for many years and has been understood for some time. In recent times, two distinct changes in the industry have brought it back into the forefront of attention:

First, advanced algorithms now provide a smart way to detect the presence/location of PIM and, where appropriate, compensate for it. Whereas previously radio designers had to choose components that met particular PIM performance requirements, with the assistance of PIM cancellation algorithms they have now gained a new degree of freedom. They have the ability to push for higher performance or, if they should choose, maintain the same performance level but with lower cost and smaller hardware components. The cancellation algorithms are digitally assisting the hardware elements.

Second, with the explosive growth in the density and diversity of base station towers, we are seeing a whole new range of challenges caused by particular system setups, such as the sharing of antennas. Algorithmic cancellation depends on knowledge of the primary transmitted signals. In instances where space on towers has a premium, various transmitters may share a single antenna, making the existence of unwanted PIM effects very likely. In such instances, an algorithm may have knowledge of certain portions of the transmitter path and may work efficiently. In cases where not all sections of a transmit path are known, the performance or implementation of the first generation of advanced PIM cancellation algorithms may be limited.

As the challenges in base station installations continue to grow, PIM detection and cancellation algorithms can be expected to deliver substantial gains and advantages to radio designers in the short term, but require development work to keep up with future challenges.

Frank Kearney

Frank Kearney [frank.kearney@analog.com] works as algorithm development manager in the Communication Systems Engineering Team at Analog Devices Limerick. He has worked at ADI since graduation in 1988. He recently returned from China where he held the position of senior applications manager for the Systems Engineering Team in Asia Pacific. He is currently a doctoral candidate at University College Dublin.

Steven Chen

Steven Chen [steven.chen@analog.com] graduated from Nankai University (Tianjin, China) in 2004, with an M.S. degree in communication engineering. He joined the Analog Devices Beijing Design Center after graduation as a digital design engineer, working in the Advanced TV and then High Speed Converter Groups. He currently works as an advanced algorithm development engineer in the Communication Systems Engineering Team. His research interests include digital signal processing, communication systems, and digitally assisted analog techniques.

Complex RF Mixers, Zero-IF Architecture, and Advanced Algorithms: The Black Magic in Next-Generation SDR Transceivers

Volume 51, Number 1

Analyzing and Managing the Impact of Supply Noise and Clock Jitter on High Speed DAC Phase Noise

By **Jarrah Bergeron**

Out of all device properties, noise can be an especially challenging topic to grasp and design for. Such challenges often lead to hearsay design rules and trial-and-error development. In this article, phase noise will be tackled with the objective of making quantitative sense of how to design around the contributions of phase noise in high speed digital-to-analog converters. The goal is to obtain a methodology that neither over-designs nor under-designs the phase noise requirement, but, rather, gets it right the first time.

Starting from a blank slate, the DAC is first treated as a block box. Noise can be generated internally, as any real component will generate some noise, or the noise can come in from external sources. The entrance from external sources can occur through any of the DAC connections, which broadly include power, clocking, and digital interfaces. These possibilities are illustrated in Figure 1. Each of these possible noise suspects will be investigated individually to understand their importance.

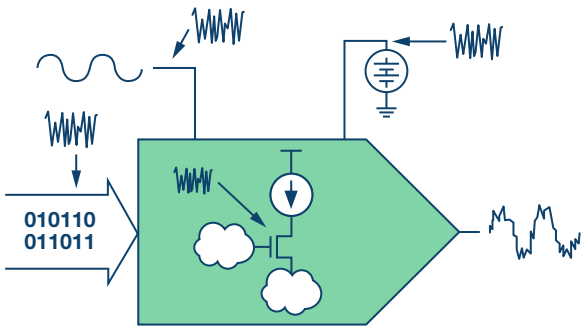


Figure 1. Sources of DAC phase noise.

The digital interface will be covered first and happens to be easiest to treat. The digital I/O is in charge of receiving samples to be outputted in the analog domain. As is commonly known, digital circuitry and the waveforms being received are noisy, as demonstrated by the eye diagrams. From this standpoint, the question that arises is: could all this noise and activity infiltrate into various regions inside the DAC and manifest itself as phase noise? Of course, the digital interface can cause noise elsewhere, but it is phase noise that is at issue here.

To prove out whether the I/O is a concern, the phase noise was compared with and without the digital interface on the AD9162 series of HSDAC parts. Without the interface, the device's NCO mode internally generates the waveform, effectively changing the DAC into a DDS generator. Figure 2 shows the experimental results.

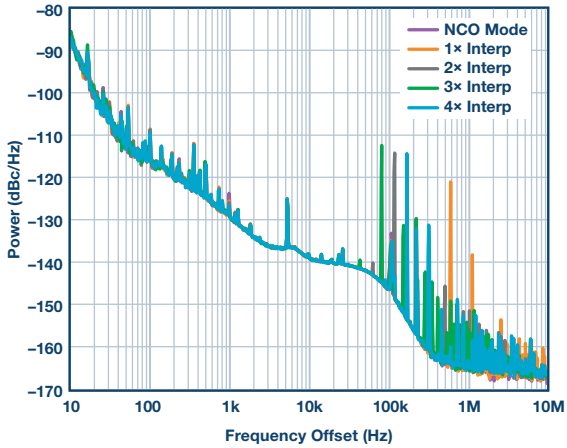


Figure 2. Phase noise at different interpolations.

Peaks do show up with the interface turned on and move around according to interface details. Now what is of interest is that the noise and all the curves are on top of each other. As a result, in this product line, the interface is not of concern, notwithstanding those spurs that may need attention depending on system requirements. Finding that the interface is of little concern leads into the next area of interest: clocking.

Clocking

Clocking is of prime concern for generating phase noise in DACs, namely the DAC clock. This clock dictates when the next sample will be sent, so any noise in the phase (or timing) directly affects the phase noise of the output, as shown in Figure 3. This process can be considered as the multiplication between each successive discrete value with a rectangular function whose timing is defined by the clock. Now, in the frequency domain, multiplication translates to the convolution operation. As a result, the desired spectrum gets smeared with the clock phase noise, as illustrated in Figure 4. The exact relationship, however, is not immediately obvious. A quick derivation follows.

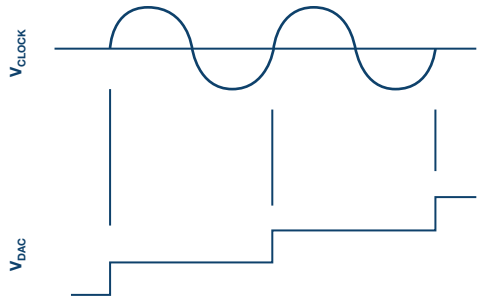


Figure 3. Clock to phase noise dependence.

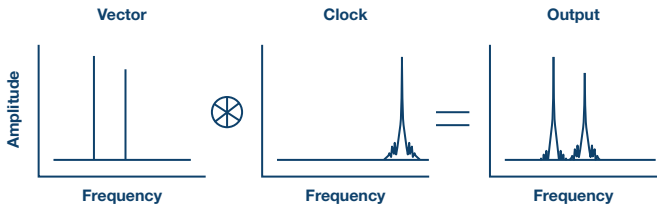


Figure 4. Phase noise convolution.

Taking a snapshot in time of the clock and output, an instance of the waveform is shown in Figure 5. The objective will be to find the ratios between the noise amplitudes of the clock and output shown as red arrows in Figure 6. Right triangles can be drawn and although none of the lengths are known, both triangles have a common horizontal side.

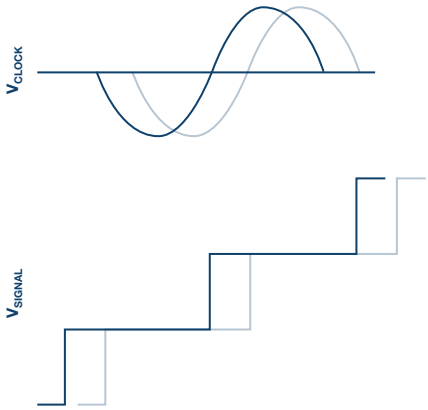


Figure 5. Waveform snapshot.

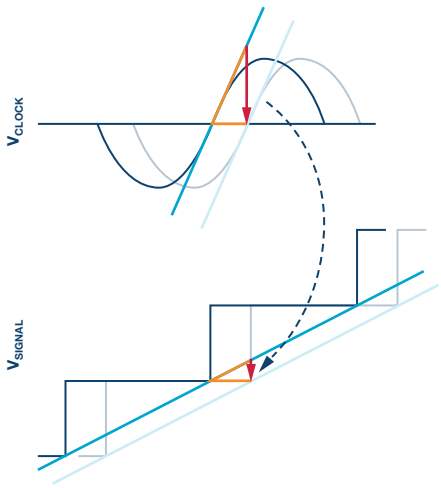


Figure 6. Phase noise relation.

Setting the slopes as derivatives of the respective waveforms, the geometry gives the following equation:

$$\frac{V_{CLK\_noise}}{\frac{\partial V_{CLK}}{\partial t}} = \frac{V_{SIG\_noise}}{\frac{\partial V_{SIG}}{\partial t}}$$

Rearranging for DAC noise yields the next equation:

$$V_{SIG\_noise} = V_{CLK\_noise} \frac{\frac{\partial V_{SIG}(t)}{\partial t}}{\frac{\partial V_{CLK}(t)}{\partial t}}$$

As we are often interested in sinusoid or near sinusoid waveforms for the DAC output and clock, the result can be simplified down. If this assumption doesn't hold, keep to the previous formulation.

$$V_{SIG\_noise} = V_{CLK\_noise}\left[\frac{V_{SIG} f_{SIG}}{V_{CLK} f_{CLK}}\right]$$

And then by reorganizing, we get this:

$$\frac{V_{SIG\_noise}}{V_{SIG}} = \frac{V_{CLK\_noise}}{V_{CLK}}\left[\frac{f_{SIG}}{f_{CLK}}\right]$$

Notice the noise relationship equates relative to the respective waveform amplitude, therefore, it is succinctly summarized relative to carrier. Also, by using logarithmic units, we arrive at the following equation:

$$N_{SIG} = N_{CLK} + 20\log_{10}\frac{f_{SIG}}{f_{CLK}}$$

The noise relative to the carrier is scaled up and down according to a ratio of the signal frequency to the clock frequency. Every halving of the signal frequency results in 6 dB improvement in noise. Examining the geometry, this makes sense as the triangle on the bottom would become more acute and shrink the vertical side. Also notice that increasing the clock amplitude doesn't improve phase noise if the noise increases at the same magnitude.

To prove this out, phase noise can be simulated by modulating the clock coming into the DAC. In Figure 7, the 5 GHz DAC clock is shown with light phase modulation at 100 kHz. Plotted on top are spectrums at 500 MHz and 1 GHz outputs. The tones do indeed follow this relationship. A 20 dB decrease is observed from the 5 GHz clock to the 500 MHz DAC output, and a 6 dB increase shows up from 500 MHz to the 1 GHz output.

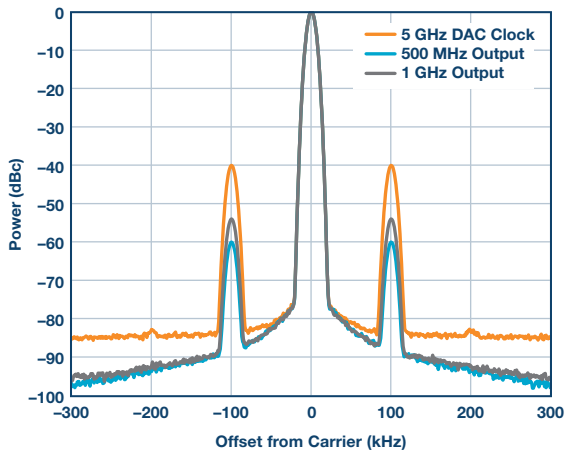


Figure 7. Clock output phase noise with 100 kHz phase modulation.

As nice as a well-controlled experiment is, real noise is of interest. Substituting the generator with the ADF4355 wideband synthesizer, Figure 8 shows the phase noise profile for the new clock source along with corresponding DAC outputs at ½ and ¼ clocking frequency. The noise behavior is preserved with 6 dB decreases each time. It should be noted that the PLL was not optimized for best phase noise. Perceptive readers will notice that some deviation from expectation occurs at small offsets, but this is expected due to differing reference sources.

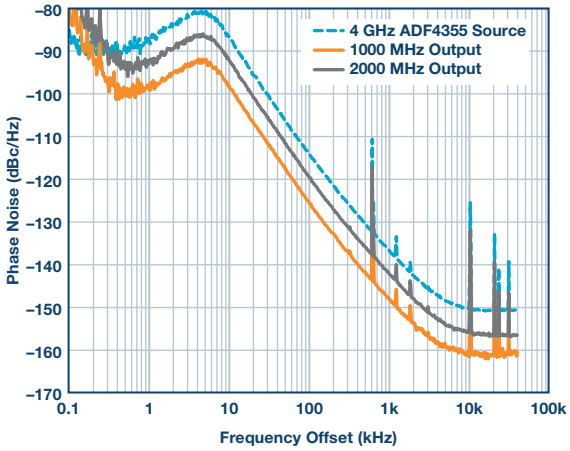


Figure 8. DAC output phase noise with wideband synthesizer clock source.

Another aspect to explore is the lack of dependency between input power and noise. It is only the difference between the noise power to the carrier that matters. This means that straight amplification of the clock yields no benefit. Figure 9 shows that this is indeed the case. The only change is a slight increase in the noise floor that is attributed to the signal generator. Now, this observation is only valid within reason; at a certain point, clocking will become so weak that other contributions such as clock receiver noise will start dominating.

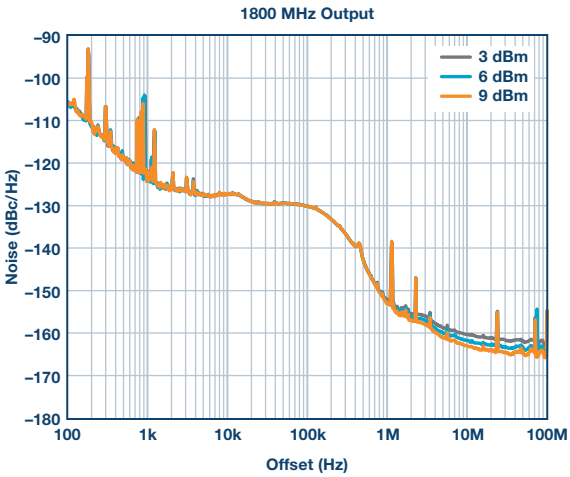


Figure 9. Phase noise vs. input power.

Lastly, the new sampling scheme, 2× NRZ, should be briefly mentioned. The AD9164 DAC series of parts introduces this new sampling mode that allows new sampled data on both the rising and falling edges of the clock. However, with these changes the phase noise characteristic stays the same. Figure 10 compares the original NRZ mode with this new mode. The curves show identical phase noise, though some noise floor rise is visible. This conclusion does assume the noise characteristics are the same on both the rising and falling edge, which is the case for most oscillators.

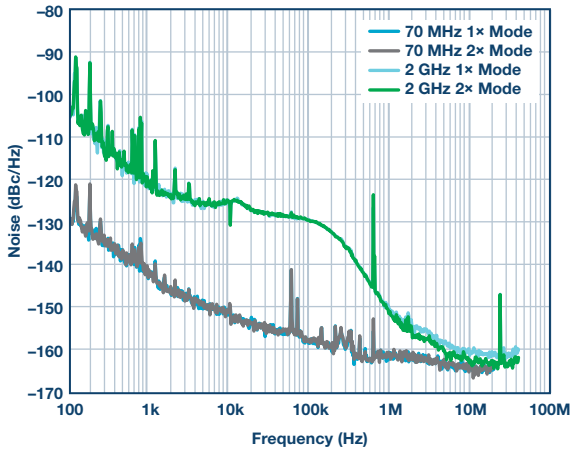


Figure 10. Phase noise and 2× NRZ.

Power Supply

The next possible point of entry for noise is through power. All of the circuitry on a die must be powered one way or another and this gives noise plenty of ways to propagate through to the output. The precise mechanism is circuit dependent but a few possibilities are highlighted below. The DAC output is typically composed of current sources with MOS switches to direct the current through either the positive or negative pin (Figure 11). As evidenced, the current source gets its power from an external supply and any noise will reflect as current fluctuations. The noise can pass through the switches to the output but that would only explain a direct coupling to baseband. To contribute to phase noise, this noise must be mixed up to the carrier frequency. This process is done by way of the switching MOSFETs, which act as a balanced mixer. Another path for noise is through the pull-up inductors. They set the dc bias from a rail and any noise present here flows to the transistors. Such fluctuations modify their operating conditions, such as source to drain voltage and current source load, leading to changes in current flow that once again gets mixed up to the RF signal. In general, any circuit is a vector for power supply noise to show up as phase noise, if switching is capable of mixing it up to the signal at hand.

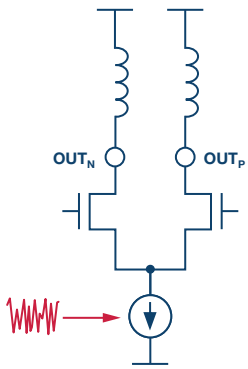


Figure 11. DAC current source.

With all of this circuitry and mixing phenomena, it gets unwieldy rather quickly to model all of this behavior. Instead, characterization of other analog blocks brings insight. In regulators, op amps, and other ICs, a power supply rejection ratio is specified. Supply rejection quantifies a load's sensitivity to supply changes and can be used here for phase noise analysis. Instead of rejection, however, a modulation ratio is used: power supply modulation ratio (PSMR). The traditional PSRR measure can still be useful in DACs in baseband applications but it is not of interest here. The next step is to obtain the data.

Measuring PSMR requires modulating the supply rail that is under investigation. A typical setup is shown in Figure 12. Supply modulation is obtained through a coupling circuit inserted between the regulator and the load, superimposing a sinusoid signal that is produced by a signal generator. The output of the coupling circuit is monitored with an oscilloscope to find the actual supply modulation. The resulting DAC output is fed to a spectrum analyzer. The PSMR is calculated by a ratio of the ac component of the supply as found from the oscilloscope to the modulated sideband voltage around the carrier.

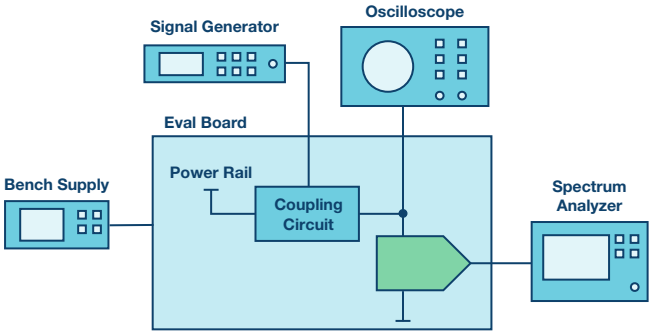


Figure 12. PSMR measurement.

Different coupling schemes are possible. Rob Reeder, Analog Devices applications engineer, provides a rundown of the use of LC circuits to measure PSMR of ADCs in the MS-2210 application note. Other options include a power op amp, transformer, or dedicated modulated power supply. The method used here was the transformer. A high turns ratio is recommended to lower the source impedance of the signal generator. A typical measurement is provided in Figure 14.

Using a 1:100 turns ratio current sense transformer and function generator, the 1.2 V clock supply was modulated at 500 kHz with a resulting peak-to-peak voltage of 38 mV. The DAC was clocked at 5 GSPS. The resulting output incurs sidebands on a full-scale, 1 GHz carrier at –35 dBm. Converting power to a voltage and then taking the ratio with the modulated supply voltage results in a PSMR of –11 dB.

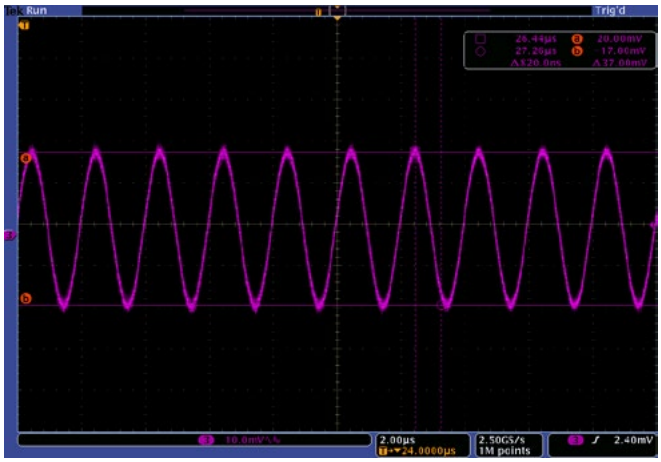


Figure 13. Clock supply modulation.

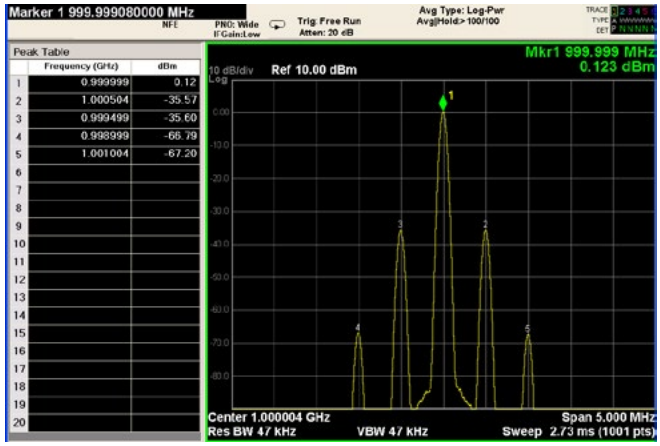


Figure 14. Modulated sidebands.

With a single data point performed, a sweep can be done over multiple frequencies. However, the AD9164 DAC includes a total of eight supplies. One option is to measure all the supplies, but the focus can be limited to the most sensitive supplies: AVDD12, AVDD25, VDDC12, and VNEG12. Some supplies, such as SERDES, aren't relevant for this analysis and therefore not included. Sweeping through multiple frequencies and supplies, the results are summarized in Figure 15.

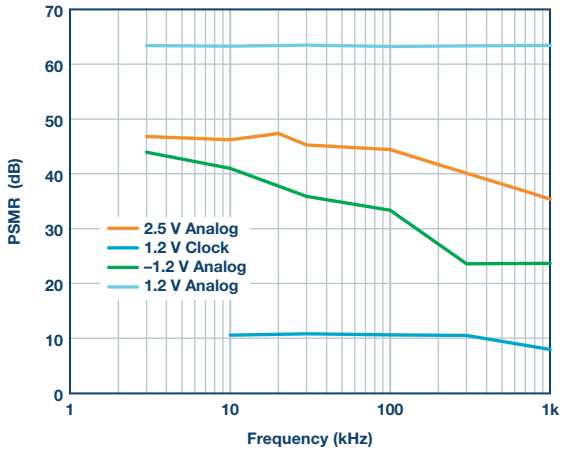


Figure 15. Supply PSMR measured over swept frequencies.

The clock supply is the most sensitive rail. The negative 1.2 V and 2.5 V analog supplies are next and then the 1.2 V analog supply, which is quite insensitive. Whereas the 1.2 V analog supply could, with appropriate consideration, be supplied by a switching regulator, the clock supply is on the complete opposite spectrum: it needs to be supplied by very low noise LDOs to obtain optimum performance.

PSMR could only be measured within a certain frequency range. On the low end, it's limited by weakening magnetic coupling. The selected transformer had a low frequency cutoff in the 10s of kHz. On the high end, decoupling caps lower the load impedance making the supply rail increasingly difficult to drive. Some caps can be removed for testing purposes as long as functionality isn't compromised.

Before using PSMR, a few aspects should be noted. Unlike PSRR, PSMR depends on the waveform power or, in the case of DACs, the digital back-off. The lower the waveform, the lower the sideband becomes, in a 1:1 ratio. However, backing off does not gain the designer anything as the sideband is constant relative to the carrier. The second aspect is the dependency over carrier frequency. A sweep of the carrier indicates linear degradation at higher bands at various rates. Interestingly, the more sensitive the rail, the steeper the slope. For instance, the clock supply is sloped at –6.4 dB/octave, while the negative analog supply is at –4.5 dB/octave. The sampling rate also influences the PSMR. Finally, PSMR only provides an upper limit on phase noise contribution as it is not differentiated from amplitude noise that is also produced.

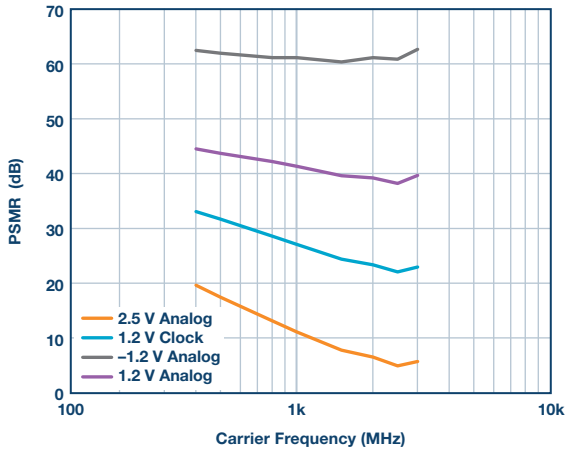


Figure 16. Supply PSMR over signal frequency.

Given these diverse noise requirements, it is helpful to look at a few powering options. The LDO is the tried and true regulator—especially for achieving utmost noise performance. However, not any LDO will do. The 15002C curve in Figure 17 demonstrates the phase noise of the initial AD9162 DAC evaluation board. The DAC output was set to 3.6 GHz, with the DAC being clocked at 4 GHz from the Wenzel source. The phase noise plateau between 1 kHz and 100 kHz was suspected to be dominated by the clock power supply noise: the ADP1740 LDO. Using this LDO's spectral noise density plots and the DAC PSMR measurements in Figure 16, the contribution can be calculated and plotted as shown in Figure 17, as well. Even though it does not precisely line up because of extrapolation, the calculated points line up reasonably to the measured noise, solidifying the clock supply effects on noise. In a redesign of the power solution, this LDO was replaced with the lower noise ADP1761. Noise was lowered by as much as 10 dB at certain offsets, approaching the clock contribution (15002D).

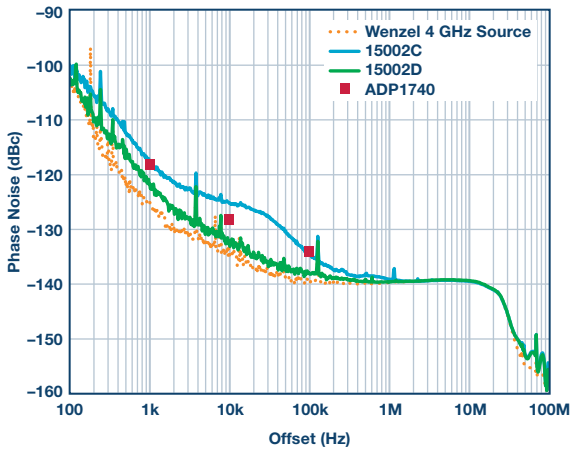


Figure 17. AD9162 evaluation board noises.

Not only does noise vary widely over various regulators, but it can also be influenced by output capacitors, output voltage, and load. Careful consideration of these factors should be taken into account, especially on sensitive rails. On the other hand, depending on the overall system requirements, LDOs are not necessarily required.

Switching regulators may provide power with appropriate LC filtering, simplifying the power solution. As with LDOs, start from the regulator NSD and design accordingly. However with LC filters, attention should be given to the series resonance. Not only can transients become unwieldly, but voltage gain in the vicinity of the resonance frequency can occur, increasing the power rail noise along with phase noise. The resonance can be tamed by de-Q-ing the circuit—that is, adding lossy elements to the circuit. The following figures show an example from another design featuring the AD9162 DAC.

In this design, the clock supply was also powered off an ADP1740 LDO but an LC filter followed it. The schematic shows the filter under consideration with an RL model for an inductor and an RC model for the main filter capacitor (C1+R1). The filter response is shown with a characteristic resonance in red on Figure 20. Not surprisingly, telltale signs of this filter show up in the phase noise response: the blue curve of Figure 21. A plateauing of the noise around 100 kHz and a steep drop off afterward from the filtering action. Fortunately, the LC filter peaking is not severe enough to cause a distinct peak, but the filter can be improved nonetheless. One scheme, which is employed here is to add a second, larger cap with an appropriate series resistance to dissipate the energy. A series circuit of a 22 µF capacitor and a 100 mΩ resistor is shown that significantly dampers down the response (the blue curve). The end result is a phase noise improvement around this frequency offset: the yellow curve in Figure 21.

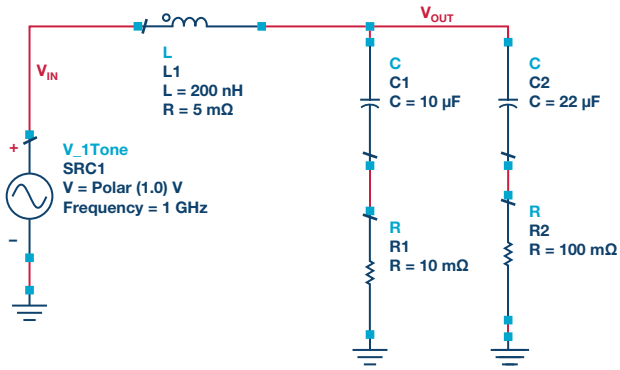


Figure 18. LC filter and de-Q network.

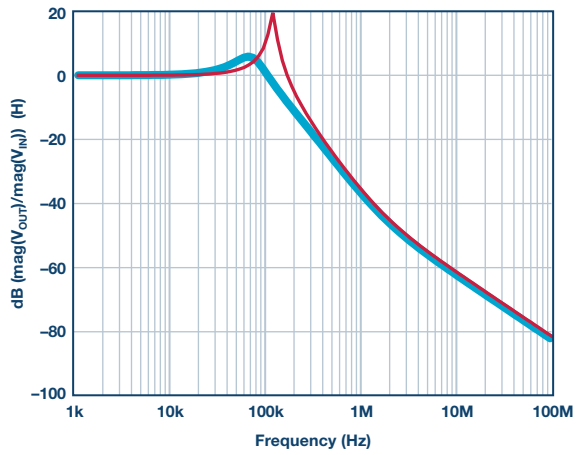


Figure 19. LC filter response.

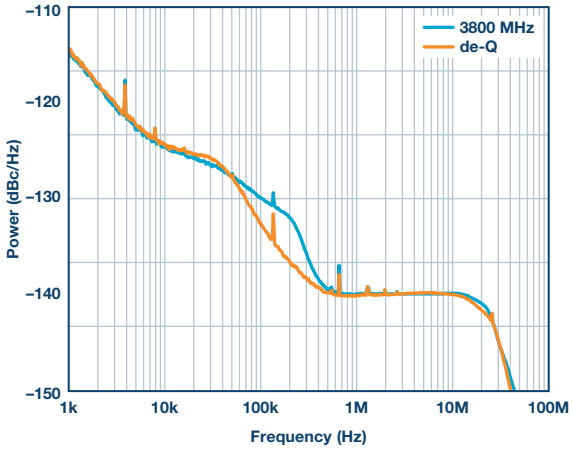


Figure 20. Phase noise response.

The final noise source to analyze is the phase noise of the part itself. The AD9164 DAC series of parts feature very low phase noise, which is challenging to quantify. By removing all the expected noise sources, the residual noise is from the DAC, as shown in Figure 22. The simulated phase noise is also plotted and aligns well with measurement. Clock phase noise still dominates in certain regions.

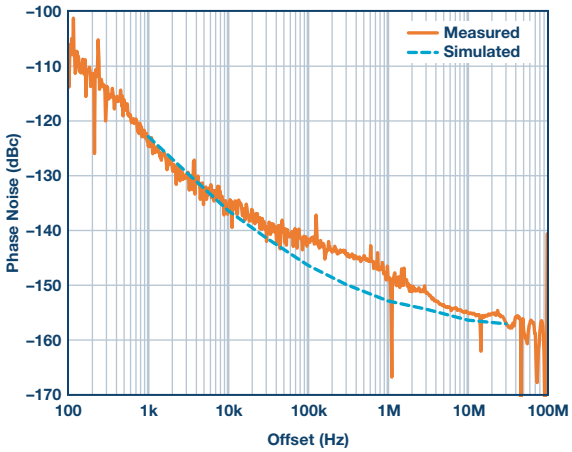


Figure 21. AD9162 phase noise.

Conclusion

Confronted with all of the previously discussed noise sources, the designer may become overwhelmed. The temptation is to follow a recommended solution; however, this approach will always be suboptimal for any particular design requirements. Analogous to RF signal chains and precision error budgets, a phase noise budget can be used in the design process. Using clock source phase noise, the PSMR results for each supply rail, LDO noise characteristics, and the DAC setup, the noise contributions from each source can be calculated and optimized. An example budget is shown in Figure 22. With all sources properly considered, phase noise can be analyzed and managed, and the signal chain designed right the first time.

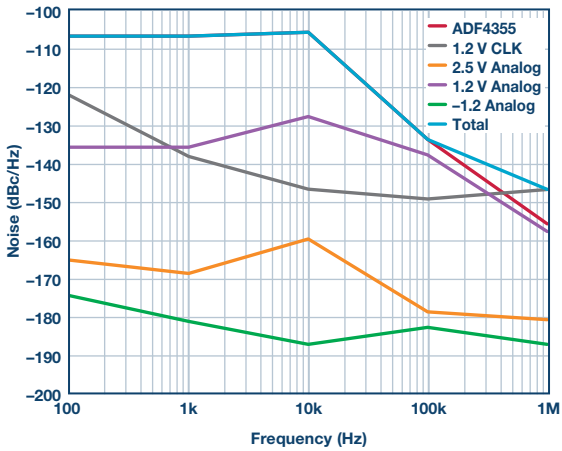


Figure 22. Example phase noise budget.

References

Brannon, Brad. Application Note AN-756, *Sampled Systems and the Effects of Clock Phase Noise and Jitter*. Analog Devices, Inc., 2004.

Reeder, Rob. "Designing Power Supplies for High Speed ADC." Analog Devices, Inc., February 2012.

Jarrah Bergeron [jarrah.bergeron@analog.com] has been an applications engineer in the High Speed Converters Group at Analog Devices since 2014. He has worked on a diverse range of projects ranging from high power microwave systems to nanoscale particle sensing, but he also makes sure to enjoy the outdoors, from rock climbing to snowboarding. Jarrah graduated from the University of Victoria with a bachelor's degree in electrical engineering.

Jarrah Bergeron

Rarely Asked Questions—139
The Case of the Misguided Gyro

By **Ian Beavers**

Share on

Question:

I have heard that it is possible for a gyroscope heading to accumulate drift error over time. Can this happen in my IMU?

Answer:

A MEMS gyroscope, which measures angular rate, has several internal contributors to error, with bias instability as one component. However, an inertial measurement unit (IMU) has several advantages over a discrete component that provides it enhanced performance. An IMU with six degrees of freedom is composed of multiple inertial MEMS sensors that are temperature compensated and calibrated to align on orthogonal axes. The internal 3-axis gyroscope capability measures rotation about a known point while a 3-axis accelerometer measures displacement. A postprocessing step using a digital signal processor or microcontroller provides an internal means for sensor fusion.

Gyroscopes are subject to bias instabilities, in which the initial zero reading of the gyroscope will cause drift over time due to integration of inherent imperfections and noise within the device. Bias repeatability can be calibrated across the known temperature range of the IMU. However, integrating constant bias instability will cause angular error. These errors will accumulate as gyroscope-based rotation or angle estimates drift over the long-term. The undesirable result of drift is that the error of a computed heading increases continuously unabated. Accelerometers, conversely, are sensitive to vibration and other nongravity accelerations.

The gyroscope drift is mainly due to the integration of two components: a slow changing, near-dc variable called bias instability and a higher frequency noise variable called angular random walk (ARW). These parameters are measured in degrees of rotation per unit of time. The yaw axis is most sensitive to this drift. A good portion of the pitch (attitude) and roll axis gyroscope drift can be removed within an IMU through the use of accelerometer feedback to monitor position relative to gravity. Filtering the gyroscope output within an IMU using a low-pass or Kalman filter is also a widely used method to cancel a portion of the drift error.

Ideally, two references are needed to correct for gyroscope drift on all axes. An IMU with nine degrees of freedom typically provides additional magnetometer sensors—about 3 axes. A magnetometer senses field strength relative to the Earth's magnetic north. These sensors can be used together with accelerometer data, as another external reference, to mitigate the impact of gyroscope drift error on the yaw axis. However, design of proper spatial magnetometers can be a less reliable vector than an accelerometer, as there are many things that create magnetic fields within the same order of magnitude as that of the Earth.

Another one of the more effective methods for canceling long-term drift is to implement a zero angular velocity update to the gyroscope. Any time the device is known to be completely stationary, the gyroscope offset can be nulled to zero for that respective axis. This opportunity can vary wildly depending on the application. But any instance that the system is at a recurring resting state can be used for nulling, such as an idling car, a stationary autonomous robot, or the time between human foot steps.

Of course, using a state-of-the-art IMU that has minimal bias instability within the design at the onset may offer the most immediate impact on gyroscope drift. The constant bias error of a gyroscope can be measured by taking the average of the output over a long period of time while the device is not rotating. An IMU Allan variance plot shows the gyroscope drift in rotational degrees per hour vs. the integration time, Tau. It is normally plotted on a log-log scale. The ADIS16490 is the latest product within ADI's portfolio of high performance, tactical grade IMUs. The ADIS16490 has an in-run bias stability of merely 1.8° per hour. This is reflected on the ADIS16490 Allan variance plot in Figure 1, where 1.8° of error is seen at the one hour point (3600 seconds).

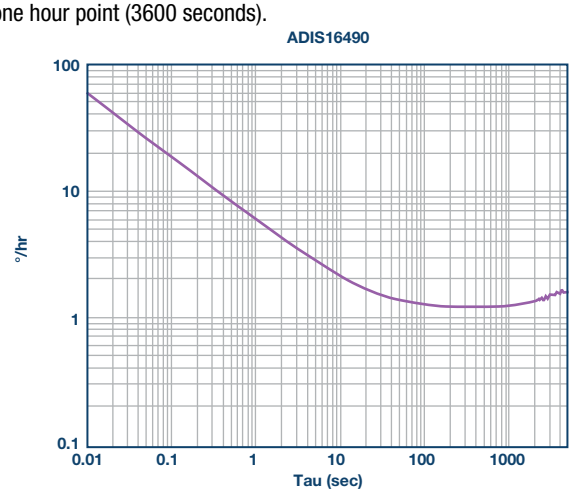


Figure 1. ADIS16490 gyroscope root Allan variance.

Ian Beavers [ian.beavers@analog.com] is a product engineering manager for the Automation Energy and Sensors Team at Analog Devices (Greensboro, NC). He has worked for the company since 1999. Ian has over 19 years of experience in the semiconductor industry. Ian earned a bachelor's degree in electrical engineering from North Carolina State University and an M.B.A. from the University of North Carolina at Greensboro.

Ian Beavers

Also by this Author:
Going Full Tilt
Volume 50, Number 3

Notes

Notes

Notes

Stay Connected with *Analog Dialogue*

Your Engineering Resource for Innovative Design

Be the first to receive the latest technology, applications, and systems-level insights needed to solve real-world design challenges by subscribing to *Analog Dialogue*. And be sure to tell a friend! *analogdialogue.com*

Like *Analog Dialogue* to discover the latest articles and author insights. *facebook.com/analogdialogue*

AnalogDialogue

EngineerZone® Support Community

Engage with the ADI technology experts in the community to ask your tough design questions, browse our rich knowledge base, or hear from our authors in the *Analog Dialogue* blog. *ez.analog.com*, *ezchina.analog.com* (China)

EngineerZone™
SUPPORT COMMUNITY

Facebook
facebook.com/AnalogDevicesInc

@ADI_News
twitter.com/ADI_News

LinkedIn
linkedin.com/company/analog-devices

Sina Weibo (China)
weibo.com/analogdevices

WeChat (China)

Analog Devices, Inc.
Worldwide Headquarters

Analog Devices, Inc.
One Technology Way
P.O. Box 9106
Norwood, MA 02062-9106
U.S.A.
Tel: 781.329.4700
(800.262.5643, U.S.A. only)
Fax: 781.461.3113

Analog Devices, Inc.
Europe Headquarters

Analog Devices GmbH
Otl-Aicher-Str. 60-64
80807 München
Germany
Tel: 49.89.76903.0
Fax: 49.89.76903.157

Analog Devices, Inc.
Japan Headquarters

Analog Devices, KK
New Pier Takeshiba
South Tower Building
1-16-1 Kaigan, Minato-ku,
Tokyo, 105-6891
Japan
Tel: 813.5402.8200
Fax: 813.5402.1064

Analog Devices, Inc.
Asia Pacific Headquarters

Analog Devices
5F, Sandhill Plaza
2290 Zuchongzhi Road
Zhangjiang Hi-Tech Park
Pudong New District
Shanghai, China 201203
Tel: 86.21.2320.8000
Fax: 86.21.2320.8222

M02000511-0-3/17

analog.com

ANALOG DEVICES
AHEAD OF WHAT'S POSSIBLE™