Acquiring an Analog Signal:

Bandwidth, Nyquist Sampling Theorem, and Aliasing

Overview

Learn about acquiring an analog signal, including topics such as bandwidth, amplitude error, rise time, sample rate, the Nyquist Sampling Theorem, aliasing, and resolution. This tutorial is part of the Instrument Fundamentals series.

Contents

What Is a Digitizer?

Scientists and engineers often use a digitizer to capture analog data in the real world and convert it into digital signals for analysis. A digitizer is any device used to convert analog signals into digital signals. One of the most common digitizers is a cell phone, which converts a voice, an analog signal, into a digital signal to send to another phone. However, in test and measurement applications, a digitizer most often refers to an oscilloscope or a digital multimeter (DMM). This article focuses on oscilloscopes, but most topics are also applicable to other digitizers.

Regardless of the type, the digitizer is vital for the system to accurately reconstruct a waveform. To ensure you select the correct oscilloscope for your application, consider the bandwidth, sampling rate, and resolution of the oscilloscope.

Bandwidth

The front end of an oscilloscope consists of two components: an analog input path and an analog-to-digital converter (ADC). The analog input path attenuates, amplifies, filters, and/or couples the signal to optimize it in preparation for digitization by the ADC. The ADC samples the conditioned waveform and converts the analog input signal to digital values that represent the analog input waveform. The frequency response of the input path causes an inherent loss of amplitude and phase information

Bandwidth describes the analog front end's ability to get a signal from the outside world to the ADC with minimal amplitude loss—from the tip of the probe or test fixture to the input of the ADC. In other words, the bandwidth describes the range of frequencies an oscilloscope can accurately measure.

It is defined as the frequency at which a sinusoidal input signal is attenuated to 70.7 percent of its original amplitude, which is also known as the -3 dB point. Figures 2 and 3 show the typical input response for a 100 MHz oscilloscope.

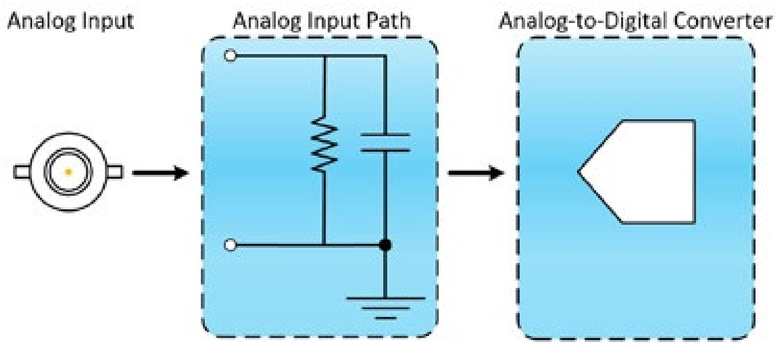


Figure 1. Bandwidth describes the frequency range in which the input signal can pass through the oscilloscope front end, which is made of two components: an analog input path and an ADC.

Bandwidth describes the analog front end's ability to get a signal from the outside world to the ADC with minimal amplitude loss—from the tip of the probe or test fixture to the input of the ADC. In other words, the bandwidth describes the range of frequencies an oscilloscope can accurately measure.

It is defined as the frequency at which a sinusoidal input signal is attenuated to 70.7 percent of its original amplitude, which is also known as the -3 dB point. Figures 2 and 3 show the typical input response for a 100 MHz oscilloscope.

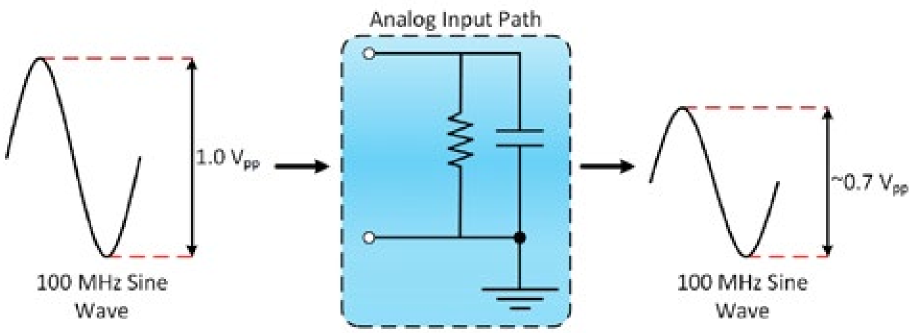


Figure 2. Bandwidth is when the input signal is attenuated to 70.7 percent of its original amplitude.

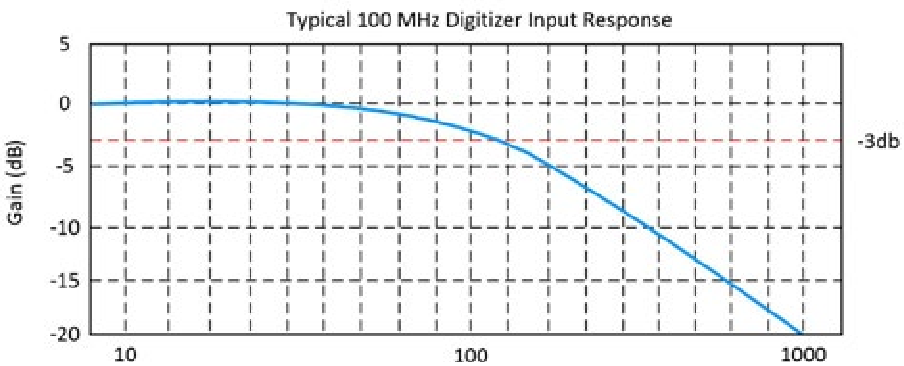


Figure 3. This graph indicates that at 100 MHz, the input signal hits the -3dB point.

Bandwidth is measured between the lower and upper frequency points where the signal amplitude falls to -3 dB below the passband frequency. This sounds complicated, but when you break it down it is actually relatively easy

First, you want to calculate your -3 dB value.

$$-3\ dB = 20 \log \frac{V_{out,pp}}{V_{in,pp}}$$

Equation 1. Calculating the -3 dB Point

$V_{in,pp}$ is the peak-to-peak voltage of the input signal and $V_{out,pp}$ is the peak-to-peak voltage of the output signal. For example, if you input a 1 V sine wave, the output voltage can be calculated as $-3 = 20 \log \frac{V_{out,pp}}{1}$ so $V_{out,pp} \approx 0.7\ V$.

Because the input signal is a sine wave, there are two frequencies at which the output signal hits this voltage; these are called the corner frequencies f_1 and f_2. These two frequencies go by many different names such as corner frequency, cut-off frequency, crossover frequency, half-power frequency, 3 dB frequency, and break frequency. However, all these terms refer to the same values. The center frequency, f_0, of the signal is the geometric mean of f_1 and f_2.

$$f_o = \sqrt{f_1 f_2}$$

Equation 2. Calculating the Center Frequency

You can calculate the bandwidth (BW) by subtracting the two corner frequencies.

$$BW = f_2 - f_1$$

Equation 3. Calculating the Bandwidth

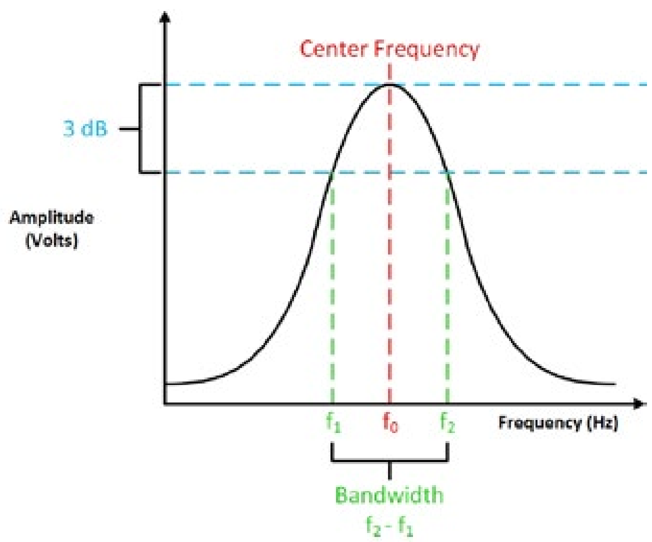


Figure 4. The bandwidth, the corner frequency, the center frequency, and the 3 dB point are all connected.

a. Calculating Amplitude Error

Another equation that is often helpful is for amplitude error.

$$Amplitude\ Error = 100\left(1 - \frac{R}{\sqrt{1+R^2}}\right), \text{where } R = \frac{BW}{f_{in}}$$

Equation 4. Calculating the Amplitude Error

The amplitude error is expressed as a percentage, and R is the ratio of the oscilloscope's bandwidth to the input signal frequency (f_{in}).

Using the example above, you have a 100 MHz oscilloscope with a 100 MHz sine wave input signal at 1 V, and BW = 100 MHz and f_{in} = 100 MHz. This means R = 1. Then you just have to solve the equation:

$$Amplitude\ Error = 100\left(1 - \frac{1}{\sqrt{2}}\right) = 29.3$$

The amplitude error is 29.3 percent. You can then determine the output voltage for the 1 V signal:

$$V_{out,pp} = 1 - \frac{V_{in,pp} * Amplitude\ Error}{100} = 0.7V$$

It is recommended that the bandwidth of your oscilloscope be three to five times the highest frequency component of interest in the measured signal to capture the signal with minimal amplitude error. For instance, for the 1 V sine wave at 100 MHz, you should use an oscilloscope with 300 MHz to 500 MHz bandwidth. The amplitude error of a 100 MHz signal at these bandwidths are:

$$Amplitude\ Error\ @300\ MHz = 100\left(1 - \frac{3}{\sqrt{1+3^2}}\right) \approx 5\%$$

$$Amplitude\ Error\ @500\ MHz = 100\left(1 - \frac{5}{\sqrt{1+25^2}}\right) \approx 2\%$$

b. Calculating Rise Time

An oscilloscope must have the appropriate bandwidth to accurately measure the signal, but it must also have sufficient rise time to accurately capture the details of rapid transitions. This is most applicable if measuring digital signals such as pulses and steps. The rise time of an input signal is the time for a signal to transition from 10 percent to 90 percent of the maximum signal amplitude. Some oscilloscopes may use 20 percent to 80 percent, so be sure to check your user manual.

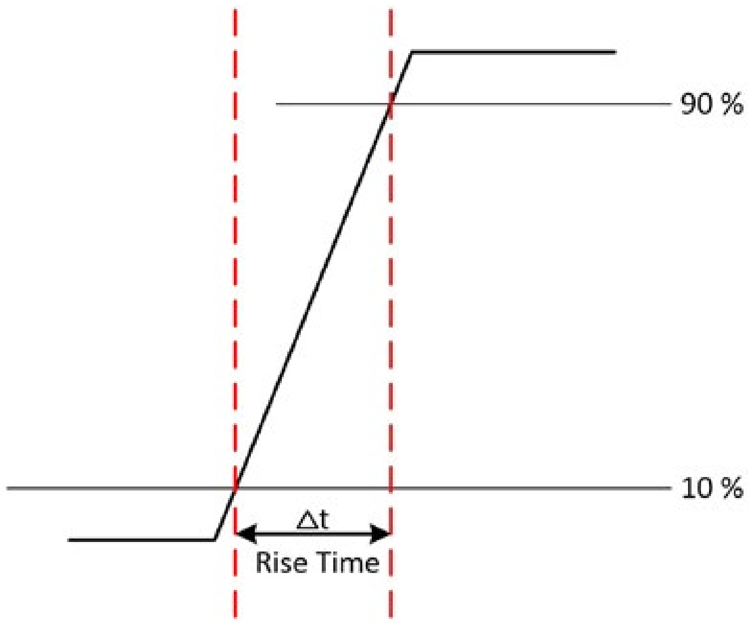


Figure 5. The rise time of an input signal is the time for a signal to transition from 10 percent to 90 percent of the maximum signal amplitude.

The rise time (Tr) can be calculated as follows:

$$T_r = \frac{k}{BW}$$

Equation 5. Calculating the Rise Time

The constant k is dependent on the oscilloscope. Most oscilloscopes with a bandwidth less than 1 GHz typically have k = 0.35, while oscilloscopes with a bandwidth greater than 1 GHz usually have a value of k between 0.4 and 0.45.

The theoretical rise time measured T_{r_m} be calculated from the rise time of the oscilloscope T_{r_o} and the actual rise time of the input signal T_{r_s}.

$$T_{r_m} = \sqrt{{T_{r_O}}^2 + {T_{r_s}}^2}$$

Equation 6. Calculating the Theoretical Rise Time Measured

It is recommended that the rise time of the oscilloscope be one-third to one-fifth the rise time of the measured signal to capture the signal with minimal rise time error.

Sample Rate

The sample rate, also referred to as sampling rate, is not directly related to the bandwidth specification. Sample rate is the frequency at which the ADC converts the analog input waveform to digital data. The oscilloscope samples the signal after any attenuation, gain, and/or filtering has been applied to the analog input path and converts the resulting waveform to digital representation. It does so in snapshots, similar to the frames of a movie. The faster the oscilloscope samples, the greater the resolution and detail that can be seen in the waveform.

a. Nyquist Sampling Theorem

The Nyquist Sampling Theorem explains the relationship between the sample rate and the frequency of the measured signal. It states that the sample rate f_S must be greater than twice the highest frequency component of interest in the measured signal. This frequency is often referred to as the Nyquist frequency, f_N.

$$f_s > 2 * f_N$$

Equation 7. The sample rate should be greater than twice the Nyquist frequency.

To understand why, take a look at a sine wave measured at different rates. In case A, the sine wave of frequency f is sampled at that same frequency. Those samples are marked on the original signal on the left and, when constructed on the right, the signal incorrectly appears as a constant DC voltage. In case B, the sample rate is twice the frequency of the signal. It now appears as a triangle waveform. In this case, f is equal to the Nyquist frequency, which is the highest frequency component allowed to avoid aliasing for a given sampling frequency. In case C, the sampling rate is at 4f/3. The Nyquist frequency in this case is:

$$f_N = \frac{f_s}{2} = \frac{{}^{4f}/_{3}}{2} = \frac{2f}{3}$$

Because f is larger than the Nyquist frequency ($\frac{4f}{3} > \frac{2f}{3}$), this sample rate reproduces an alias waveform of incorrect frequency and shape.

Thus, to accurately reconstruct the waveform, the sample rate fs must be greater than twice the highest frequency component of interest in the measured signal. Usually, you want to sample around five times greater than the signal frequency.

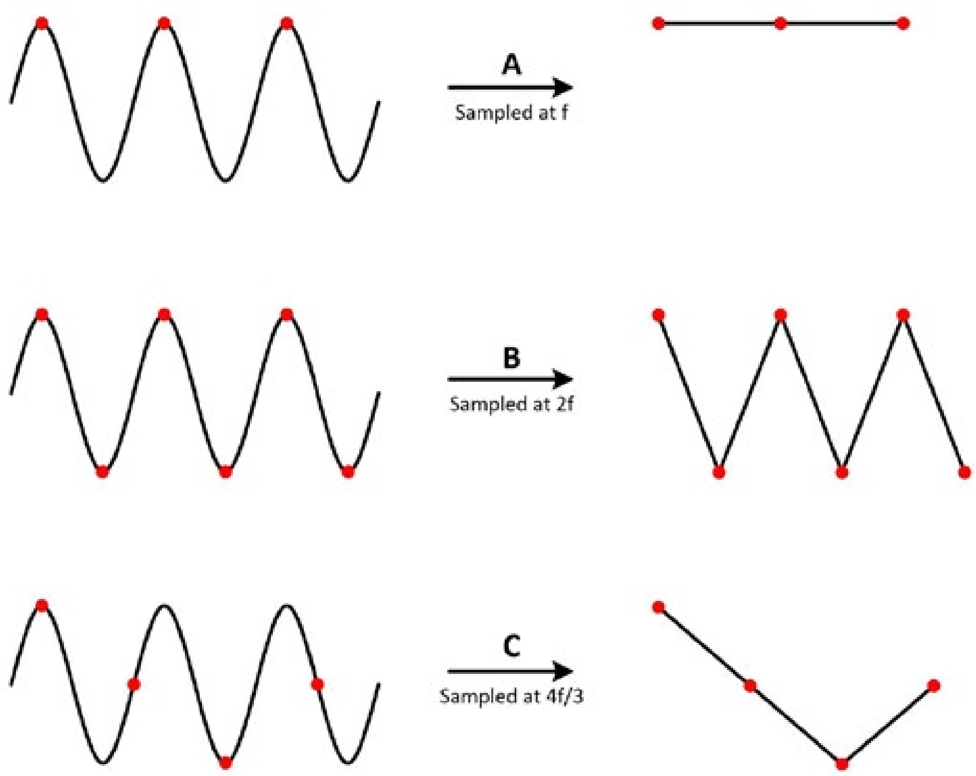


Figure 6. Too low a sample rate can cause inaccurate reconstruction of the waveform.

b. Aliasing

If you need to sample at a certain rate to avoid aliasing, then what exactly is aliasing? If a signal is sampled at a sampling rate smaller than twice the Nyquist frequency, false lower frequency components appear in the sampled data. This phenomenon is referred to as aliasing. The following figure shows an 800 kHz sine wave sampled at 1 MS/s. The dotted line indicates the aliased signal recorded at that sample rate. The 800 kHz frequency aliases back in the passband, falsely appearing as a 200 kHz sine wave.

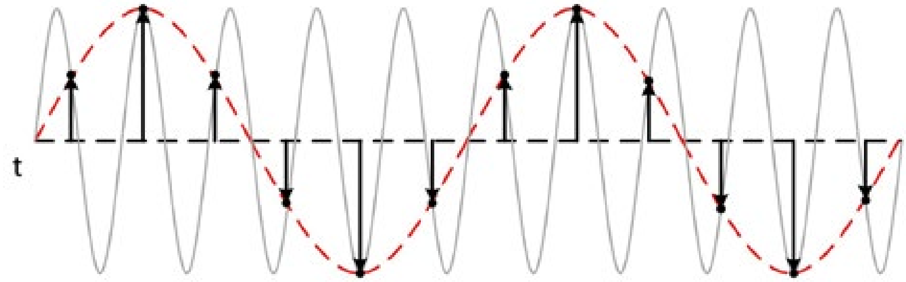


Figure 7. Aliasing occurs when a sample rate is too low and reproduces an inaccurate waveform representation.

The alias frequency fa can be calculated to determine how an input signal at a frequency over the Nyquist frequency appears. It is the absolute value of the closest integer multiple of the sample frequency minus the frequency of the input signal.

$$f_a = |closest\ integermultiple\ of\ f_s * f_s - f|$$

Equation 8. Calculating the Alias Frequency

For example, consider a signal with a sample frequency of 100 Hz, and the input signal contains the following frequencies: 25 Hz, 70 Hz, 160 Hz, and 510 Hz. Frequencies below the Nyquist frequency of 50 Hz are sampled correctly; those over 50 Hz appear as alias.

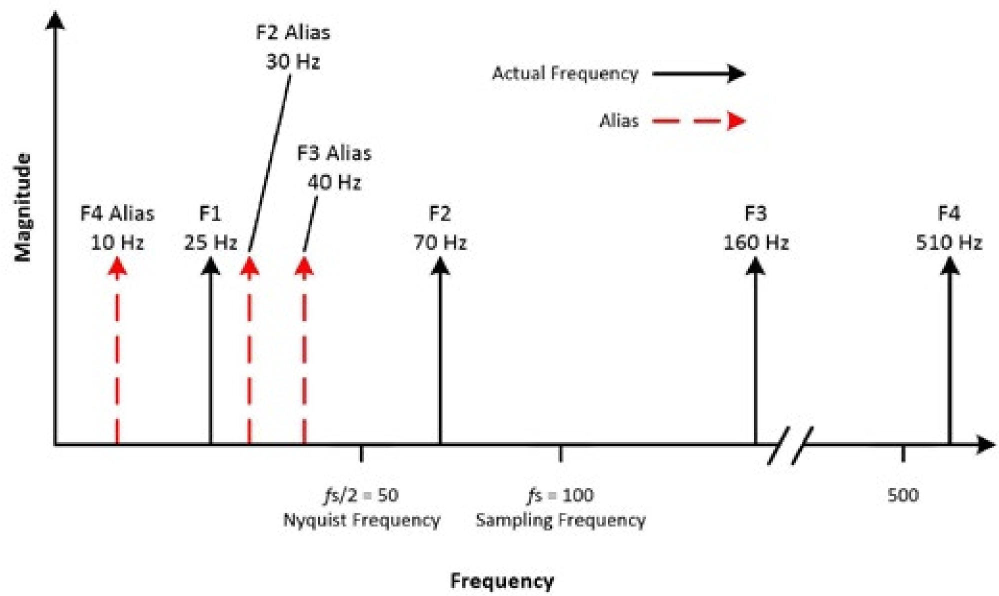


Figure 8. Different frequency values are measured, some of which are alias frequencies and some of which are actual frequencies from the waveform.

Here are the calculations for the alias frequencies:

$$Alias\ F1 = |100 - 70| = 30\ Hz$$

$$Alias\ F2 = |2 * 100 - 160| = 40\ Hz$$

$$Alias\ F3 = |5 * 100 - 160| = 40\ Hz$$

In addition to increasing the sample rate, aliasing can also be prevented by using an antialiasing filter. This is a lowpass filter that attenuates any frequencies in the input signal that are greater than the Nyquist frequency, and must be introduced before the ADC to restrict the bandwidth of the input signal to meet the sampling criteria. Analog input channels can have both analog and digital filters implemented in hardware to assist with aliasing prevention.

Resolution

Another factor to consider when selecting an oscilloscope for an application is the resolution. Bits of resolution refers to the number of unique vertical levels that an oscilloscope can use to represent a signal. One way to understand the concept of resolution is by comparison with a yardstick. Divide a meter yardstick into millimeters; what is the resolution? The smallest tick on the yardstick is the resolution–or 1 out of 1,000.

The resolution of an ADC is a function of how many parts the maximum signal can be divided into. The amplitude resolution is limited by the number of discrete output levels an ADC has. A binary code represents each division; as such, the number of levels can be calculated as follows:

$$\#\ of\ levels = 2^{Resolution}$$

Equation 9. Calculating the Discrete Output Levels of an ADC

For example, a 3-bit oscilloscope has 23 or eight levels. A 16-bit oscilloscope on the other hand has 216 or 65,536 levels. The minimum detectable voltage change or code width can be calculated as follows:

$$Code\ width = \frac{Device\ Input\ Range}{2^{Resolution}}$$

Equation 9. Calculating the Discrete Output Levels of an ADC

The code width is also referred to as the least significant bit (LSB). If the device input range is 0 to 10 V, then a 3-bit oscilloscope has a code width of 10/8 = 1.25 V while a 16-bit oscilloscope has a code width of 10/65,536 = 305 µV. This can mean a big difference in how the signal is displayed.

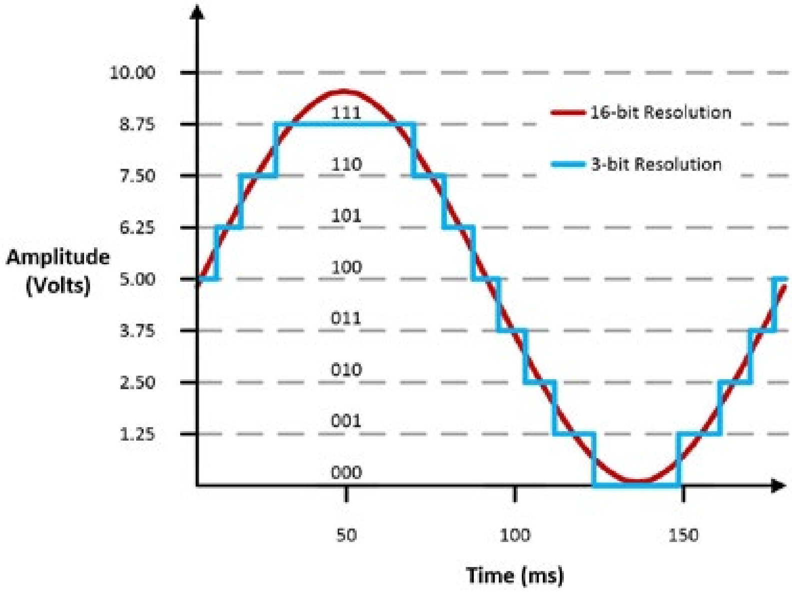


Figure 9. Difference of a Waveform Between 16 Bits and 3 Bits of Resolution

The resolution you need depends on your application; the higher the resolution, the more the oscilloscope costs. Keep in mind that an oscilloscope with high resolution doesn't necessarily mean that it has high accuracy. However, the achievable accuracy of an instrument is limited by the resolution. Resolution limits the precision of a measurement; the higher the resolution (number of bits), the more precise the measurement.

Some oscilloscopes use a method called dithering to help smooth out signals to get the appearance of a higher resolution. Dithering involves the deliberate addition of noise to the input signal. It helps by smearing out the little differences in amplitude resolution. The key is to add random noise in a way that makes the signal bounce back and forth between successive levels. Of course, this in itself just makes the signal noisier. But, the signal smoothes out by averaging this noise digitally once the signal is acquired

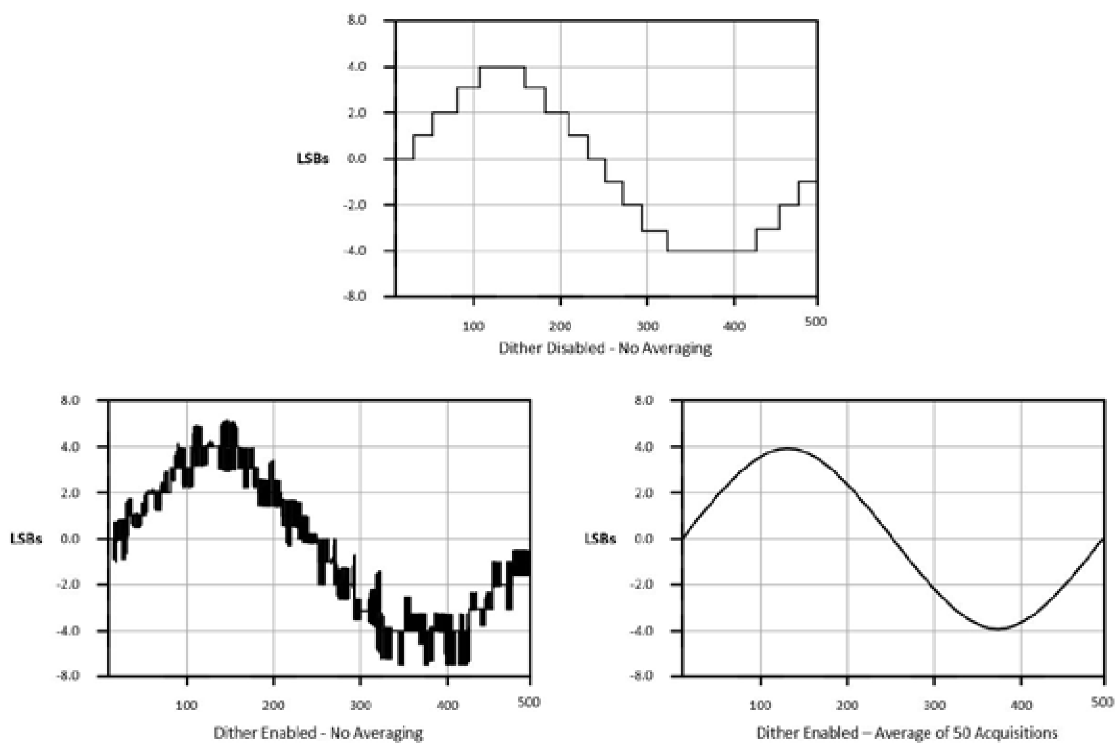


Figure 10. Dithering can help smooth out a signal.

Summary

- **Bandwidth** describes the range of frequencies an oscilloscope can accurately measure. It is defined as the frequency at which a sinusoidal input signal is attenuated to 70.7 percent of its original amplitude, which is also known as the -3 dB point.
- **Bandwidth** is the difference between the ***corner frequencies***.
- **Amplitude error** is a percentage that is the ratio of the bandwidth to the input signal frequencies that assists with determining the noise in a system.
- It is recommended that the bandwidth of your oscilloscope be **three to five times** the highest frequency component of interest in the measured signal to capture the signal with minimal amplitude error.
- The **rise time** of an input signal is the time for a signal to transition from 10 percent to 90 percent of the maximum signal amplitude.
- It is recommended that the rise time of the oscilloscope be **one-third to one-fifth** the rise time of the measured signal to capture the signal with minimal rise time error.
- **Sample rate** is the frequency at which the ADC converts the analog input waveform to digital data.
- The sample rate should be **at least twice** the highest frequency of interest in the signal, but most of the time should be **around five times greater**.
- **Aliasing** is when false frequency components appear in sampled data.
- Bits of **resolution** refers to the number of unique vertical levels that an oscilloscope can use to represent a signal.
- The higher the resolution of an instrument, the greater the precision.

Analog Sample Quality:
Accuracy, Sensitivity, Precision, and Noise

Overview
Learn about sensitivity, accuracy, precision, and noise in order to understand and improve your measurement sample quality. This tutorial is part of the Instrument Fundamentals series.

Contents

Measurement Sensitivity

When referring sample to quality, you want to evaluate the accuracy and precision of your measurement. However, it is important to understand your oscilloscope's sensitivity first. Sensitivity is the smallest change in an input signal that can cause the measuring device to respond. In other words, if an input signal changes by a certain amount—by a certain sensitivity—then you can see a change in the digital data.

Don't confuse sensitivity with resolution and code width. The resolution defines the code width; this is the discrete level at which the instrument displays values. However, the sensitivity defines the change in voltage needed for the instrument to register a change in value. For example, an instrument with a measurement range of 10 V may be able to detect signals with 1 mV resolution, but the smallest detectable voltage it can measure may be 15 mV. In this case, the instrument has a resolution of 1 mV but a sensitivity of 15 mV.

In some cases, the sensitivity is greater than the code width. At first, this may seem counterintuitive—doesn't this mean that the voltage changes by an amount that can be displayed and yet not be registered? Yes! To understand the benefit, think about a constant DC voltage. Although it would be great if that voltage was really exactly constant with no deviations, there is always some slight variation in a signal, which is represented in Figure 1. The sensitivity is denoted with red lines, and the code width is depicted as well. In this example, because the voltage is never going above the sensitivity level, it is represented by the same digital value—even though it is greater than the code width. This is beneficial in that it doesn't pick up noise and more accurately represents the signal as a constant voltage.

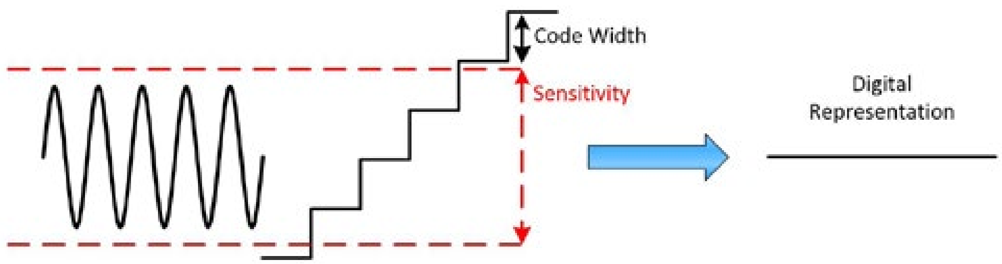


Figure 1. Sensitivity that ~~is~~ greater than the code width can help smooth out a noisy signal.

Once the signal actually starts to rise, it crosses the sensitivity level and then is represented by a different digital value. See Figure 2. Keep in mind that your measurement can never be more accurate than the sensitivity.

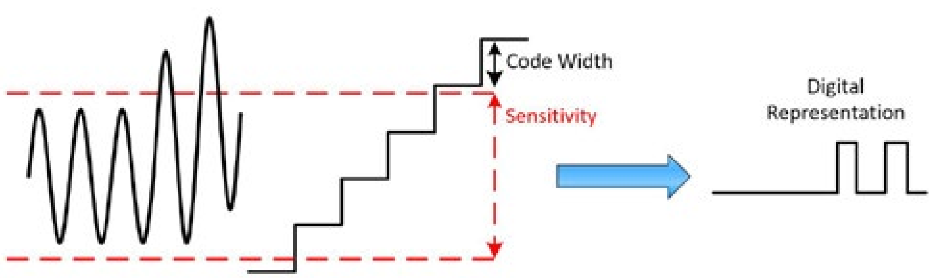


Figure 2. Once the signal crosses the sensitivity level, it is represented by a different digital value.

There is also some ambiguity in how the sensitivity of an instrument is defined. At times, it can be defined as a constant amount as in the example above. In this case, as soon as the input signal crosses the sensitivity level, the signal is represented by a different digital value. However, sometimes it is defined as a change in signal. After the signal has changed by the sensitivity amount specified, it is represented by a different signal. In this case, it doesn't matter the absolute voltage but rather the change in voltage. In addition, some instruments define the sensitivity as around zero.

Not only does the exact definition of the term sensitivity change from company to company, but different products at the same company may use it to mean something slightly different as well. It is important that you check your instrument's specifications to see how sensitivity is defined; if it isn't well documented, contact the company for clarification.

Accuracy

Accuracy is defined as a measure of the capability of the instrument to faithfully indicate the value of the measured signal. This term is not related to resolution; however, the accuracy can never be better than the instrument's resolution.

Depending on the instrument or digitizer, there are different expectations for accuracy. For instance, in general, a digital multimeter (DMM) is expected to have higher accuracy than an oscilloscope. How accuracy is calculated also changes by device, but always check your instrument's specifications to see how your particular instrument calculates accuracy.

a. Accuracy of an Oscilloscope

Oscilloscopes define the accuracy of the horizontal and vertical system separately. The horizontal system refers to the time scale or the X axis; the horizontal system accuracy is the accuracy of the time base. The vertical system is the measured voltage or the Y axis; the vertical system accuracy is the gain and offset accuracy. Typically, the vertical system accuracy is more important than the horizontal.

The vertical accuracy is typically expressed as a percentage of the input signal and a percentage of the full scale. Some specifications break down the input signal into the vertical gain and offset accuracy. Equation 1 shows two different ways you might see the accuracy defined.

$$\begin{aligned} &Vertical\ Accuracy = \\ &\qquad \pm(DC\ Gain\ Accuracy + DC\ Vertical\ Offset\ Accuracy + \%\ Full\ Scale) \\ &Vertical\ Accuracy = \pm(\%\ of\ Input\ at\ \%\ of\ Range) \end{aligned}$$

Equation 1. Calculating the Vertical Accuracy of an Oscilloscope

For example, an oscilloscope can define the vertical accuracy in the following manner:

$$Vertical\ Accuracy = \pm 2\%\ of\ Input, \pm 1\%\ Full\ Scale$$

With a 10 V input signal and using the 20 V range, you can then calculate the accuracy:

$$Vertical\ Accuracy = \pm(2\%\ of\ Input + 1\%\ Full\ Scale)$$
$$Vertical\ Accuracy = \pm(0.02 * 10 + .01 * 20) = \pm 0.4V$$

b. Accuracy of a DMM and Power Supply

DMMs and power supplies usually specify accuracy as a percentage of the reading. Equation 2 shows three different ways of expressing the accuracy of a DMM or power supply.

$$Accuracy = (\%\ of\ Reading) + Offset$$
$$Accuracy = (\%\ of\ Reading) + (\%\ of\ Range)$$
$$Accuracy = \pm(ppm\ of\ Reading + ppm\ of\ Range)$$

Equation 2. Calculating the Vertical Accuracy of a DMM or Power Supply

The term ppm means parts per million. Most specifications also have multiple tables for determining accuracy. The accuracy depends on the type of measurement, the range, and the time since last calibration. Check your specifications to see how accuracy is calculated.

As an example, a DMM is set to the 10 V range and is operating 90 days after calibration at 23 °C ±5 °C, and is expecting a 7 V signal. The accuracy specifications for these conditions state ±(20 ppm of Reading + 6 ppm of Range). You can then calculate the accuracy:

$$Accuracy = \pm(20\ ppm\ of\ Reading + 6\ ppm\ of\ Range)$$
$$Accuracy = \pm(ppm\ of\ 7\ V + 6\ ppm\ of\ 10\ V)$$
$$Accuracy = \pm\left(\frac{20}{1{,}000{,}000} * 7 + \frac{6}{1{,}000{,}000} * 10\right)$$
$$Accuracy = 0.0002\ V = 200\ \mu V$$

In this case, the reading should be within 200 µV of the actual input voltage.

c. Accuracy of a DAQ Device

DAQ cards often define accuracy as the deviation from an ideal transfer function. Equation 3 shows an example of how a DAQ card might specify the accuracy.

$$Accuracy = (Reading * Gain\ Error) + (Reading * Offset\ Error) + (Noise\ Uncertainty)$$

Equation 3. Calculating the Accuracy of a DAQ Device

It then defines the individual terms:

$$\begin{aligned} Gain\ Error = {} & Residual\ AI\ Gain\ Error \\ & + (Gain\ Temperature\ Coefficient \\ & * Temperature\ Change\ From\ Last\ Internal\ Calibration) \\ & + (Reference\ Temperature\ Coefficient \\ & * Temepratrure\ Change\ From\ Last\ External\ Calibration) \end{aligned}$$

$$\begin{aligned} Offset\ Error = {} & Residual\ AI\ Offset\ Error \\ & + (Offset\ Temperature\ Coefficient \\ & * Temperature\ Change\ From\ Last\ Internal\ Calibration) + INL\_Error \end{aligned}$$

The majority of these terms are defined in a table and based on the nominal range. The specifications also define the calculation for noise uncertainty. Noise uncertainty is the uncertainty of the measurement because of the effect of noise in the measurement and is factored into determining the accuracy.

In addition, there may be multiple accuracy tables for your device, depending on if you are looking for the accuracy of analog in or analog out or if a filter is enabled or disabled.

Precision

Accuracy and precision are often used interchangeably, but there is a subtle difference. Precision is defined as a measure of the stability of the instrument and its capability of resulting in the same measurement over and over again for the same input signal. Whereas accuracy refers to how closely a measured value is to the actual value, precision refers to how closely individual, repeated measurements agree with each other.

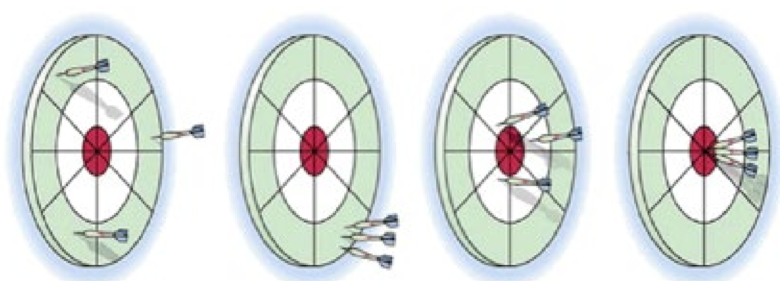


Figure 3. Precision and accuracy are related but not the same.

Precision is most affected by noise and short-term drift on the instrument. The precision of an instrument is often not provided directly, but it must be inferred from other specifications such as the transfer ratio specification, noise, and temperature drift. However, if you have a series of measurements, you can calculate the precision.

$$Precision = 1 - \frac{|Offset\ From\ Input\ Signal|}{|Input\ Signal|}$$

Equation 4. Calculating Precision

For instance, if you are monitoring a constant voltage of 1 V, and you notice that your measured value changes by 20 μV between measurements, then your measurement precision can be calculated as follows:

$$Precision = 1 - \frac{|20\ \mu V|}{|1\ V|} = 1 - \frac{|20|}{|1{,}000{,}000|} = 0.99998$$

Typically, precision is expressed as a percentage. In this example, the precision is 99.998 percent.

Precision is meaningful primarily when relative measurements (relative to a previous reading of the same value), such as device calibration, need to be taken.

Noise and Noise Sources

Don't confuse sensitivity with resolution and code width. The resolution defines the code width; this is the discrete level at which the instrument displays values. However, the sensitivity defines the change in voltage needed for the instrument to register a change in value. For example, an instrument with a measurement range of 10 V may be able to detect signals with 1 mV resolution, but the smallest detectable voltage it can measure may be 15 mV. In this case, the instrument has a resolution of 1 mV but a sensitivity of 15 mV.

a. Thermal Noise

An ideal electronic circuit produces no noise of its own, so the output signal from the ideal circuit contains only the noise that was in the original signal. But real electronic circuits and components do produce a certain level of inherent noise of their own. Even the simple fixed-value resistor is noisy.

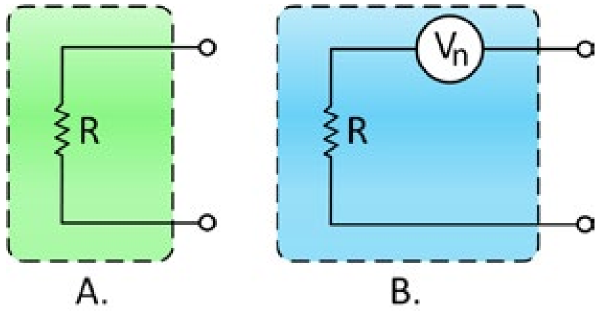


Figure 4. An ideal resistor is reflected in A, but, practically, resistors have internal thermal noise as represented in B.

Figure 4A shows the equivalent circuit for an ideal, noise-free resistor. The inherent noise is represented in Figure 4B by a noise voltage source Vn in series with the ideal, noise-free resistance Ri. At any temperature above absolute zero (0 °K or about -273 °C), electrons in any material are in constant random motion. Because of the inherent randomness of that motion, however, there is no detectable current in any one direction. In other words, electron drift in any single direction is cancelled over short time periods by equal drift in the opposite direction. Electron motions are therefore statistically decorrelated. There is, however, a continuous series of random current pulses generated in the material, and those pulses are seen by the outside world as a noise signal. This signal is called by several names: Johnson noise, thermal agitation noise, or thermal noise. This noise increases with temperature and resistance, but as a square root function. This means you have to quadruple the resistance to double the noise of that resistor.

b. Flicker or $\frac{1}{F}$ Noise

Semiconductor devices tend to have noise that is not flat with frequency. It rises at the low end. This is called $\frac{1}{F}$ noise, pink noise, excess noise, or flicker noise. This type of noise also occurs in many physical systems other than electrical. Examples are proteins, reaction times of cognitive processes, and even earthquake activity. The chart below shows the most likely source of the noise, depending on the frequency the noise occurs for a particular voltage; knowing the cause of the noise goes a long way in reducing the noise.

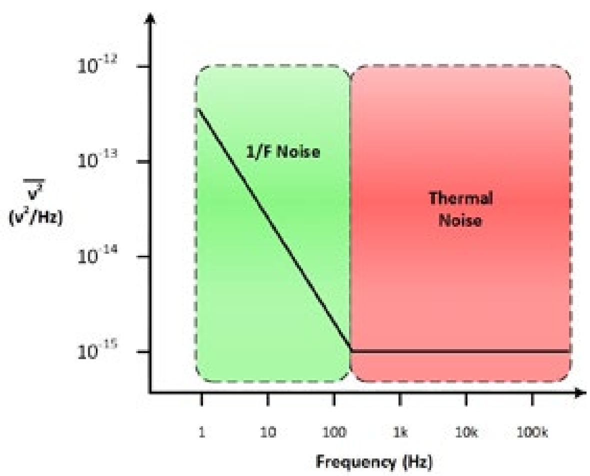


Figure 4. An ideal resistor is reflected in A, but, practically, resistors have internal thermal noise as represented in B.

Noise Reduction Strategies

Although noise is a serious problem for the designer, especially when low signal levels are present, a number of commonsense approaches can minimize the effects of noise on a system. Here are some strategies to help reduce noise:

- Keep the source resistance and the amplifier input resistance as low as possible. Using high value resistances increases thermal noise proportionally.
- Total thermal noise is also a function of the bandwidth of the circuit. Therefore, reducing the bandwidth of the circuit to a minimum also minimizes noise. But this job must be done mindfully because signals have a Fourier spectrum that must be preserved for accurate measurement. The solution is to match the bandwidth to the frequency response required for the input signal.
- Prevent external noise from affecting the performance of the system by appropriate use of grounding, shielding, cabling, careful physical placement of wires, and filtering.
- Use a low-noise amplifier in the input stage of the system.
- For some semiconductor circuits, use the lowest DC power supply potential that does the job.

Summary

- Sensitivity is the smallest change in an input signal that causes the measuring device to respond.
- Accuracy is defined as a measure of the capability of the instrument to faithfully indicate the value of the measured signal.
- The accuracy and sensitivity are documented in the specifications document; because companies and products at the same company may use these terms differently, always check the documentation and contact the company for clarification if needed.
- Precision is defined as a measure of the stability of the instrument and its capability of resulting in the same measurement over and over again for the same input signal.
- Noise is any unwanted signal that interferes with the wanted signal.
- There are different types of noise and different strategies to help reduce noise.

Digital States, Voltage Levels, and Logic Families

Overview

Learn about digital states, voltage logic levels, and logic level families for digital signals. This tutorial is part of the Instrument Fundamentals series.

Contents

Digital States

In digital devices, there are only two states: on and off. Using only these two states, devices can communicate a great deal of data and control various other devices. In binary, these states are represented as a 1 or 0. Binary 1 is typically considered a logic high, and 0 is a logic low.

a. Voltage Levels

Digital devices, however, are often driven by analog devices with an infinite number of states. How do you turn an infinite number of states into only two? The answer is by creating voltage logic levels, which define the voltage to represent a logic high or logic low.

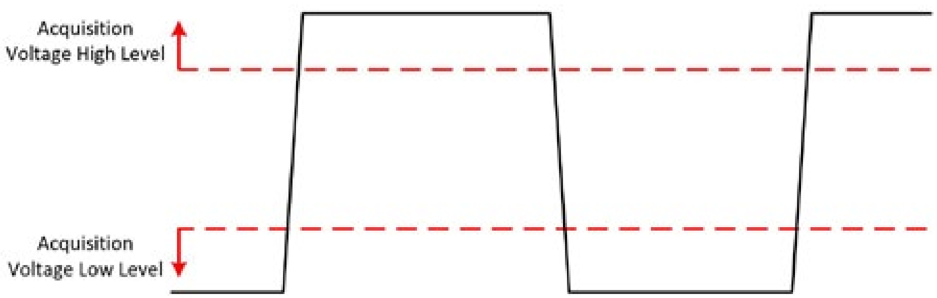


Figure 1. Voltage levels define the analog voltage that represents a logic high or low.

A system can define the voltage logic levels at any value it chooses, but many circuits represent a logic high by +5 V or +3.3 V to ground and a logic low as ground or 0 V. This type of system is called a positive or active-high. It describes how the pin is activated—for an active-high pin, you connect it to your high voltage.

A negative or active-low system is the reverse. The higher voltage represents a logic low and the lower voltage represents a logic high. For an active-low pin, you must pull that pin low by connecting it to ground. Data sheets often denote a pin is active-low by putting a line above the pin name, such as $\overline{\text{EN}}$.

Although a high and low are specified, in most systems there is actually a range so as to be more practical. For example, a logic high might be any value between 2 V and 5 V and a low might be any value from 0 V to 1 V. Voltages outside those ranges are considered invalid and occur only in a fault condition or during a logic-level transition.

b. Z and X States

Although a digital signal can have only two states—on and off—you can use additional states to assist with acquiring and generating digital signals. With tristate logic, there is a third possible condition: a high-impedance state where the output is disconnected from the line. This state isn't a high or low, but rather a floating or high-impedance state. It has the designation Z and is often used as an enable line.

The most common use of the Z state is to test one or more digital lines that can be driven by multiple transmitters. The data port on a memory chip is a good example of this. When the computer writes into the memory device, the computer needs to drive the data to be written into the memory chip on the data pins of the memory device (either 0 or 1). Later, when the computer processor needs to read out the contents of the memory, the memory device needs to drive the previously stored data value back to the computer processor (usually a Z state on the data pins).

A fourth state you might see is the hold state that is designated with an X. When generating digital signals, you might find it useful to have the device simply maintain the channel at its current state, regardless of which state that might be. This state is useful when setting initial or idle states.

When you acquire data, the X state has a different designation of don't care. This state is useful when you are comparing an acquired digital signal to an expected signal. For instance, on a signal you may care about only the first four values in a 10-value signal. You can use the X state for the last six values and compare just the first four.

State	Designation
0	Logic Low
1	Logic High
Z	High Impedance
X	Hold State or Don't Care

Table 1. A digital signal can be in only a high or low state; however, the Z and X states can assist in applications that generate or acquire digital signals.

Logic Families

Standardized logic families make it easier to work with circuits and components. They provide a standardized voltage level that constitutes a logic high or logic low. All circuits within a logic family are compatible with other circuits within that same family because they share the same characteristics.

a. Single-Ended Logic Families

Single-ended logic families specify voltage levels in relation to ground. The four levels are defined as:

- **V_{OH} (output high-level voltage)**–This is also known as the generation voltage high level. When configured for active drive generation, this is the voltage produced by the device when it generates a logic high. When configured for open collector generation, this is the equivalent to setting the data channel to a high-impedance state.
- **V_{OL} (output low-level voltage)**–This is also known as the generation voltage low level. This is the voltage produced by the device when it generates a logic low.
- **V_{IH} (input high-level voltage)**–This is also known as the acquisition voltage high level. This is the voltage level necessary to send to the device for it to read a logic high.
- **V_{IL} (input low-level voltage)**–This is also known as the acquisition voltage low level. This is the voltage level necessary to send to the device for it to read a logic low.

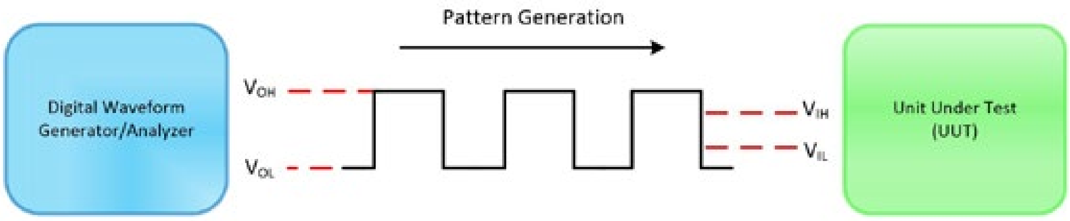


Figure 2. Single-ended logic levels are specified for output and input.

To accurately communicate with a device, be sure to configure the digital device such that the following conditions are met:

- $V_{OH} \geq$ DUT V_{IH}
- $V_{OL} \leq$ DUT V_{IL}
- $V_{IH} \leq$ DUT V_OH
- $V_{IL} \geq$ DUT V_{OL}
- $V_{IH} > V_{IL}$

There is usually a cushion between the output voltage of one device and the input of another. This is referred to as the noise margin or the noise immunity level (NIM). If you are in a noisy environment and having difficulty with incorrect data bits, consider increasing this value.

There are several single-ended logic families. Transistor-transistor logic (TTL) is very common for integrated circuits and is used in many applications such as computers, consumer electronics, and test equipment. Circuits built from bipolar transistors achieve switching and maintain logic states. A TTL must also meet specific current specifications and rise/fall times, which you can read more about in [What Is the Definition of a TTL-Compatible Signal?](#)

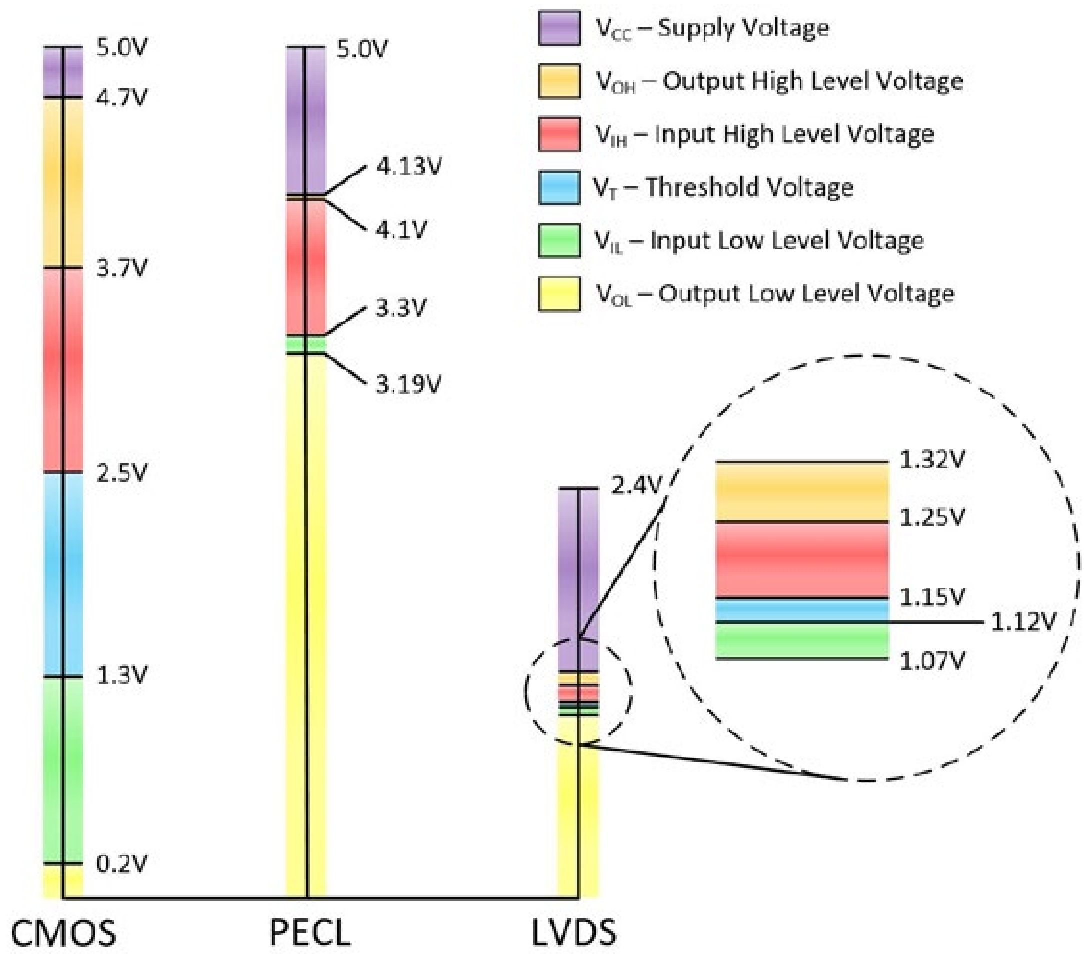


Figure 3. Standard 5 V TTL Voltage Levels

Another common IC family is CMOS. These devices have high noise immunity, require less power consumption, and have a lower based voltage. Most of the voltage levels are similar to TTL devices for greater compatibility. This makes it easy to switch from a TTL to a CMOS device, but going the other direction can be trickier. Too high a voltage to a CMOS could damage the chip. In this case, you can use a voltage divider to reduce the voltage.

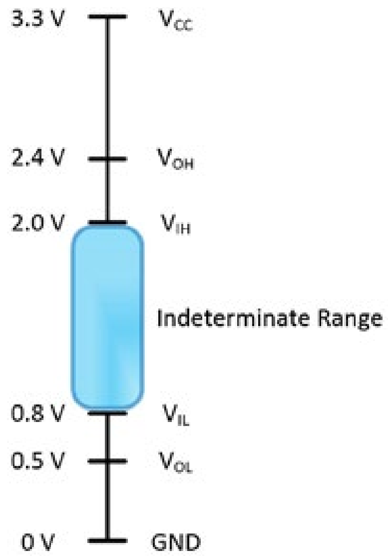


Figure 4. Standard CMOS Voltage Levels

It is always important to check you device's data sheet for voltage levels.

b. Differential Logic Families

Single-ended logic families use a set voltage level in relation to ground; however, differential logic families use the difference between two values and not a reference to ground. For the differential signal to be interpreted as a logic low, the signal must be less than its complementary signal by more than a particular value known as the threshold value (V_{TH}). Because the signals are referenced and transmitted together, you can achieve higher noise immunity in your signals than using single-ended logic families.

Voltage levels for differential logic families are typically specified from a differential rather than an absolute voltage. The four levels are defined as:

- V_{OD} **(output differential voltage)**–This is the difference in voltage between the signals.
- V_{OS} **(offset voltage)**–This is the common mode of the differential signal. Think of this as the average of the two signals. It is a reference to ground.
- V_{TH} **(threshold voltage)**–This is the difference in voltage needed for the device to register a valid logic state.
- V_{RANGE} **(input voltage range)**–This is the absolute voltage referenced from ground allowed by the device.

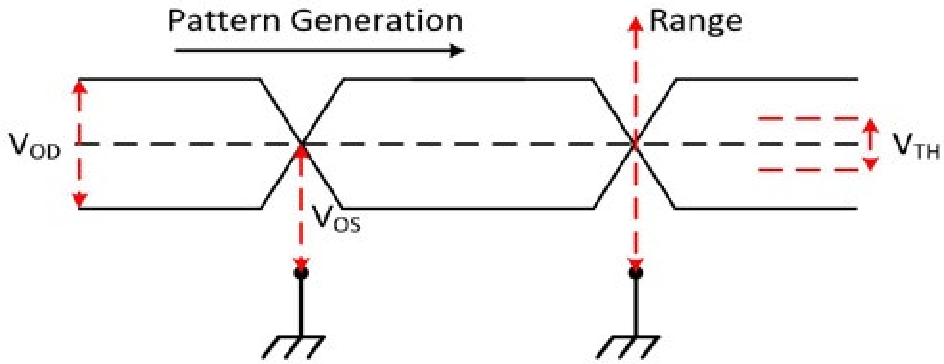


Figure 5. Voltage levels for differential logic families are typically specified from a differential rather than an absolute voltage.

Low-voltage differential signaling (LVDS) is a low-noise, low-power, and low-amplitude differential method. A current source is used to drive the signals. The electrical characteristics of an LVDS signal offer many performance improvements compared to single-ended standards. For example, because the received voltage is a differential between two signals, the voltage swing between the logic high-level and low-level state can be smaller, allowing for faster rise and fall times and thus faster toggle and data rates. Also, the differential receiver is less susceptible to common-mode noise than single-ended transmission methods.

Emitter-coupled logic (ECL) circuits use a design that uses transistors to steer current through gates, which compute logical functions. Because the transistors are always in the active region, they can change state very rapidly, so ECL circuits can operate at very high speeds. Low-voltage positive emitter-coupled logic (LVPECL) circuits are a type of ECL circuit that require a pair of signal lines for each channel. The differential transmission scheme is less susceptible to common-mode noise than single-ended transmission methods. LVPECL circuits are typically designed for use with VCC = 3 V or 3.3 V. To learn more about interfacing with ECL circuits, read Interfacing NI PXI-655x Digital Waveform Generator/Analyzers to ECL Logic Families.

Summary

- A **voltage logic level** defines the voltage to represent a logic high or a logic low.
- Many circuits represent a **logic high** by +5 V or +3.3 V to ground and a **logic low** as ground or 0 V. This type of system is called a **positive or active-high**.
- In **tristate logic**, the **Z state** is a high-impedance state and is often used as an enable line.
- In digital generation, the **X state** maintains the current logic level. In digital acquisition, it indicates a *don't care* state.
- **Logic families** provide a standardized voltage level that constitutes a logic high or logic low.
- **TTL** is based on V_{cc} = 5 V.
- **CMOS** is based on V_{cc} = 3.3V.

- **Differential logic families** use the difference between two values and not a reference to ground.
- **LVDS** is a low-noise, low-power, and low-amplitude differential method with V_{cc} = 3.3 V.
- **LVPECL** circuits are a type of ECL circuit that require a pair of signal lines for each channel (V_{cc} = 3 or 3.3 V).

Digital Timing:

Clock Signals, Jitter, Hystereisis, and Eye Diagrams

Overview

Learn about digital timing of clock signals and common terminology such as jitter, drift, rise and fall time, settling time, hysteresis, and eye diagrams. This tutorial is part of the Instrument Fundamentals series.

Contents

Clock Signals

When sending digital signals, a 0 or 1 is being sent. However, for different devices to communicate, timing information needs to be associated with the bits sent. Digital waveforms are referenced to clock signals. You can think of a clock signal as a conductor that provides timing signals to all parts of a digital system so that each process may be triggered at a precise moment.

A clock signal is a square wave with a fixed period. The period is measured from the edge of one clock to the next similar edge of the clock; most often is it measured from one rising edge to the next. The frequency of the clock can be calculated by the inverse of the clock period.

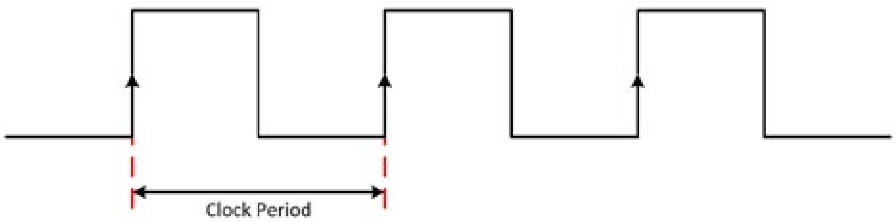


Figure 1. Digital waveforms are referenced to clock signals, which have a fixed period to synchronize digital transmitters and receivers during data transfer.

The duty cycle of a clock signal is the percentage of the waveform period that the waveform is at a logic high level. Figure 2 shows the difference between two waveforms with different duty cycles. You can see that the 30 percent duty cycle waveform is at a logic high level for less time than the 50 percent duty cycle.

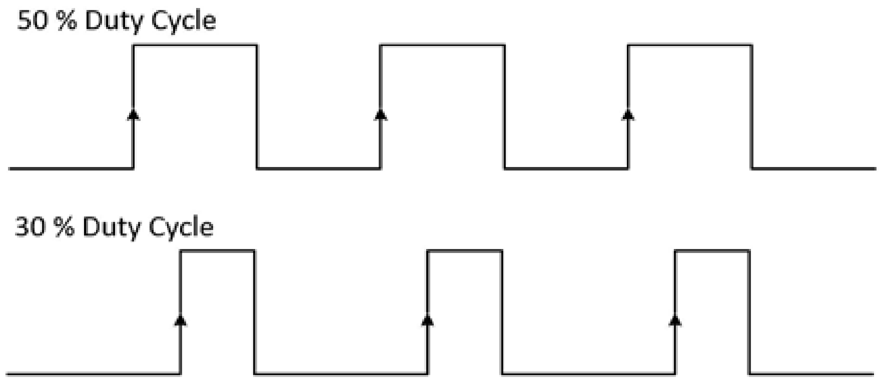


Figure 2. The duty cycle of a signal is the percentage of time the waveform is at a logic high level.

Clock signals are used to synchronize digital transmitters and receivers during data transfer. For example, the transmitter can use each rising edge of the clock signal to send each bit of data and the receiver can use the same clock to read the data. In this scenario, the assertion edge of the device is the rising edge (low to high). For other devices, it is the falling edge (high to low). The assertion edge of the clock is also called the active clock edge. Digital transmitters drive new data samples on each active clock edge while receivers sample data on each active clock edge. Newer devices are beginning to use both the rising and falling edge of the clock; these are called double date rate (DDR) devices. In actuality, the data is transmitted after a small delay from the assertion edge of the clock; this delay is called the clock-to-out time.

When a receiver samples the data on the digital lines, there are two timing parameters to be aware of in order to receive data reliably. Setup time (ts) is the amount of time the data must be at a valid logic level uninterrupted while the receiver sets itself up to receive the input. The hold time (tH) specifies the amount of time the data needs to hold the state before the it can change after it has been sampled by the receiver. Together, the setup time and hold time set up a stable window around the assertion edge of the receiver's clock for the receiver to reliably sample the data. Figure 3 shows setup and hold times with reference to a rising-edge clock signal. Often, digital signals switch the voltage midway between the supply rails; for this reason, the time reference markers are positioned at the midpoint of the signal edges.

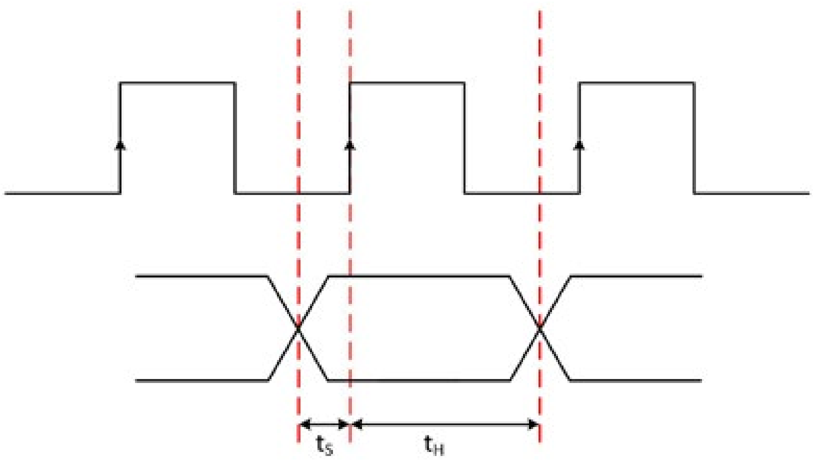


Figure 3. The duty cycle of a signal is the percentage of time the waveform is at a logic high level.

Common Terminology

In digital systems, timing is one of the most important factors. The reliability and accuracy of digital communications are based on the quality of its timing. However, in the real world, nothing is ever ideal. Below are some common terms and ways to better understand the timing of your particular digital signal.

a. Jitter

Jitter is the deviation from the ideal timing of an event to the actual timing of an event. To understand what this means, imagine you are sending a digital sine wave and plotting it on graph paper. Each square corresponds to a clock pulse; because the vertical lines are equidistant apart, you end up with a perfectly periodic clock signal. At each clock pulse, you receive three bits and plot that point on your graph paper. Because of the periodic nature, it ends up as a nice sine wave.

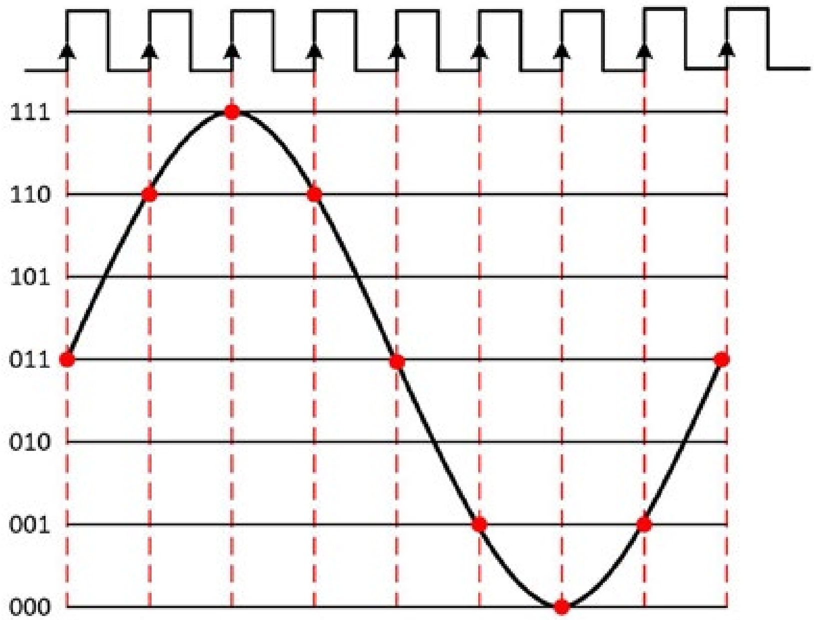


Figure 4. A sample clock that is periodic allows a digital system to communicate correctly and accurately.

Now, imagine that those lines aren't equidistant apart. This would make your clock signal less periodic. When you plot your data, it isn't at the same intervals and, thus, doesn't look correct.

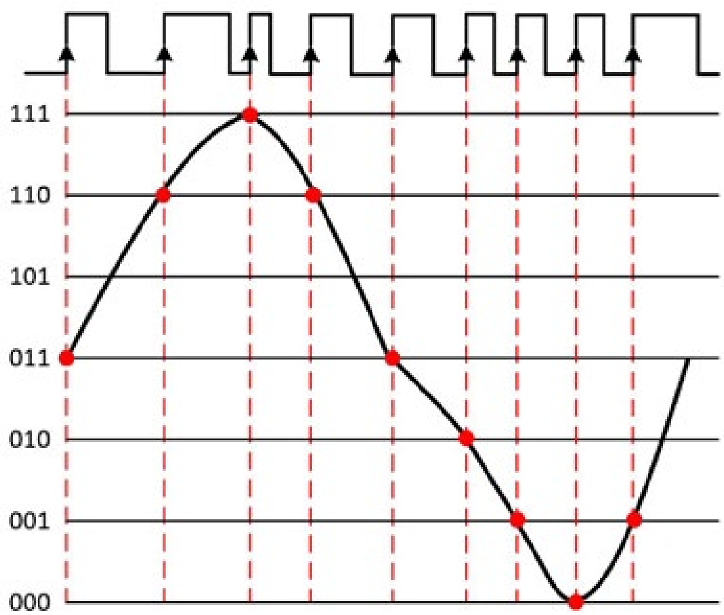


Figure 5. If a clock signal has jitter, it results in distortion of the digital waveform.

In Figure 5, you can see that the distance between the transitions in the clock signal is uneven; this is jitter in the clock. Although the above figure has an exaggerated amount of jitter, it does show how a jittery clock can cause samples to be triggered at uneven intervals. This unevenness introduces distortion into the waveform you are trying to record and reproduce.

Now look at jitter in terms of a digital signal with only 1s and 0s. Remember, jitter is the deviation from the ideal timing of an event to the actual timing of an event. Taking a look at a single pulse, jitter is the deviation in edge timing from the actual signal to the ideal positions in time.

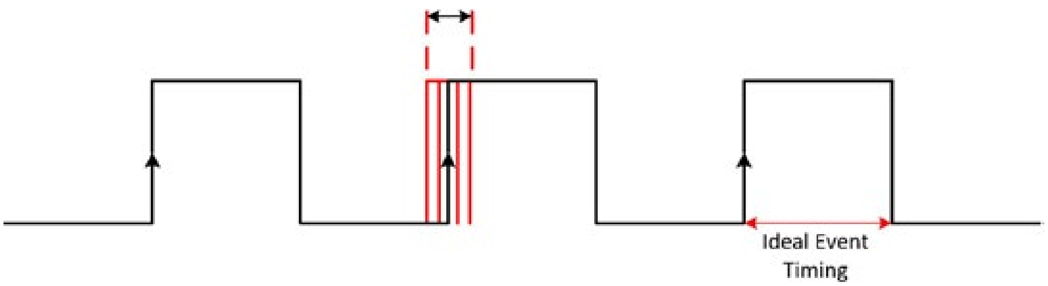


Figure 6. Jitter of a single pulse is the deviation in edge timing.

Jitter is typically measured from the zero-crossing of a reference signal. It typically comes from cross-talk, simultaneous switching outputs, and other regularly occurring interference signals. Jitter varies over time, so measurements and quantification of jitter can range from a visual estimate on an oscilloscope in the range of jitter in seconds to a measurement based on statistics such as the standard deviation over time.

b. Drift

Another common timing issue is drift. Clock drift occurs when the transmitter's clock period is slightly different from that of the receiver. At first, it may not make much of a difference. However, over time, the difference between the two clock signals may become noticeable and cause loss of synchronization and other errors.

c. Rise Time, Fall Time, and Aberrations

Even with drift, in theory, when a digital signal goes from a 0 to a 1, it would happen instantaneously. However, in reality, it takes time for a signal to change between high and low levels. Rise time (trise) is the time it takes a signal to rise from 20 percent to 80 percent of the voltage between the low level and high level. Fall time (tfall) is the time it takes a signal to fall from 80 percent to 20 percent of the voltage between the low level and high level.

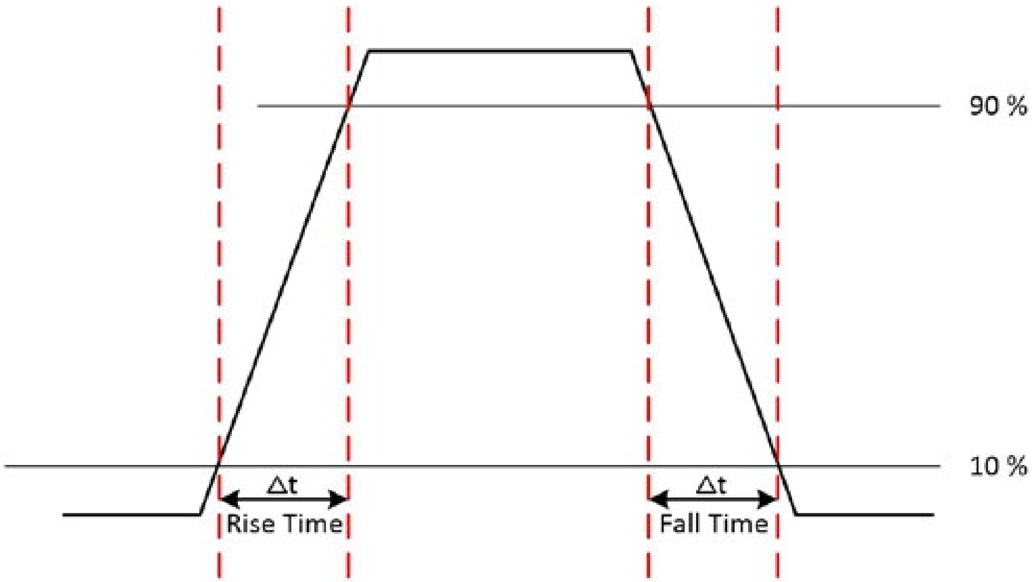


Figure 7. Rise time and fall time indicate the length of time a signal takes to change voltage between the low level and high level.

In addition, in the real world, a signal rarely hits a voltage level and stays there in a clean fashion. When a signal actually exceeds the voltage level following an edge, the peak distortion is called overshoot. If the signal exceeds the voltage level preceding an edge, the peak distortion is called preshoot. In between edges, if the signal drifts short of the voltage level it is called undershoot.

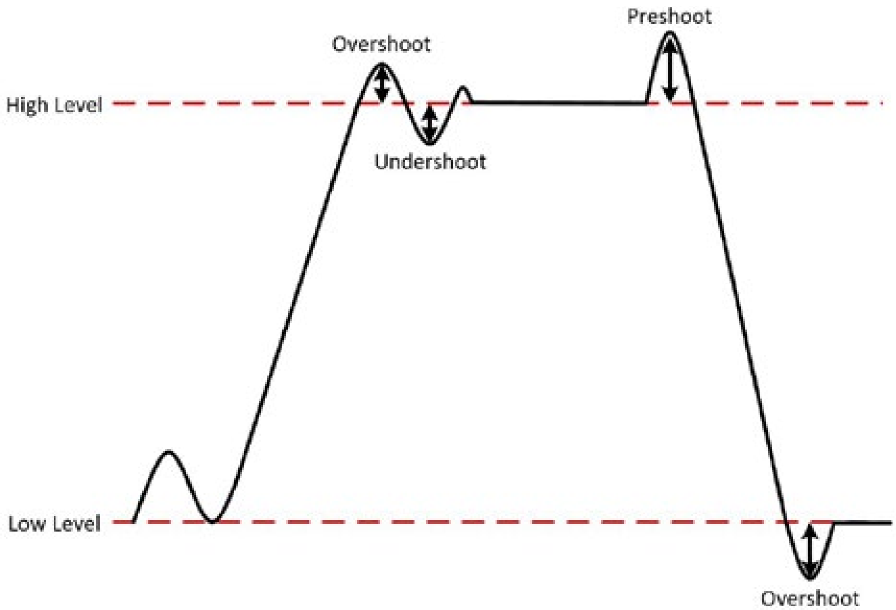


Figure 8. Overshoot, preshoot, and undershoot are collectively called aberrations.

Together, overshoot, preshoot, and undershoot are called aberrations. Aberrations can result from board layout problems, improper termination, or quality problems in the semiconductor devices themselves.

d. Settling Time

After a digital signal has reached a voltage level, it bounces a little and then settles to a more constant voltage. The settling time (ts) is the time required for an amplifier, relay, or other circuit to reach a stable mode of operation. In the context of digital signal acquisition, the settling time for full-scale step is the amount of time required for a signal to reach a certain accuracy and stay within that range.

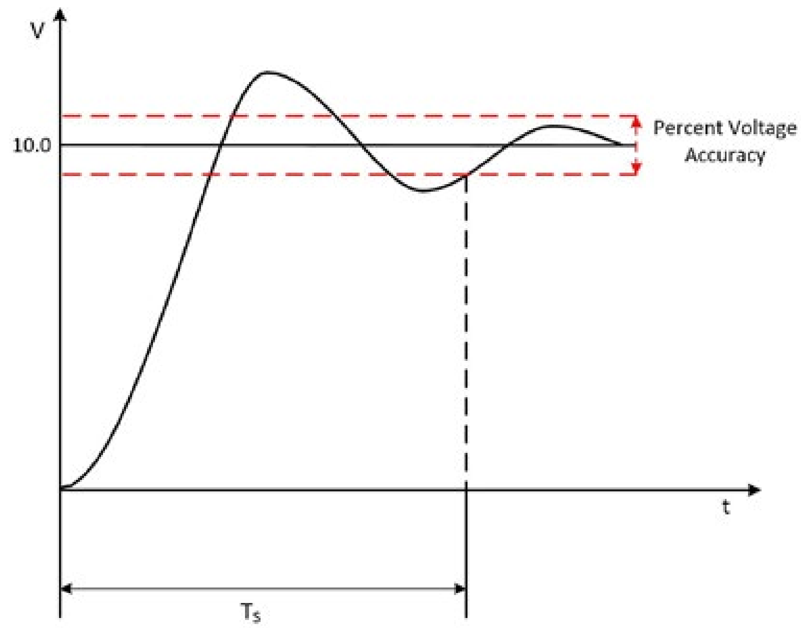


Figure 9. Settling time is the amount of time for a signal to reach a certain accuracy and stay within that range.

e. Hysteresis

Hysteresis refers to the difference in voltage levels between the detection of a transition from logic low to logic high, and the transition from logic high to logic low. It can be calculated by subtracting the input high voltage from the input low voltage.

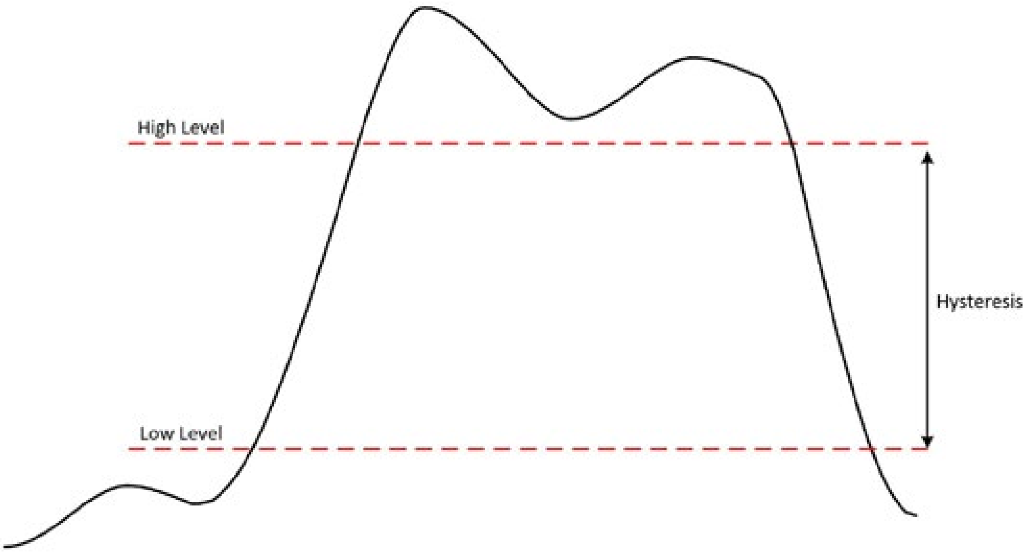


Figure 10. Hysteresis is the difference in voltage levels between the detection of a transition from one logic value to another.

Hysteresis is a useful property for digital devices, because it naturally provides some amount of immunity to high-frequency noise in your digital system. This noise, often caused by reflections from the high-edge rates of logic level transitions, could cause the digital device to make false transition detections if only a single voltage threshold determined a change in logic state. You can see this in Figure 11. The first sample is acquired as a logic low level. The second sample is also a logic low level because the signal has not yet crossed the high-level threshold. The third and fourth samples are logic high levels, and the fifth is a logic low level.

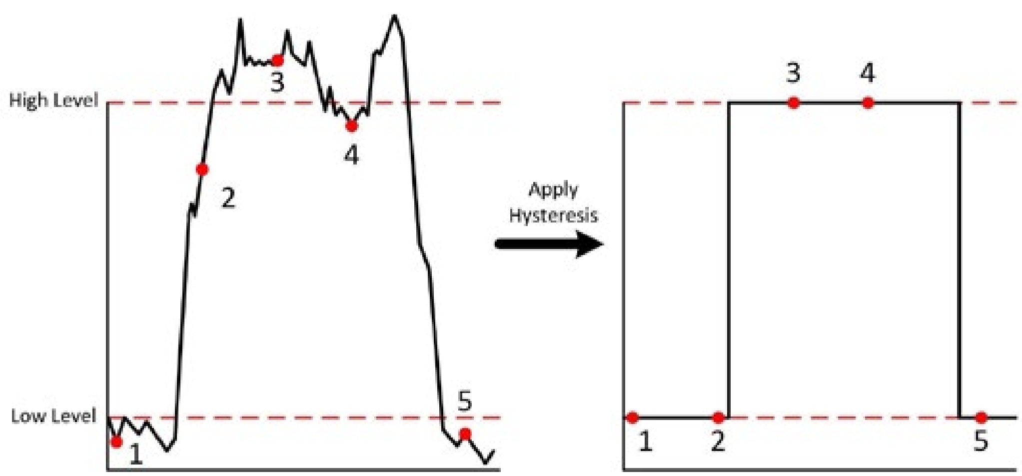


Figure 11. Hysteresis provides an amount of immunity to high-frequency noise in your digital system.

For devices with fixed voltage thresholds, the noise immunity margin (NIM) and hysteresis of your system are determined by your choice of system components. Both system NIM and hysteresis give your system levels of noise immunity, but for a specific logic family, there is always a trade-off between these two—the larger the hysteresis, the smaller the NIM, and vice versa. To determine how to set your voltage thresholds, you should carefully examine the signal quality in your system to determine whether you need more noise immunity from your high and low logic levels (greater NIM) or need more noise immunity on your logic level transitions (greater hysteresis).

f. Skew

Skew is when the clock signal arrives at different components at different times. Unlike drift, the clock signals have the same period; they just arrive at different times. This can be caused by a variety of factors including wire length, temperature variation, or differences in input capacitance. Channel-to-channel skew generally refers to the skew across all data channels on a device. When each sample is acquired, the point in time at which each data channel is sampled with respect to every other data channel is not identical, but the difference is within some small window of time called the channel-to-channel skew.

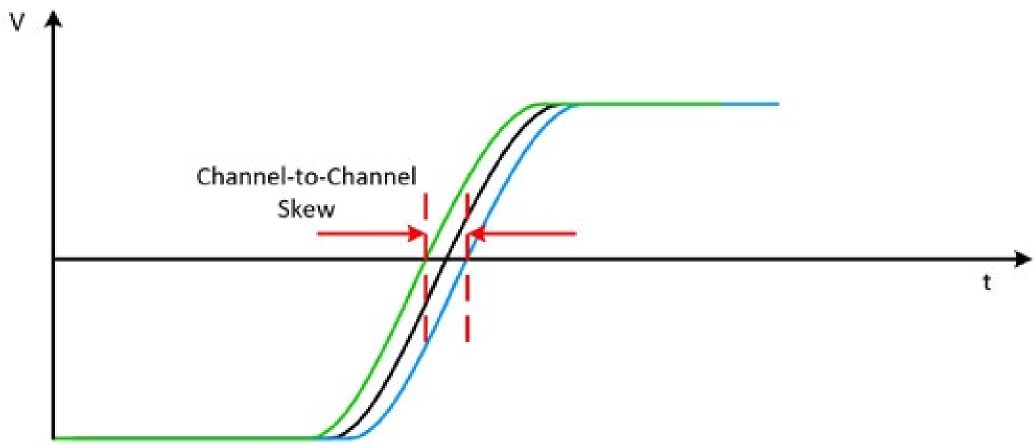


Figure 12. Channel-to-channel skew generally refers to the skew across all data channels on a device.

g. Eye Diagram

An eye diagram is a timing analysis tool that provides you with a good visual of timing and level errors. In real life, errors, like jitter, are difficult to quantify because they change so often and are so small. Therefore, an eye diagram is an excellent tool for finding the maximum jitter as well as measuring aberrations, rise times, fall times, and other errors. As these errors increase, the white space in the center of the eye diagram decreases

An eye diagram is created by overlaying sweeps of different segments of a digital signal. It should contain every possible bit sequence from simple high to low transitions to isolated transitions after long runs of consistency. When overlapped, it looks like an eye. Eye diagrams are a visual way to understand the signal integrity of a design. Keep in mind that an eye diagram shows parametric information about a signal, but does not detect logic or protocol problems such as when it is supposed to transmit a high but sends a low.

Figure 13 shows common terminology of an eye diagram.

A. **High level,** also called the one level, is the main value of a logic high. The calculated value of a high level comes from the mean value of all the data samples captured in the middle 20 percent of the eye period.

B. **Low level,** also called the zero level, is the main value of a logic low. This level is calculated in the same region as the high level.

C. **Amplitude** of the eye diagram is the difference between the high and low levels.

D. **Bit period,** also referred to as the unit interval (UI), is a measure of the horizontal opening of an eye diagram at the crossing points of the eye. It is the inverse of the data rate. When creating eye diagrams, using the bit period on the horizontal axis instead of time, gives you the ability to compare diagrams with different data rates easily.

E. **Eye height** is the vertical opening of an eye diagram. Ideally, this would equal the amplitude, but this rarely occurs in the real world because of noise. As such, the eye height is smaller the more noise in the system. The eye height indicates the signal-to-noise ratio of the signal.

F. **Eye width** is the horizontal opening. It is calculated as the difference between the statistical mean of the crossing points of the eye.

G. **Eye crossing percentage** shows duty cycle distortion or pulse symmetry problems. An ideal signal is 50 percent; as the percentage deviates, the eye closes and indicates degradation of the signal.

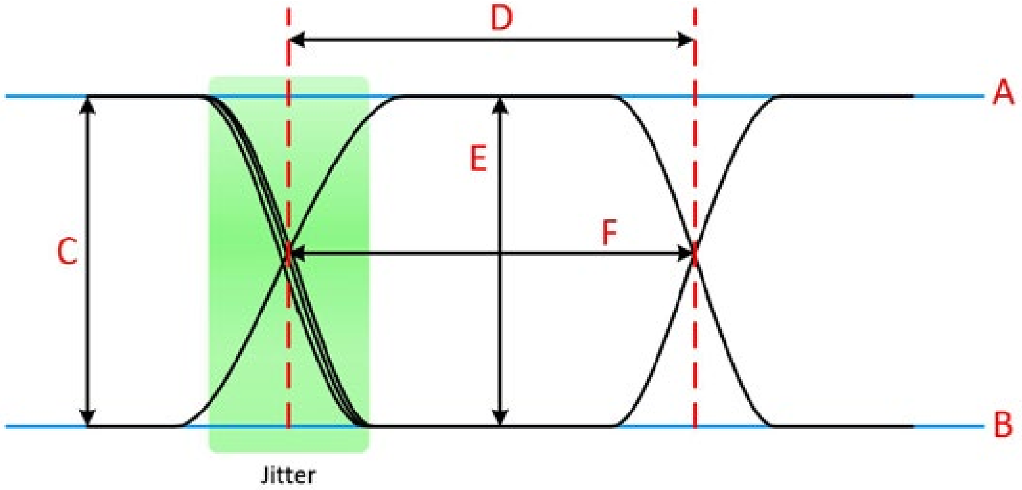


Figure 13. The image shows the high level (A), low level (B), amplitude (C), bit period (D), eye height (E), eye width (F), and eye crossing percentage (G) on an eye diagram.

Figure 14 shows additional measurements on an actual eye diagram.

A. **Rise time** in the diagram is the mean of the individual rise times. The slope indicates sensitivity to timing error; the smaller the better.

B. **Fall time** in the diagram is the mean of the individual fall times. The slope indicates sensitivity to timing error; the smaller the better.

C. The width of the logic high value is the amount of **distortion** in the signal (set by the signal-to-noise ratio).

D. The **signal-to-noise ratio** at the sampling point is from the eye width to the bottom or the logic high-voltage range.

E. **Jitter** of the signal.

F. The most open part of the eye is when there is the best signal-to-noise ratio and is thus the **best time to sample**.

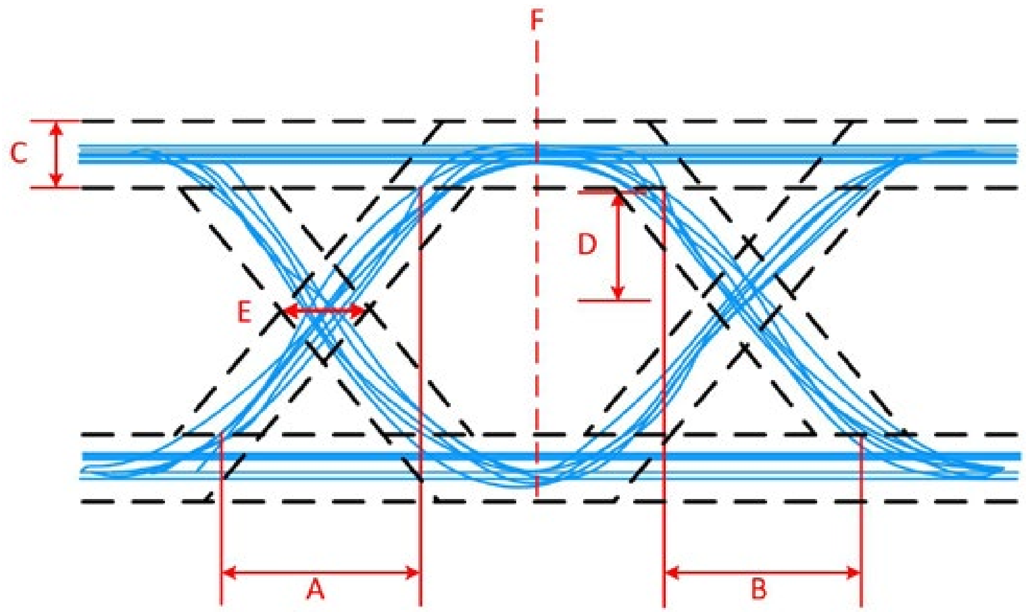


Figure 14. The image shows the rise time (A), fall time (B), distortion (C), signal-to-noise ratio (D), jitter (E), and best time to sample (F) on an eye diagram.

Summary

- Digital waveforms are referenced to **clock signals**, which have a fixed period to synchronize digital transmitters and receivers during data transfer.
- The **duty cycle** of a clock signal is the percentage of the waveform period that the waveform is at logic high level.
- The assertion edge of the clock is called the **active clock edge**.
- The **setup time and hold time** set up a stable window around the assertion edge of the receiver's clock for the receiver to reliably sample the data.
- **Jitter** is the deviation from the ideal timing of an event to the actual timing of an event; jitter in timing can cause distortion in the signal.
- Clock **drift** occurs when the transmitter's clock period is slightly different from that of the receiver and can cause loss of synchronization and other errors.
- **Rise time** and **fall time** indicate the length of time a signal takes to change voltage between the low level and the high level.
- Overshoot, preshoot, and undershoot are called **aberrations** and are an indication of errors in the system.
- **Settling time** is the amount of time for a signal to reach a certain accuracy and stay within that range.
- **Hysteresis** provides an amount of immunity to high-frequency noise in your digital system.
- **Skew** is when the clock signal arrives at different components at different times.
- An **eye diagram** is a timing analysis tool that provides you with a good visual of timing and level errors.

Direct Digital Synthesis

Overview

Learn the fundamentals and theory behind direct digital synthesis and how it applies to function generators and arbitrary function generators. This tutorial is part of the Instrument Fundamentals series.

Contents

Introduction

The Generating a Signal white paper explores the beginnings of how signal generators, such as function generators and arbitrary function generators (AFGs), output a wanted analog signal and how some signal generators use direct digital synthesis (DDS) technology to output signals at precise frequencies. This article discusses the components and technology that give signal sources the ability to achieve sub-Hertz accuracy in signal generation.

Theory of Operation

Signal generators that use DDS generate signals at precise frequencies through a unique memory access and clocking mechanism, which differs from the traditional method of outputting each sample in the order of which the waveform is stored. Arbitrary waveform generators (AWGs) use the traditional signal generation method. AWGs can produce complex user-defined waveforms, but are limited in the frequency precision at which the waveform is generated. This is because of the constraints that the waveform must produce point by point from the AWG's memory and the sample clock controlling the time between each point generated has a finite number of frequencies.

Function generators and AFGs that use DDS store a large amount of points for a single cycle of a periodic waveform in memory. DDS technology gives the function generator or AFG the ability to choose which sample to output from memory. Because the function generator or AFG is not restricted on choosing the next sample in the waveform, it can produce signals at precise frequencies. Figure 1 graphically represents how a function generator or AFG can produce a 21 MHz sine wave, which is not an integer division of the 100 MHz sample clock. The 100 MHz sample clock still drives the update rate of the DAC output; therefore, the faster the sample clock, the more accurate the shape of the created signal.

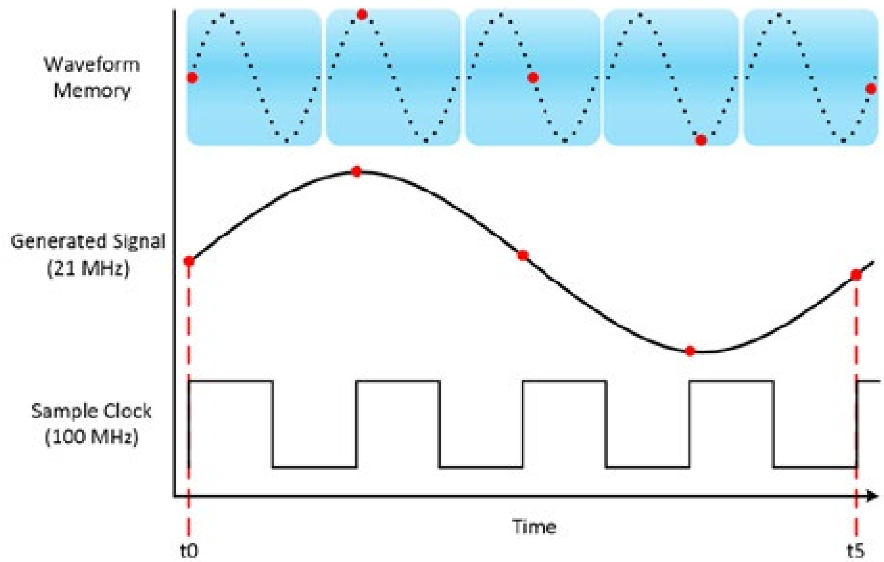


Figure 1. In DDS-capable hardware, the samples are not necessarily chosen in the order they are stored in memory. This allows the 100 MHz sample clock to accurately create the 21 MHz sine wave.

In the specific case above, the AFG uses the 100 MHz sample clock to drive the DAC but the frequency of the signal generated is created by the method of which the samples are chosen from the waveform memory location. The next sections discuss the components that implement the controlling logic behind the sample choice.

Functional Overview

DDS implementation requires three main hardware building blocks: a (a) sample clock, (b) phase accumulator, and (c) lookup table, which is an implementation of a programmable read-only memory. Figure 2 shows the higher level flow from hardware block to hardware block.

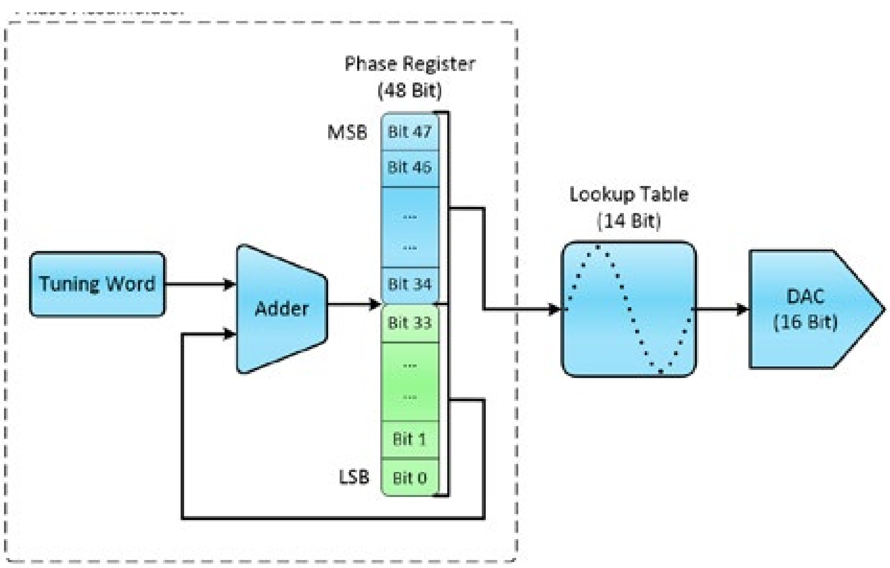


Figure 2. Hardware Block Diagram for the DDS Architecture

a. Sample Clock

The sample, or reference, clock is used to create the frequency tuning word, update the phase accumulator value, and drive the digital-to-analog conversion. The sample clock determines when a sample is output by the DAC, but it does not directly determine the frequency of the output signal.

b. Phase Accumulator

The phase accumulator is a collection of components that allows a function generator or AFG to output at precise frequencies. To create the signal at a precise frequency, the phase accumulator uses three general components. First, the phase accumulator uses the tuning word to specify the frequency of the signal. The tuning word is a 24- to 48-bit digital word that specifies how many samples to jump in the waveform memory. The second component, the adder, takes the tuning word and sums it to the phase register remainder. This new digital value is output to the phase register. The final component of the phase accumulator, the phase register, takes the new digital word and uses it to specify the memory address of the next sample point to be output in the lookup table. The phase register takes the remaining most significant bits not used in the lookup table memory address and provides them back to the adder to ensure frequency precision over time.

c. Lookup Table

The output of the phase register only looks like a digital ramp as the memory address increases over time, which is changing at the rate specified by the tuning word. Therefore, to output the wanted waveform, the output of the phase register points to the needed waveform sample address in the lookup table. The lookup table then provides the digital word at the provided memory address, which is the digital word of the correct amplitude and phase for the DAC to produce.

Frequency agility, or the ability to change the waveform's frequency very rapidly and phase continuously, is one of the main benefits to the DDS architecture. An AFG using DDS can change the waveform's frequency very rapidly because only the tuning word needs to be changed in order to change the waveform's frequency.

Common Applications

As discussed above, DDS technology provides two main benefits. One major benefit of DDS technology is the frequency accuracy of the generated signal. This capability opens the door to extremely accurate component testing because you can rely on the frequency accuracy of the function generator or AFG-created signal.

The capability to change the generated signal's frequency extremely rapidly and phase continuously is the second main benefit of DDS technology. This allows for more efficient component testing over specific ranges because you can implement the frequency change quickly and also stress test devices by pushing the limits on what signal they are providing to the device under test.

A specific example where AFGs with DDS technology are extremely valuable is accurate filter characterization. The characterization of the filter is only accurate if the signal provided to the filter is generated precisely by the AFG and if the filtered signal is accurately measured by an oscilloscope. Figure 3 represents a typical test setup for filter characterization.

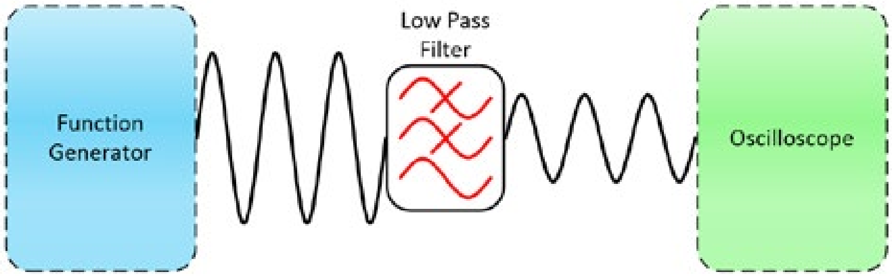


Figure 3. Filter Characterization Application Block Diagram With a DDS-Capable Function Generator, a Lowpass Filter, and an Oscilloscope

Summary

- Signal generators without DDS technology produce waveforms by outputting the stored waveform point by point at the frequency of the sample clock.
- Signal generators with DDS technology can produce periodic waveforms at many frequencies with extreme frequency accuracy. This is because of the unique memory access and clocking mechanism.
- DDS technology is implemented with three higher level hardware blocks: the sample clock, the phase accumulator, and the lookup table.
- The **sample clock** creates the **frequency tuning word**, updates the phase accumulator value, and drives the DAC output rate
- The **phase accumulator** takes the frequency tuning word as input and provides the digital memory address of the next sample to be output in the lookup table.
- The **lookup table** stores the periodic waveforms as digital samples. The lookup table takes the memory address from the phase accumulator and provides the digital waveform sample at that memory address to the DAC.
- Signal generators with DDS technology should be used for applications that require precise frequency generation or frequency agility.
- Applications that require extremely large, complex, and user-defined waveforms may be best served by arbitrary waveform generators instead of arbitrary function generators with DDS technology.

DMM Measurement Types and Common Terminology

Overview

Learn how to correctly use and understand a digital multimeter (DMM). Article topics include display digits, AC and DC voltage measurements, AC and DC current measurements, resistance measurements, and continuity and diode testing. This tutorial is part of the Instrument Fundamentals series.

Contents

Display Digits

Digital multimeters (DMMs) can be useful for a variety of measurements. When choosing a DMM or understanding one you are using, the first things to be aware of are the display digits of the instrument.

It is important that a DMM has enough digits to be precise enough for your application. The number of display digits on a DMM is not related to the resolution, but can help determine the number of significant values that can be displayed and read. DMMs are said to have a certain number of digits, such as 3 ½ digits or 3 ¾ digits. A full digit represents a digit that has 10 states, 0 to 9. A fractional digit is the ratio of the maximum value the digit can attain over the number of possible states. For example, a ½ digit has a maximum value of one and has two possible states (0 or 1). A ¾ digit has a maximum value of 3 with four possible states (0, 1, 2, or 3).

$$\textit{Fractional Digit} = \frac{\textit{Maximum Value Digit Can Attain}}{\textit{Number of Possible States}}$$

Equation 1. DMMs often have fractional digits, which can display only a limited number of states.

The fractional digit is the first digit displayed, with the full digits displayed after. For instance, on the 2 V range, the maximum display for a 3 ½ digit DMM is 1.999 V.

Typically, ½ digit displays have full scale voltages of 200 mV, 2 V, 20 V, and 200 V while ¾ digit displays have full scale voltages of 400 mV, 4 V, 20 V, and 400 V.

Voltage Measurements

Practically every DMM has a DC and an AC measurement function. Voltage testing is commonly used to test and verify the outputs of instruments, components, or circuits. Voltage is always measured between two points, so two probes are needed. Some DMM connectors and probes are colored; red is intended for the positive point that you want to actually take a measurement of and black is intended for the negative point that is typically a reference or ground. However, voltage is bidirectional, so if you were to switch the positive and negative points, the measured voltage would simply be inverted.

There are usually two different modes for measuring voltage: AC and DC. Typically, DC is denoted with a V with one dashed line and one solid line while AC is denoted with a V with a wave. Be sure to select the correct range and mode for your application.

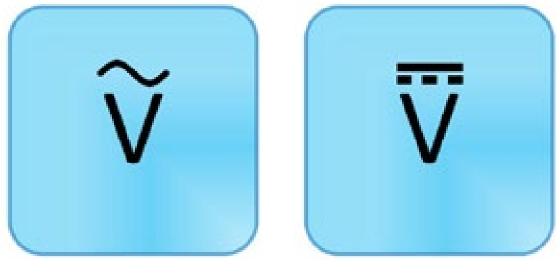


Figure 1. AC voltage (left) and DC voltage (right) measurements are commonly used to test and verify outputs of instruments, components, or circuits.

There are several terms and concepts you should be familiar with when measuring AC or DC voltage.

a. Input Resistance

An ideal voltmeter has an infinite input resistance so that the instrument does not draw any current from the test circuit. However, in reality, there is always some resistance that affects measurement accuracy. To minimize this problem, a DMM's voltage measurement subsystems are often designed to have impedances in the 1s to 10s of MΩ. If you are measuring low voltages, even this resistance can be enough to add unacceptable inaccuracies to your measurement. For this reason, lower voltage ranges often have a higher impedance option such as 10 GΩ.

With some DMMs, you can select the input resistance. For most applications, it can be said the higher the impedance, the more accurate the measurement. However, there are a few cases where you might choose the lower impedance. For instance, a conduit that has many different wires inside might have coupling across the wires. Even though the wires are open and floating, the DMM still reads a voltage. The higher impedance isn't sufficient to eliminate these ghost voltages, but a low impedance provides a path for this built-up charge and allows the DMM to correctly measure 0 V. An example of this at a lower voltage range is if you had traces close together on a circuit.

b. Crest Factor

When measuring AC signals (voltage or current), the crest factor can be an important parameter when determining accuracy for a specific waveform. The crest factor is the ratio of the peak value to the rms value and is a way to describe waveform shapes. Typically, the crest factor is used for voltages, but can be used for other measurements such as current. It is technically defined as a positive real number, but most often it is specified as a ratio.

$$Crest\ Factor = \frac{|V_{peak}|}{V_{rms}}$$

Equation 2. The crest factor is a measure of how extreme the peaks are in a waveform.

A constant waveform with no peaks has a crest factor of 1 because the peak value and the rms value of the waveform are the same. For a triangle waveform, it has a crest factor of 1.732. Higher crest factors indicate sharper peaks and make it more difficult to get an accurate AC measurement.

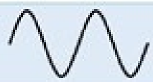
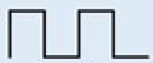
Wave Type	Waveform	RMS Value	Crest Factor
Sine Wave		$\frac{1}{\sqrt{2}} \approx 0.707$	$\sqrt{2} \approx 1.414$
Half-Wave Rectified Sine		$\frac{1}{2} = 0.5$	2
Full-Wave Rectified Sine		$\frac{1}{\sqrt{2}} \approx 0.707$	$\sqrt{2} \approx 1.414$
Triangle Wave		$\frac{1}{\sqrt{3}} \approx 0.577$	$\sqrt{3} \approx 1.732$
Square Wave		1	2

Figure 2. The crest factor of an AC signal can affect the accuracy.

An AC multimeter that measures using true rms specifies the accuracy based on a sine wave. It indicates, through the crest factor, how much distortion a sine wave can have and still be measured within the stated accuracy. It also includes any additional accuracy error for other waveforms, depending on their crest factor.

For example, if a given DMM has an AC accuracy of 0.03 percent of the reading. You are measuring a triangle waveform, so you need to look up any additional error with a crest factor of 1.732. The DMM specifies that for crest factors between 1 and 2, there is additional error of 0.05 percent of the reading. Your measurement then has an accuracy of 0.03 percent + 0.05 percent for a total of 0.08 percent of the reading. As you can see, the crest factor of a waveform can have a large affect on the accuracy of the measurement.

c. Null Offset

Most DMMs offer the ability to do a null offset. This is useful for eliminating errors caused by connections and wires when making a DC voltage or resistance measurement. First, you select the correct measurement type and range. Then connect your probes together and wait for a measurement to read. Then select the null offset button. Subsequent readings subtract the null measurement to provide a more accurate reading.

d. Auto Zero

In addition to performing a null offset, another way to improve voltage and resistance measurement accuracy is by enabling a feature called auto zero. Auto zero is used to compensate for internal instrument offsets. When the feature is enabled, the DMM makes an additional measurement for every measurement you take. This additional measurement is taken between the DMM input and its ground. This value is then subtracted from the measurement taken, thus subtracting any offsets in the measurement path or ADC. Although it can be very helpful in improving the accuracy of the measurement, auto zero can increase the time it takes to perform a measurement.

Current Measurements

Another common measurement function is DC and AC current measurements. Although voltage is measured in parallel with the circuit, current is measured in series with the circuit. This means that you need to break the circuit—physically interrupt the flow of current—in order to insert the DMM into the circuit loop to take an accurate measurement. Similar to voltage, current is bidirectional. The notation is similar as well, but with an A symbol instead of a V. The A stands for amperes, the unit of measure for current. Be sure to select the correct range and mode for your application.

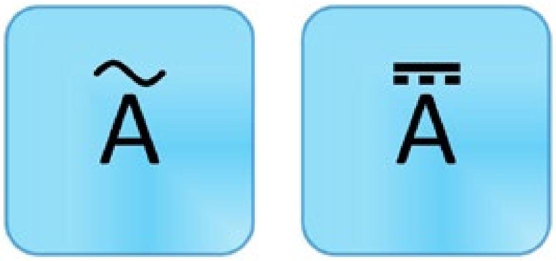


Figure 3. DC current (left) and AC current (right) measurements are helpful for troubleshooting circuits or components.

DMMs have a small resistance at the input terminals, and it measures the voltage. It then uses Ohm's law to calculate the current. The current is equal to the voltage divided by the resistance. To protect your multimeter, avoid switching out of the current measurement function when currents are flowing through the circuit. You should also be careful not to accidentally measure voltage while in the current measurement mode; this can cause the fuse to blow. If you do accidentally blow the fuse, you can often replace it. See your instrument's instruction manual for detailed information.

Resistance Measurements

Resistance measurements are commonly used to measure resistors or other components such as sensors or speakers. Resistance measuring works by applying a known DC voltage over an unknown resistance in series with a small internal resistance. It measures the test voltage, then calculates the unknown resistance. Because of this, test the device only when it isn't powered; otherwise, there is already voltage in the circuit and you can get incorrect readings. Also keep in mind that a component should be measured before it is inserted into the circuit; otherwise, you are measuring the resistance of everything connected to the component instead of just the component by itself.

One of the nice things about resistance is that it is nondirectional, meaning if you switch the probes the reading is still the same. The symbol for a resistance measurement is an Ω, which represents the resistance unit of measure. Be sure to select the correct range and mode for your application. If the display reads OL, this means the reading is over the limit or greater than the meter can measure in that range. As discussed earlier, using the null offset can improve your measurement readings.

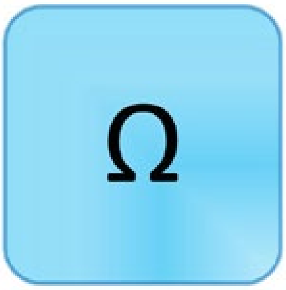


Figure 4. Resistance measurements are commonly used to measure resistors or other components.

Additional Measurements

Many DMMs offer two additional measurement functions: diode testing and continuity testing.

a. Continuity Testing

Continuity testing helps you identify when two points are electrically connected. This can be very helpful when troubleshooting wire breaks, printed circuit board (PCB) traces, or solder joints. When testing for continuity, it is essential to monitor exactly where the probes are touching. As such, most DMMs emit a sound when they detect a closed circuit, so you don't have to look up from your probes. As such, the symbol for continuity looks like a sound wave.

Figure 5. Continuity testing helps you identify when two points are electrically connected.

Continuity testing works just like a resistance measurement; as such, it is essential that your device not be powered when you are testing. It can also be helpful to make sure everything is connected first by brushing the test tips together to verify the beep. If you don't hear a sound, then check that the probes are firmly connected, your DMM has sufficient battery life, and that you are in the correct mode. You should also look in your user manual to determine the level of resistance required to trigger the sound as it varies from model to model.

If you are testing a circuit that has a large capacitor, you may hear a quick beep and then silence. This is because the voltage the DMM is applying to the circuit is charging up the capacitor and, during that time, the DMM thinks it is a closed circuit when it isn't really.

b. Diode Testing

Diode testing displays the forward voltage drop of the diode in volts. The symbol, not surprisingly, is the diode symbol.

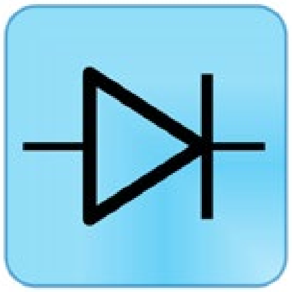


Figure 6. Diode testing displays the forward voltage drop of the diode in volts.

The DMM forces a small current through the diode and measures the voltage drop between the two test leads. When measuring a diode, you want the positive probe on the anode side and the negative on the cathode side. The voltage reading typically is about 0.7 V for silicon but can range from 0.5 to 0.9 V and still be a working diode. Germanium diodes are typically around 0.3 V.

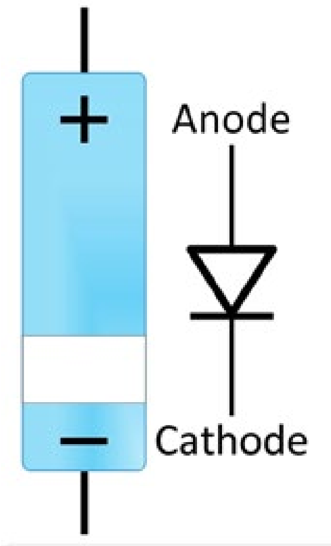


Figure 7. Typically, test a diode with the positive probe on the anode side and the negative on the cathode side. However, switching them can also be illuminating.

Next, switch probes so the negative is on the anode side and the positive is on the cathode side. If the diode is working properly, the multimeter should show that there is an open circuit indicated by OL.

If a diode is defective, it can defect to be either a short or an open diode. If the diode has failed to open, the DMM shows OL in both the forward and reverse bias because the current flowing through is zero and is an equivalent to an open circuit. If the diode is shorted, the DMM indicates 0 V as there is no voltage drop across the diode.

Noise Rejection Parameters

It is always important to consider noise when taking a measurement. There are two additional parameters you should be familiar with to better understand your instrument and the associated noise of the measurement.

The normal-mode rejection ration (NMRR) describes the DMM's ability to reject noise that appears between the two input terminals or, in other words, the noise mixed in with the measured signal. Most of this noise is a line frequency and its harmonics. NMRR, which is often used to indicate the capability of the instrument to reject a power line noise of 50 or 60 Hz, is valid only at the specified frequency and is useful when making DC measurements. Normal-mode noise can also be reduced through the use of shielding or filtering.

The common-mode rejection ratio (CMRR) describes the DMM's ability to reject noise that is common to both input terminals, such as from a noisy environment. Common-mode noise is usually less severe than normal-mode noise.

NMRR and CMRR are typically specified at 50 Hz and 60 Hz, and CMRR is often specified at a DC value as well. Typical values are greater than 80 dB and 120 dB, respectively.

Summary

- The number of **display digits** on a DMM is not related to the resolution, but can help determine the number of significant values that can be displayed and read.
- For most applications, it can be said the higher the impedance, the more accurate the voltage measurement.
- Higher **crest factors** indicate sharper peaks and make it more difficult to get an accurate AC measurement.
- **Null offset** can be used to eliminate errors caused by connections and wires when making a DC voltage or resistance measurement.
- **Auto zero** is used to compensate for internal instrument offsets.
- **Current measurements** require you to break the circuit in order to insert the DMM into the circuit loop.
- Accidentally measuring voltage while in the current mode can cause **a fuse** to blow.
- **Resistance measurements** and **continuity testing** should be taken when the circuit does not have power.
- The **normal-mode rejection ratio (NMRR)** describes the DMM's ability to reject noise that appears between the two input terminals.
- The **common-mode rejection ratio (CMRR)** describes the DMM's ability to reject noise that is common to both input terminals, such as from a noisy environment.

Generating a Signal:

Types of Function Generators, DAC Considerations, and Other Common Terminology

Overview

Learn about how signal generators generate analog signals and other topics such as types of signal generators, bit resolution, bandwidth, attenuation, digital gain, digital filtering, and analog filtering. This tutorial is part of the Instrument Fundamentals series.

Contents

Types of Signal Generators

The broad category of signal generators can include many different types of devices. At a higher level, there are two main groups: (1) signal generators, also called arbitrary/function generators and arbitrary waveform generators, and (2) logic sources, also known as pulse or pattern generators. Signal generators create waveforms with analog characteristics and logic sources generate digital waveforms that are commonly used to test computer buses. This article focuses on signal generators.

a. Function Generators

Function generators produce a limited number of predefined periodic waveforms at precise frequencies. More recent function generators employ a technology called direct digital synthesis (DDS), which gives the device the ability to produce the waveforms at precise frequencies. Function generators using DDS can change their output waveform frequency during generation with a short response time. To read more about DDS, refer to the Direct Digital Synthesis (DDS) white paper in the Instrument Fundamentals series. Function generators often have very limited memory size, because they store only a small amount of periodic waveforms. Common waveforms such as sine, square, pulse, ramp, and sweep are included in a function generator's memory; however, depending on the device, there could be more or less waveform options available. Function generators are cost-effective devices for applications such as stimulus-response testing, filter characterization, and clock source simulation, which require only periodic waveforms.

b. Arbitrary Function Generators

Arbitrary function generators (AFGs) are similar to function generators with one important additional capability: onboard memory space dedicated for a user-defined waveform. This gives you the capability to define a waveform, store it to the AFG onboard memory, and then output the waveform using DDS. Similar to function generators, AFGs also have predefined sets of waveforms stored on the device's onboard memory that can be output using DDS. Therefore, AFGs are extremely valuable devices if you are working with the same kind of applications suited for function generators, but you benefit from defining a more unique waveform than the predefined waveforms from the vendor. Before purchasing, always verify your user-defined waveform fits on the device's user-available memory.

c. Arbitrary Waveform Generators

Arbitrary waveform generators (AWGs) can produce the standard waveforms as well as large, complex, user-defined waveforms. Some AWGs also have the added capability of linking and looping combinations of waveforms to effectively produce sequences of waveforms to output. To output complex or sequenced waveforms, AWGs must use a large amount of onboard memory to store these waveforms. Therefore, if you plan to use a specific complex waveform for your application, make sure to purchase an AWG with enough memory to store the applicable waveforms. In addition to increased memory space, AWGs also employ a different clocking scheme than function generators or AFGs using DDS. An AWG's clocking scheme allows the device to output points only in the order that they are placed in memory; therefore, they cannot change the frequency of output in a short time.

Digital-to-Analog Conversion Characteristics

a. Bit Resolution

The bit resolution, or vertical resolution, of a signal generator is defined by the resolution of the digital-to-analog convertor (DAC) used. A DAC can only produce an output waveform using discrete voltage steps, or levels. You can find the number of discrete voltage levels a DAC can produce by raising two to the power of the DAC resolution. Figure 1 demonstrates the difference of varying DAC resolutions by comparing a sine wave created by a theoretical 3-bit DAC to a sine wave created by a 16-bit DAC.

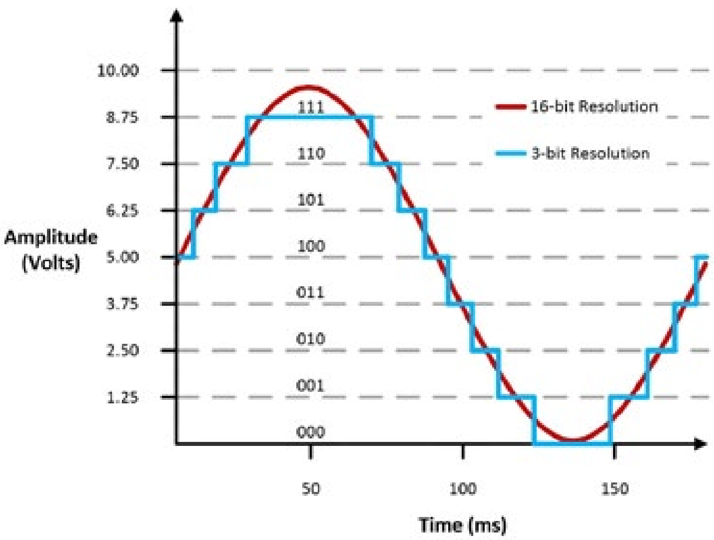


Figure 1. Differences of Two Different DAC Resolutions in Analog Signal Creation

Table 1 shows the number of discrete voltage levels each DAC can produce and Equation 1 shows how the number of discrete voltage levels for a DAC is calculated.

$$Discrete\ Voltage\ Levels\ of\ DAC = 2^{DAC\ Resolution}$$

Equation 1. Calculating the Discrete Voltage Levels for a DAC

DAC Resolution	Discrete Voltage Levels
16-bit	65,536
3-bit	8

Table 1. Discrete Voltage Levels of 3-Bit and 16-Bit DAC

The 3-bit DAC can output only eight discrete voltage levels; therefore, if the DAC had a 0 V to 10 V signal range, it can produce voltages in only 1.25 V increments, as seen in Figure 1. The 16-bit DAC can produce voltages at 152.6 μV increments and that is why the signal appears much smoother. Equation 2 shows the general formula and how the voltage increment, or commonly referred to as code width, is calculated for the 16-bit DAC.

$$Code\ Width = \frac{Device\ Output\ Range}{2^{DAC\ Resolution}}$$

$$Code\ Width = \frac{(10V - 0V)}{2^{16}}$$

$$Code\ Width = \frac{10V}{65{,}536}$$

$$Code\ Width = 152.6 * 10^{-6}$$

$$Code\ Width = 152.6\ \mu V$$

Equation 2. Code Width General Formula and Code Width Calculation Example for a 16-Bit DAC

Note that if you zoomed in to a small enough scale, the sine wave produced by the 16-bit DAC would also exhibit a stepwise appearance but with 152.6 μV increments.

b. Bandwidth

The bandwidth of an AFG or AWG describes the maximum frequency the device's analog circuitry can output without significant attenuation. The maximum frequency for the bandwidth specification is defined as the frequency where a sinusoidal output signal is attenuated to 70.7 percent of the signal's original amplitude. This frequency is also known as the -3 dB point on a bode plot.

The bandwidth specification determines the maximum frequency of sinusoidal output and other specifications such as overshoot and rise time for the instrument. This becomes critical when generating square waves or pulse signals with the signal generator. As seen in Figure 2, a signal generator with a higher bandwidth can produce square waves with smaller overshoot and faster rise times.

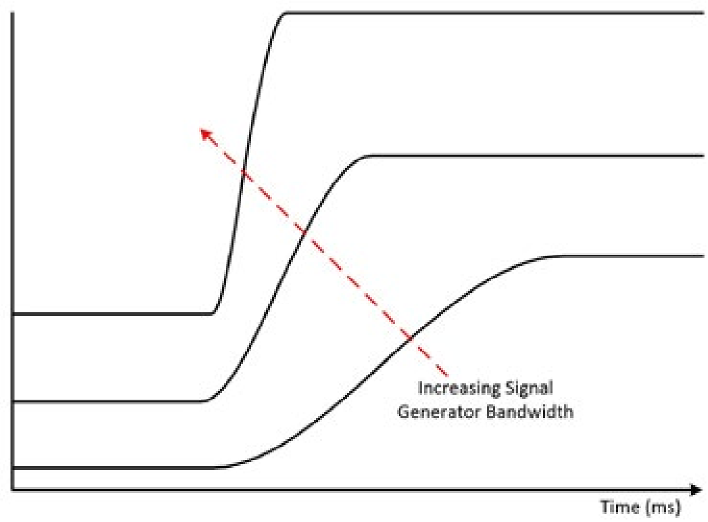


Figure 2. Higher signal generator bandwidth allows for a better representation of the signal. In this figure, the signal is a square wave.

Attenuation and Digital Gain

Signal generators are designed to produce waveforms at a variety of voltage ranges and they can switch between these voltage ranges quickly. Depending on the supported voltage ranges and how they are implemented, it is possible that a change in voltage range could require a relay switch to change the physical routing for a signal. This affects the output signal and a glitch may be observed. To accomplish this task, signal generators can employ the following techniques.

a. Attenuation

Attenuating the DAC output signal gives the signal generator the ability to alter the generated signal's amplitude while using the dynamic range of the DAC. To illustrate this, consider a situation where a 16-bit DAC with a 0 to 10 V range is used but the wanted output signal ranges from 0 to 1 V. To produce the wanted output signal, the digital data is written to the DAC at the full 0 to 10 V range, and then the analog signal at the output of the DAC is attenuated 10x. This effectively decreases the voltage resolution to 15.26 µV because the full resolution of the 16-bit DAC was used. If the 0 to 1 V signal was produced by only writing digital words to the DAC that represented values between 0 to 1 V at the 0 to 10 V range, the voltage resolution would stay at 152.6 µV as shown in Equation 2. Although attenuation uses the full resolution of the DAC, it is often a slower technique because it involves switching combinations of resistor networks.

b. Digital Gain

Digital gain is a technique that involves multiplying the waveform digital data by a factor before that data reaches DAC. Because digital gain is applied during waveform generation while the digital data is transferred from the signal generator memory, the delay associated with applying digital gain is minimal in comparison to analog gain methods. However, the output resolution of the DAC is a function of digital gain, which means that only analog gain uses the full resolution of the DAC.

Filtering and Interpolation

To generate a signal with the proper frequency, the device update rate, or sampling rate, must be twice the maximum frequency component of the generated signal. Strictly following this criterion would leave you with only a generated signal with the correct frequency, but to generate the most accurate representation of the waveform shape, the DAC operation must be accounted for. DACs use a sample-and-hold technique, which introduces high-frequency images even in a highly oversampled waveform. The sample-and-hold output can be seen visually in Figure 3 in the time domain when a sinusoid is sampled at 20 times the sine wave frequency. The sample-and-hold output gives the *stepped* waveform appearance.

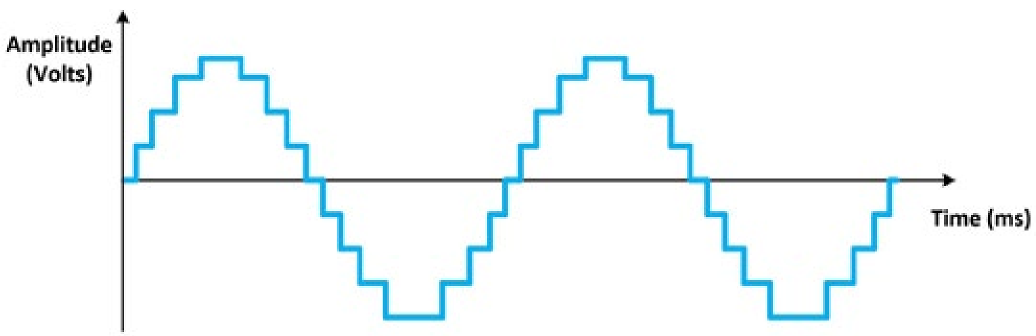


Figure 3. This time domain graph of the generated sine waveform showcases the sample-and-hold technique used by DACs.

The time domain signal still resembles a sine wave; however, inspection of the frequency domain reveals the high-frequency images created by the DAC. These images occur at integer multiples of the sampling rate plus or minus the fundamental tone. For example, a 20 MHz sine wave generated by a 100 MHz sample clock has images at 80 MHz, 120 MHz, 180 MHz, 220 MHz, and so on. Figure 4 shows the frequency domain of a generated sine wave with high-frequency images.

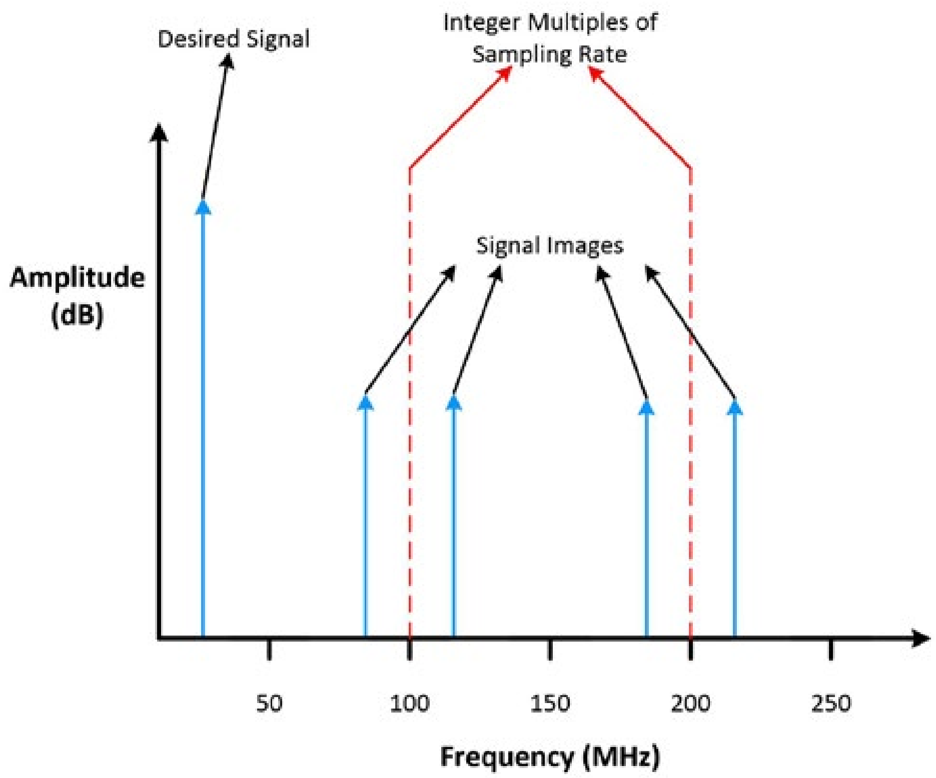


Figure 4. This frequency domain graph of the generated sine wave shows the high-frequency images.

Signal generators can use a combination of digital and analog filters to remove these images to create a more spectrally pure signal.

a. Digital Filtering and Interpolation

A signal generator can use a digital finite-impulse response (FIR) filter to provide points that interpolate between generated samples. This increases the effective sampling rate, which in turn alters the location of high-frequency images in the frequency domain. For an explanation of this concept, think of the original example of a 20 MHz sine wave generated by a 100 MHz sample clock. If the FIR filter interpolates the signal by 4x, you can now find the images using 400 MHz as the sample clock rate, thus producing images at 380 MHz, 420 MHz, 780 MHz, 820 MHz, and so on when they were originally, in Figure 4, at 80 MHz, 120 MHz, 180 MHz, 220 MHz, and so on. As Figure 5 below illustrates, interpolation does not eliminate spectral images, but it does shift these images farther from the fundamental tone.

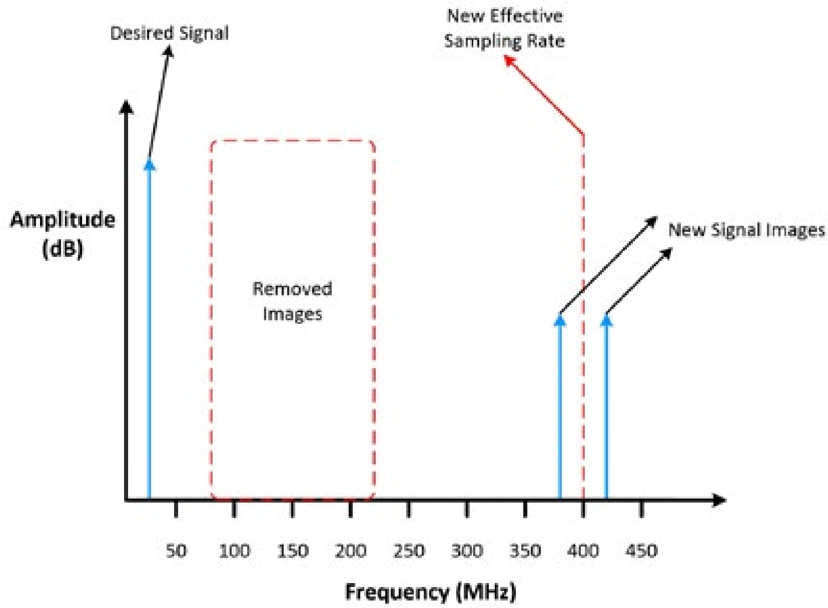


Figure 5. In this frequency domain graph of the generated sine wave, the digital filtering has moved the high-frequency images further away from the fundamental tone.

b. Analog Filtering

To produce the most spectrally pure signal, an analog filter can be applied after the interpolated signal. Because the digital FIR filter has pushed the high-frequency images farther from the fundamental tone, the requirements for the analog filter have relaxed. The analog filter does not need as steep a cut-off frequency, which would have given the circuit poor passband flatness. As seen in Figure 6, after the digital FIR filter and analog filter are applied, the high-frequency images are removed from the frequency domain.

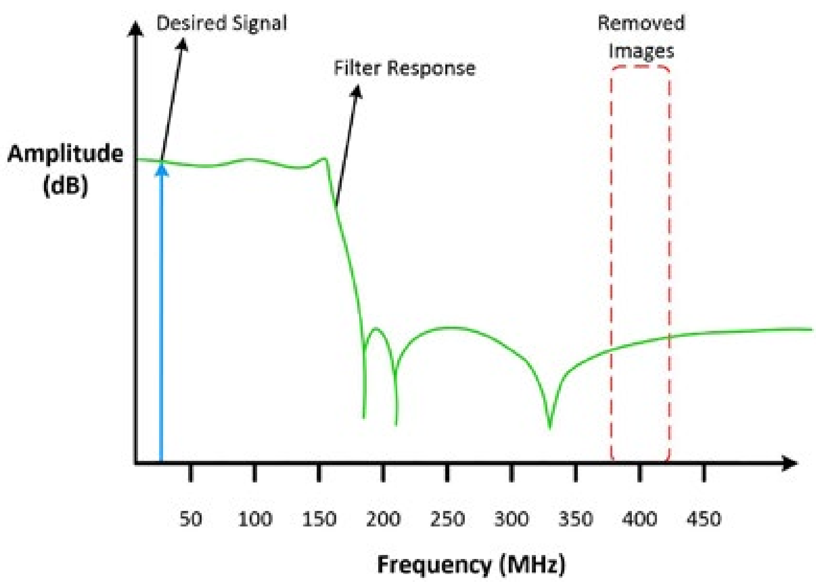


Figure 6. This is the frequency domain graph of the generated sine wave after digital and analog filtering are applied.

Because the digital FIR filter and analog filter have effectively removed the high-frequency images, you can inspect the sinusoid waveform, in Figure 7, in the time-domain once again.

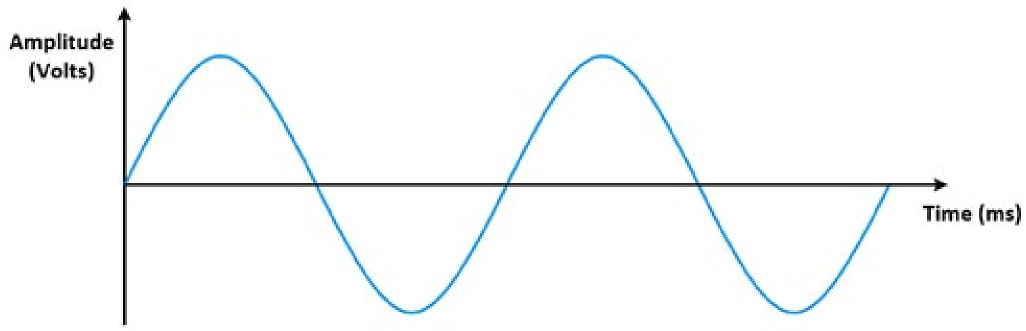


Figure 7. This is the time domain graph of the generated sine waveform after digital and analog filtering.

Notice the stepped waveform appearance created by the high-frequency images is removed and the generated sinusoid signal appears to be a more pure sinusoid waveform compared to the sinusoid in Figure 1.

You have now seen how bit resolution, bandwidth, attenuation, gain, and filtering affect the output signal from a signal generator. While looking at the specifications sheet of your signal generator, make sure to keep these specifications in mind and match them to the requirements of your application.

Summary

- **Function generators** produce a limited and predefined set of periodic waveforms at precise frequencies.
- **Arbitrary function generators (AFGs)** have the same capabilities as a function generator with the added benefit that you can use the accessible onboard memory to add a user-defined waveform.
- **Arbitrary waveform generators (AWGs)** produce the standard waveforms as well as large, complex, user-defined waveforms and they do this by using a much larger onboard memory in comparison to function generators and AFGs.
- The bit resolution, also **vertical resolution**, of a signal generator defines the number of discreet voltage levels the DAC can produce.
- **Bandwidth** describes the range of frequencies a signal generator can output. It is defined by the frequency at which a sinusoidal input signal is attenuated to 70.7 percent of its original amplitude, which is also known as the -3 dB point.
- **Attenuation** is a technique that alters the generated signals amplitude without sacrificing the dynamic range or losing digital bits of representation.
- **Digital gain** is a technique that involves multiplying the waveform digital data by a factor before the DAC. This gives the generated signal the ability to change amplitude nearly instantly; however, it may not use the full resolution of the DAC.
- **Interpolation** and **analog filtering** can be used to increase the effective sampling rate and remove the high-frequency images of a signal generated by a DAC.

Grounding Considerations for Improved Measurements

Overview

Learn how grounding can determine how a measurement system should be connected for the most accurate measurements. Article topics include grounded signal sources, floating signal sources, differential measurements, single-ended measurements, and the proper signal and measurement configuration. This tutorial is part of the Instrument Fundamentals series.

Contents

Grounding and Measurements

Measurement systems can have the ability to use different grounding configurations because signal sources can also have different grounding configurations. This capability is essential to ensure the most accurate measurement; however, this flexibility adds some difficulty when choosing the grounding configuration of the measurement system.

Figure 1 shows a block diagram of components used to make a measurement. On the right a measurement system is comprised of an instrument and signal conditioning. It should be noted that signal conditioning can be integrated in your instrument or it could be external to the instrument. On the left we have the signal source, which could be a single transducer producing a voltage from a physical phenomena or it could be a device under test. In this article we will talk about grounding on the signal source, grounding on the measurement system, and finally how to choose a measurement system configuration to ensure minimal measurement noise and error.

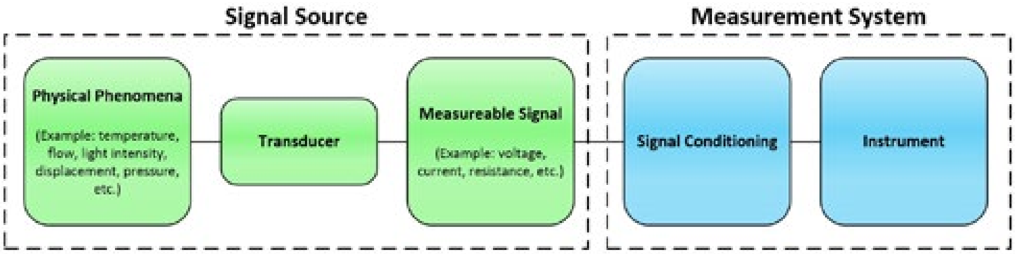


Figure 1. A signal source is fed into a measurement system that comprises an instrument and signal conditioning.

Signal Sources

There are two main categories of signal sources that should be considered for this grounding discussion; they are shown in schematic form in Figure 2.

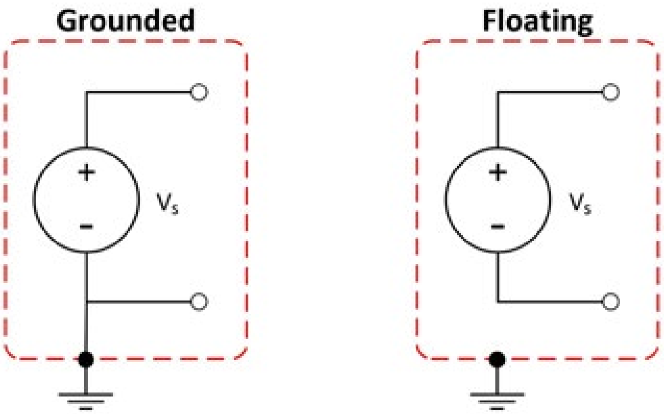


Figure 2. It is important to know if your signal is grounded or floating.

a. Grounded or Ground-Referenced Signal Sources

A grounded signal source is when a voltage signal is referenced to a system ground, such as earth or building ground. This is represented by the schematic on the left of Figure 2 above because the voltage signal has a direct electrical path to the system ground. The most common examples of grounded sources are devices that plug into the building ground through three pronged wall outlets such as signal generators and power supplies.

It is important to know that the grounds of two independently grounded signal sources are typically not at the same potential. The difference in ground potential between two systems connected to the same building ground can be 10 mV, 200 mV, or more.

b. Ungrounded or Floating Signal Sources

An ungrounded or floating signal source is one in which the voltage signal is not referenced to a system ground, such as earth or building ground. This is represented on the right in Figure 2. Note that neither the positive nor negative terminal has a direct electrical path to a ground. Some common floating signal sources are batteries, thermocouples, and transformers.

Measurement Systems

You can configure instruments in one of three modes: differential (DIFF), referenced single-ended (RSE), or nonreferenced single-ended (NRSE).

a. Differential Measurement Systems

A differential instrument requires two inputs where neither input to the instrumentation amplifier is referenced to a system ground. This is illustrated in Figure 3, where CH0+ and CH0- are wired into the positive and negative terminals of the instrumentation amplifier respectively, but they are not connected to the measurement system ground (AI GND).

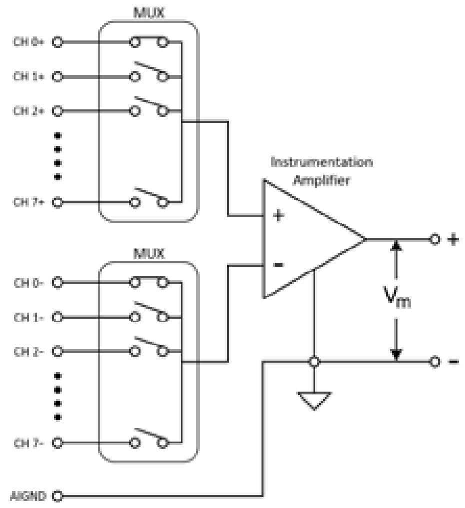


Figure 3. An ideal differential acquisition system responds to only the voltage difference between its two terminals.

An ideal differential acquisition system responds to only the voltage difference between its two terminals, the positive (+) and negative (-) inputs. The differential voltage across the circuit pair is the desired signal, yet an unwanted signal that is common to both sides of a differential circuit pair can exist. This voltage is known as common-mode voltage. An ideal differential measurement system completely rejects, instead of measures, the common-mode voltage for more accurate measurements. Practical devices, however, have limitations described by specifications such as common-mode voltage range and common-mode rejection ratio (CMRR).

The common-mode voltage range is the maximum allowable voltage swing on each input with respect to the instrument ground. To violate this constraint results in not only measurement error but also possible damage to instrument components. Here is the formula to calculate the common-mode voltage:

$$V_{CM} = \frac{V_{IN+} + V_{IN-}}{2}$$

Equation 1. Common Mode Voltage Calculation

Where:

V_{CM} = Common-mode voltage
V_{IN+} = Voltage at noninverting input terminal with respect to measurement ground
V_{IN-} = Voltage at inverting input terminal with respect to measurement ground

An example of violating the common-mode voltage range specification would be to attempt a differential measurement with one lead at 110 V and the other lead at 100 V. Although the differential measurement is 10 V, which may be within specification for the device, the common-mode voltage would be 105 V and this may not be within the specification of the instrument.

CMRR describes the ability of a measurement system to reject common-mode voltages. Amplifiers with higher CMRRs are more effective at rejecting common-mode voltages and are, therefore, more desirable for accurate measurements. The CMRR can be described as a ratio of the differential gain over the common-mode gain, seen in Equation 2. CMRR can also be described in dB as shown in Equation 3.

$$CMRR = \left|\frac{Differential\ Gain}{Common\ Mode\ Gain}\right|$$

$$CMRR_{dB} = 20log_{10}\left(\frac{Differential\ Gain}{Common\ Mode\ Gain}\right)$$

Equation 3. CMRR Expressed in dB

For example, if the instrument has a CMRR of 100,000:1 (or 100 dB) and the common-mode voltage is 5 V, you can distinguish voltage differences greater than 50 µV on the differential leads.

Common-mode rejection is critical because noise sources from the environment are present on both lines of the differential measurement. However, if the noise is present on both lines, it is cancelled out by the differential measurement. For this reason, differential configurations lead to more accurate measurements in comparison to single-ended measurements, but differential measurements require double the channel count compared to single-ended measurements.

b. Single-Ended Measurement Systems

Single-ended configurations are commonly the default configuration for instruments. They differ from differential configurations because only one analog input channel is required for the measurement. All channels on the instrument use the negative input to the instrumentation amplifier as the common reference, which can be seen in Figure 4. Because single-ended configurations use only one input, they can take twice the number of measurements compared to a differential configuration system with the same number of physical channels. On the other hand, this leaves single-ended measurements susceptible to ground loops, which can decrease the accuracy of the measurements.

Below are two different types of single-ended measurement systems:

- **Ground RSE (GRSE) or RSE systems** have the common reference channel connected to the instrument ground. In the example RSE system shown in Figure 4, the instrument ground channel is labeled AI GND.

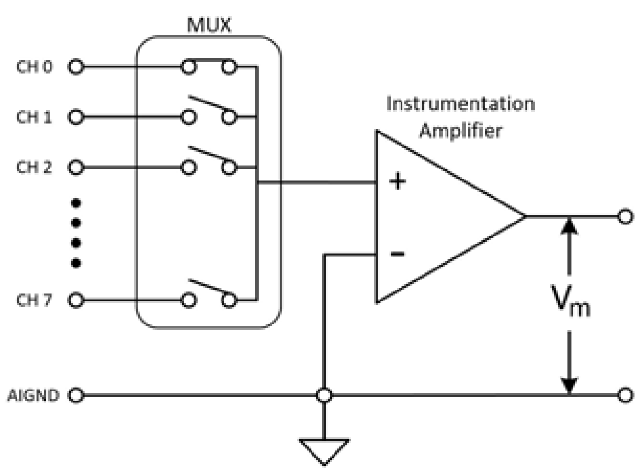
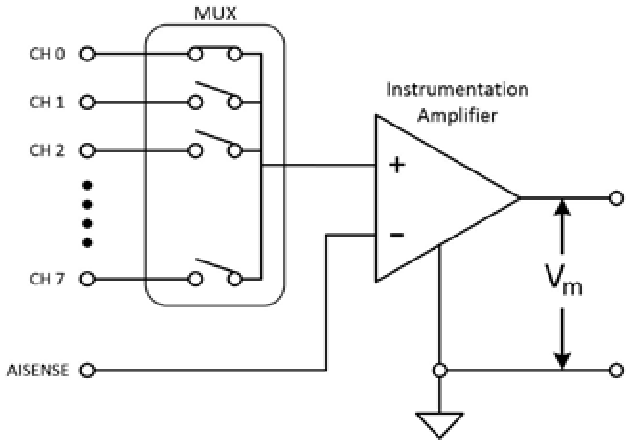


Figure 4. A GRSE or RSE system's common reference channel is connected to the instrument ground.

- **NRSE** instruments reference a common point; however, the common point is the voltage provided at the negative terminal of the instrumentation amplifier. In the NRSE example shown in Figure 5, the common reference is the AI SENSE line; therefore, the measured voltage is the potential difference between CH X and the voltage at the AI SENSE channel.



igure 5. An NRSE instrument's common point is the voltage provided at the negative terminal of the instrumentation amplifier.

Signal Source–Measurement System Configurations

After characterizing both the signal source grounding types and the instrument configurations, we will now discuss which combinations of signal sources and instrument configurations can yield the most accurate results.

a. Measuring Grounded Signal Sources

A grounded signal source is most accurately measured with a differential or NRSE instrument configuration because an additional ground is not introduced into the entire system. An additional ground added to the system can result in ground loops, which are common sources of noise in measurement applications.

Ground loops occur when two connected terminals in a circuit are at different ground potentials, causing current to flow between the two points. The ground of the signal source can be several volts above or below the ground of the instrument. This additional voltage can cause error in the measurement itself and the flowing current can also induce voltages on nearby wires causing additional measurement error. These errors can appear as scalar or periodic signals added to the measured signal. For example, if a ground loop is formed with a 60 Hz AC power line, the standard power line frequency in the United States and some other countries, the unwanted 60 Hz AC signal can appear as a periodic voltage error in the measurement.

To calculate the measured voltage, V_m, use Equation 4 below:

$$V_m = V_s + \Delta V_g$$

Equation 4. Measured Voltage With a Ground Loop Present

Where:

V_m = Measured voltage

V_s = Signal voltage

ΔV_g = Voltage difference between the signal source ground and the measurement system ground

Using Equation 4 above mathematically gives you the measured voltage when a ground loop is present. If you continue to use the 60 Hz power line example, ΔV_g is a value that changes with time instead of a scalar offset. Therefore, the measured signal looks periodic instead of like a simple offset error for the measured voltage.

Figure 6 shows what a system with a ground loop looks like in schematic form. If you are measuring the voltage source V_s with an instrument using an RSE configuration, you can simplify the schematic on the left of the equation with the schematic on the right of the equation in Figure 6, which agrees with the calculations in Equation 4.

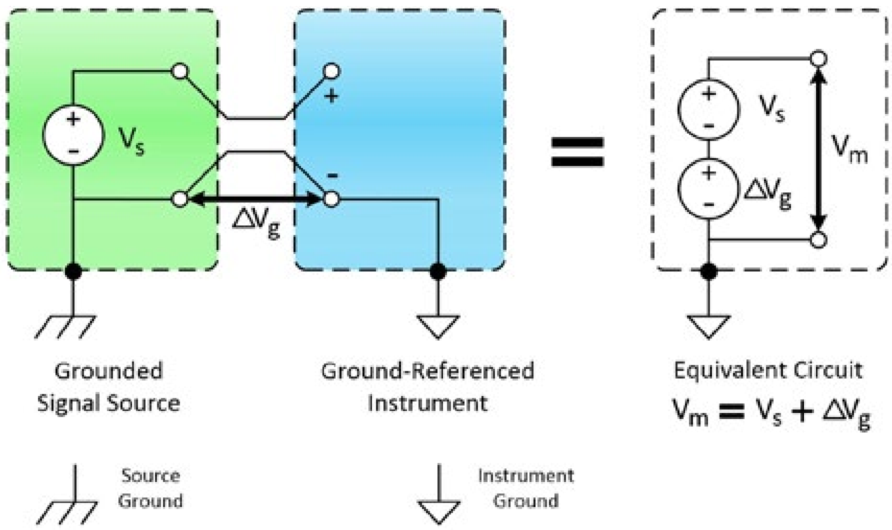


Figure 6. A grounded signal source measured with a ground-referenced system introduces ground loops and measurement error.

To avoid ground loops as shown in Figure 6, ensure only one ground reference exists in the signal source and the measurement system by using a differential or NRSE instrument configuration or by using isolated measurement hardware, which is discussed in the Isolation Types and Considerations when Taking a Measurement white paper of the Instrument Fundamentals Series.

b. Measuring Floating Signal Sources

You can measure floating signal sources with any of the measurement configurations discussed: differential, GRSE/RSE, or NRSE. Note that when using differential or NRSE measurement configurations with a floating source, you must include bias resistors from each lead, positive (+) and negative (-), to the instrument ground (see Figure 7).

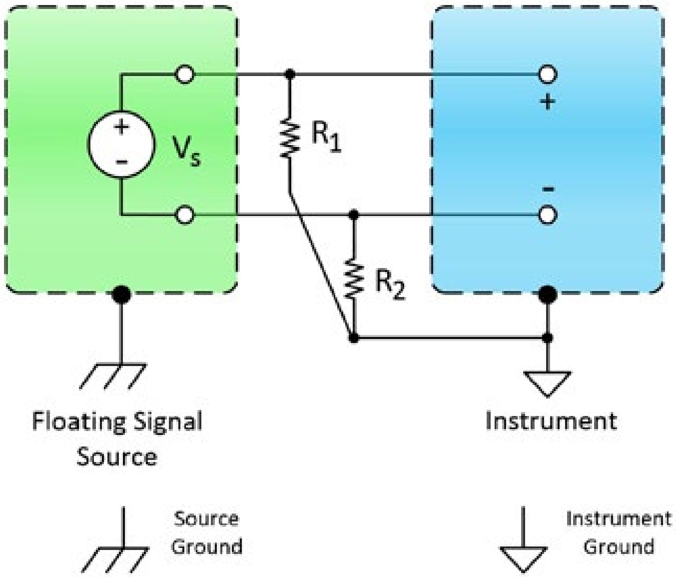


Figure 7. When measuring a floating signal source with a differential or NRSE instrument configuration, bias resistors are needed.

Bias resistors provide a DC path from the instrument amplifier inputs to the instrument amplifier ground. Bias resistors should have a high enough resistance to not load the signal source and to allow the signal source to float with respect to the instrument reference. However, the bias resistors should be small enough to keep the voltage within the range of the instrument. This typically results in bias resistors with a range of 10 kΩ to 100 kΩ to satisfy the conditions. You should always double-check the specifications guide of your device to ensure you use a bias resistor value that is within the suitable range.

If bias resistors are not used in a differential or NRSE configuration when measuring floating signal sources, the measured signals can be unstable or at positive or negative full-scale range of the instrument.

When a GRSE/RSE configuration is used to measure a floating signal source, bias resistors are not necessary. To get the best measurement results when using single-ended instrument configurations, the following is recommended:

- The input signals are equal to or greater than 1 V.
- Signal cabling is relatively short and travels through a noise-free environment (or is properly shielded).
- All input signals can share a common, stable, and known reference signal–generally a point in the system where the voltage is at 0 V.

For a summary of the recommended combinations of signal sources and instrument configurations, refer to Figure 8.Grounding and Measurements

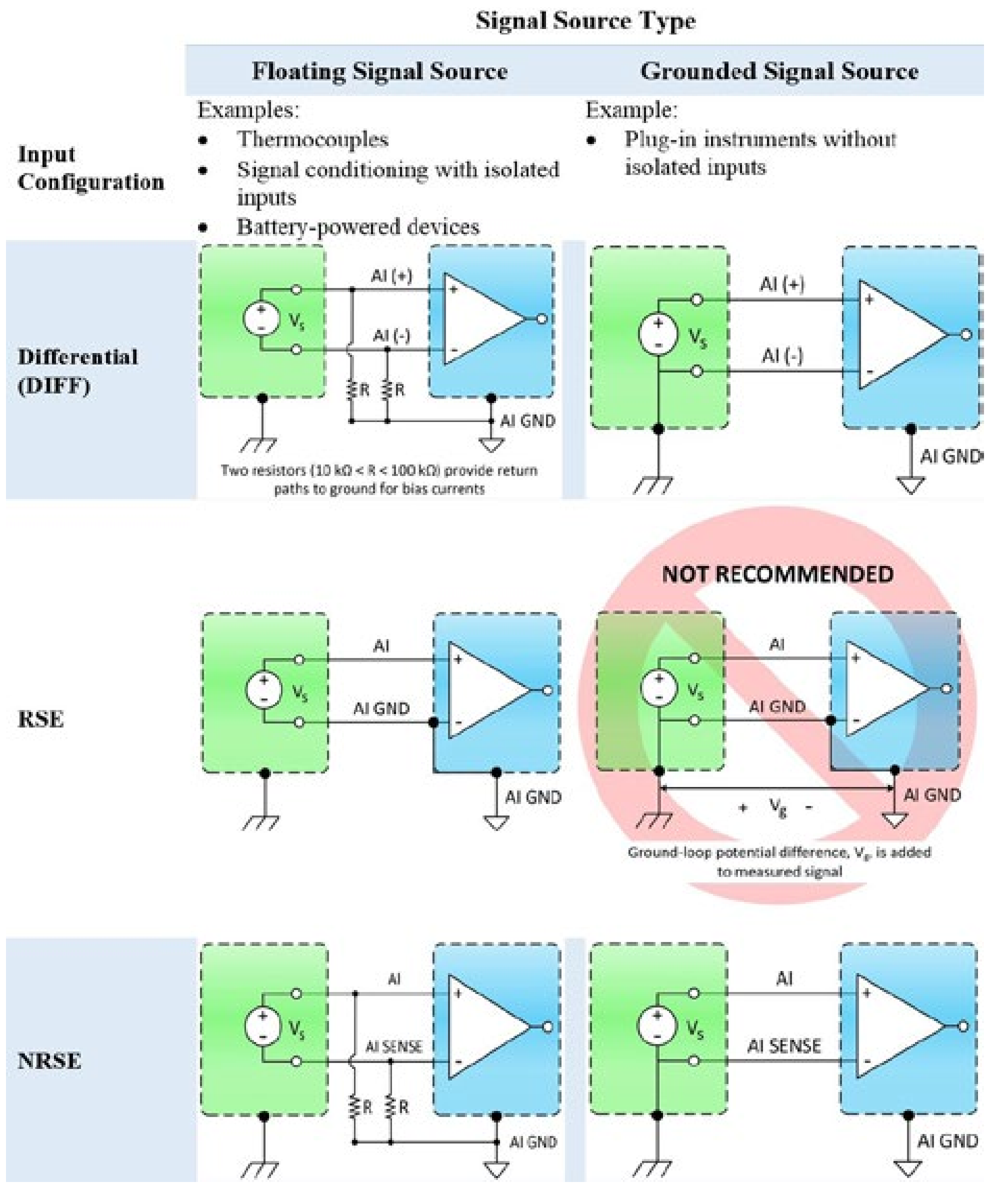


Figure 8. Instrument Configuration Versus Signal Source Type Summary.

Summary

- Measurement systems include an instrument and signal conditioning. Depending on the instrument, signal conditioning can be a part of the instrument or external.
- Two main categories of signal sources:
  - **Grounded signal source:** Signal has a direct electrical path to ground
  - **Floating signal source:** Signal does not have a direct electrical path to ground
- Instruments can have three main measurement configurations:
  - **Differential:** A measurement that takes two input channels and is the most accurate configuration because it removes common-mode voltages
  - **Ground referenced single-ended (GRSE)** or **referenced single-ended (RSE):** A measurement that uses only one channel and the instrument ground; however, this single-ended measurement type is susceptible to noise
  - **Nonreferenced single-ended (NRSE):** A type of measurement that uses only one channel and a common reference point, which is not ground; however, this system is more susceptible to noise in comparison to differential measurements
- Differential or NRSE instrument configurations are recommended to measure a grounded signal source.
- Differential, GRSE/RSE, or NRSE instrument configurations are recommended configurations to measure a floating signal source.
  - **Bias resistors** must be used in differential or NRSE instrument configurations to measure a floating signal source.

Isolation Types and Considerations When Taking a Measurement

Overview

Learn about isolation topologies used in instruments and the positive benefits that isolation can provide. Article topics include ground loops, common-mode voltage, isolation topologies, analog isolation, digital isolation, and isolation types. This tutorial is part of the Instrument Fundamentals series.

Contents

What Is Isolation?

Isolation is a method of physically and electrically separating two distinct parts of an instrument. When the term isolation is used with instruments, it most likely refers to electrical isolation, which means that current does not flow between the two parts of the system that are isolated from each other. There are several advantages of electrical isolation but one of the largest advantages, in regards to measurement accuracy, is that isolation breaks ground loops.

Isolation also uses the physical and electrical barriers to provide safety benefits by keeping high voltages or high transient voltages away from the user or away from important circuit components, which we will discuss in later sections.

First, here's a quick review of ground loops, which are covered in more detail in the Grounding Considerations for Improved Measurements white paper in the Instrument Fundamental Series.

a. Ground Loops

Ground loops are the most common source of noise in acquisition applications. They occur when two connected terminals in a circuit are at different ground potentials, which causes current to flow between the two points. This potential difference causes error in the measured voltage, V_m, which can be calculated using Equation 1.

$$V_m = V_s + \Delta V_g$$

Equation 1. Measured Voltage With a Ground Loop Present

Where:

V_m = Measured voltage

V_s = Signal voltage

ΔV_g = Voltage difference between the signal source ground and the instrument ground

The Grounding Considerations for Measurements white paper discusses how to eliminate ground loops by ensuring only one ground reference exists in the signal source and measurement system setup. However, using isolated hardware also removes ground loops, because it eliminates the path for current to flow between the ground of the signal source and the ground of the measurement system.

Isolation Topologies

In general, there are three different types of isolation topologies, from a low level of protection to a high level of protection, respectively:

- Channel-to-earth isolation
- Bank (channel-to-bus) isolation
- Channel-to-channel isolation

a. Channel-to-Earth Isolation

This is the lowest protection level of isolation for an instrument. See Figure 1 for a schematic of channel-to-earth isolation. The voltages present at AI 1, AI 2, and AI Ground are not isolated from each other; however, they are isolated from the instrument ground. This isolation topology breaks ground loops between AI 1 and the earth ground, but it is possible that a current present on AI 1 could induce a voltage on AI 2, because they are not isolated from each other.

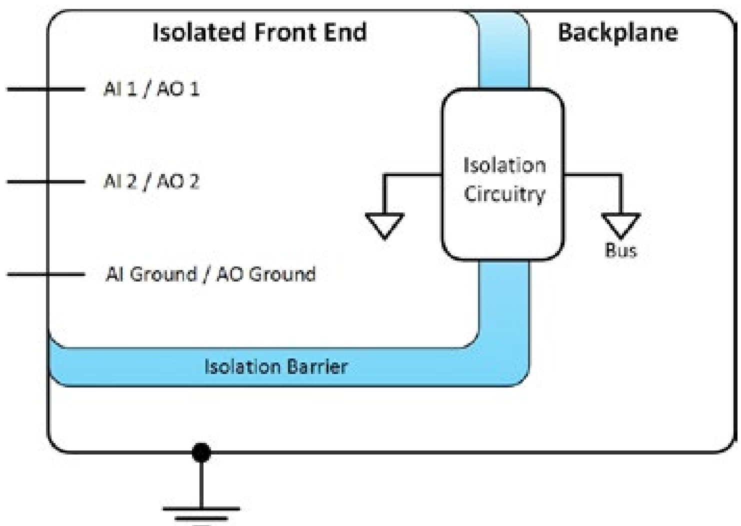


Figure 1. Channel-to-earth isolation does not isolate channels from each other but does isolate the channels from instrument ground.

b. Bank (Channel-to-Bus) Isolation

In bank isolation, also known as channel-to-bus isolation, several physical lines are built into groups called banks. See Figure 2 for this architecture. Because isolation barriers exist between channels in different banks, the ground loop protection is high between banks. However, it is still possible in this topology that signals on channels within a bank can affect each other.

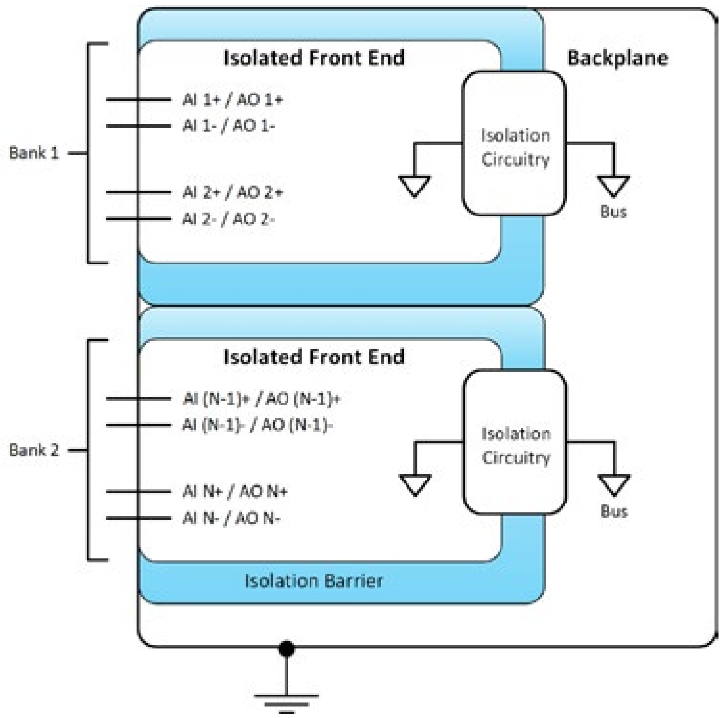


Figure 2. In bank isolation, the ground loop protection is high between different banks.

c. Channel-to-Channel Isolation

This topology provides the most comprehensive protection for the signals on the instrument lines because not only are all channels isolated from earth ground, but each channel is also isolated from all other individual channels. See this topology in Figure 3.

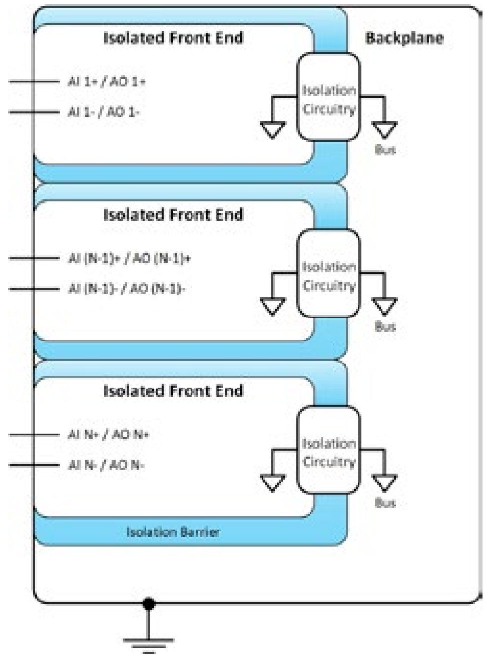


Figure 3. In channel-to-channel isolation, each channel is isolated from all other individual channels.

Analog Versus Digital Isolation

Analog input or output channels can be isolated using two different methods regardless of the instrument isolation topology. The difference between the two methods lies in the location of the isolation circuitry in the instrument. Analog isolation is where the isolation circuitry is in the path prior to the analog-to-digital convertor (ADC) and it acts on the analog signal. Digital isolation is where the isolation circuitry is after the ADC, because it acts on the newly digitized data.

a. Analog Isolation

An isolation amplifier is one of the more common parts used to provide isolation in the analog front end of an instrument. As shown in Figure 4, the analog data passes from the sensor into the I/O connector through the gain amplifier into the isolation amplifier and then to the ADC.

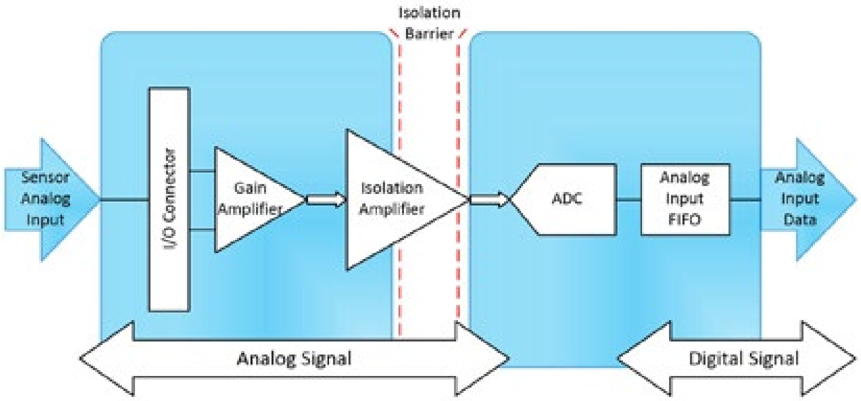


Figure 4. An isolation amplifier is one of the more common parts used to provide isolation in the analog front end of an instrument.

One large benefit of analog isolation is that it protects the ADC. Because the isolation is provided before the ADC, the ADC is less likely to be damaged by transient or high voltages. Analog isolation does, however, have disadvantages. First, because analog isolation is not perfect and it lies before the ADC, it can add gain, nonlinear, or offset error to the analog signal before it reaches the ADC. This is not ideal and can decrease the accuracy of the measurement. In addition, analog isolation components can introduce longer settling times and are often more expensive than their digital isolation counterparts.

b. Digital Isolation

As opposed to analog isolation, digital isolation circuitry is placed after the ADC in the instrument, as shown in Figure 5.

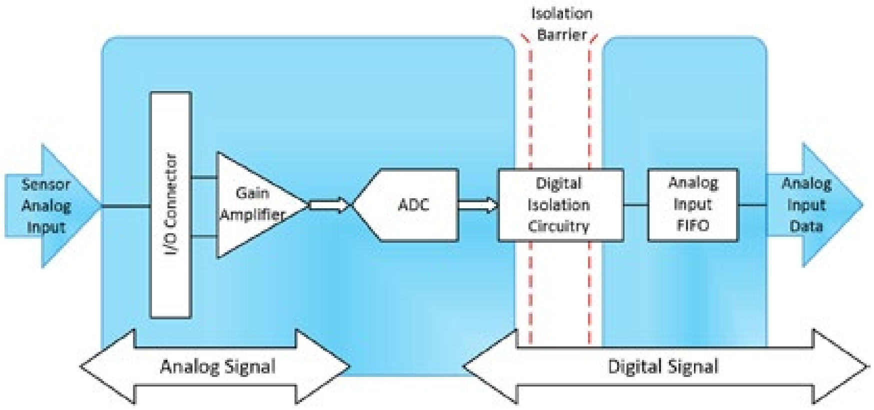


Figure 5. As opposed to analog isolation, digital isolation circuitry is placed after the ADC in the instrument.

Digital isolation can lead to better performance and accuracy, in comparison to analog isolation circuitry, because the measured signal is less altered before it is digitized by the ADC. Digital isolation circuitry also has advantages over analog isolation circuitry because it is typically lower in overall cost and it performs at higher data transfer speeds. However, because digital isolation circuitry is after the ADC, the ADC is more susceptible to the damage a voltage spike can cause.

Isolation Types

We have talked about common isolation topologies for instruments and where the isolation can be applied to the signal within the instrument, but we have not talked about the isolation barrier itself or how the signal crosses the isolation barrier. In this section we will quickly cover the isolation barrier and then we will move into three common isolation types, which use different techniques to transmit the signal data across the isolation barrier.

Physical isolation is the most basic form of isolation, meaning that there is a physical barrier between two electrical systems. This can be in the form of insulation, an air gap, or any nonconductive path between two electrical systems. With pure physical isolation, you can imply that no signal transfer exists between electrical systems. When dealing with isolated measurement systems, the signal of interest needs to cross the isolation barrier with the benefits of removing ground loops. Therefore, you must have a transfer, or coupling, of the signal's energy across the isolation barrier. Three common techniques of transferring the signal across the isolation are discussed below.

a. Capacitive Isolation

Capacitive isolation, as seen in Figure 6, uses an electrical field as the form of energy to transfer the signal across the isolation barrier. The electric field changes the level of charge on the capacitor. This charge is detected across the isolation barrier and the charge detected is proportional to the level of the measured signal.

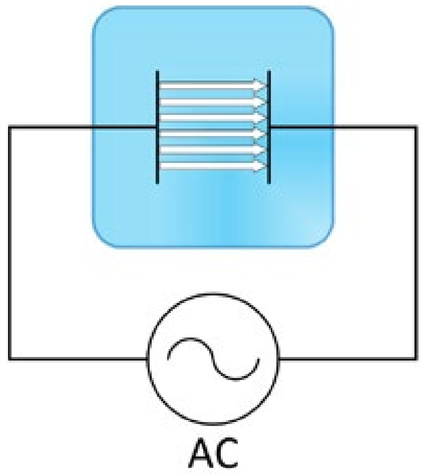


Figure 7. Inductive isolation uses a transformer, notated with the above symbol, to transfer a signal across an isolation barrier.

b. Inductive Isolation

Inductive isolation uses a transformer, shown in Figure 7, to transfer a signal across an isolation barrier. The transformer generates an electromagnetic field, proportional to the measured signal, as the form of energy to cross the isolation barrier.

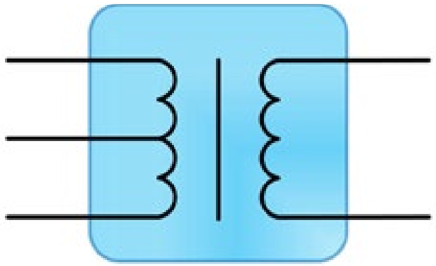


Figure 6. Capacitive isolation uses an electrical field as the form of energy to transfer the signal across the isolation barrier.

As in capacitive coupling, inductive isolation can provide relatively high-speed data transmission rates. In addition to high-speed transmission, inductive coupling uses low power for the data transmission. However, inductive coupling is susceptible to interference from surrounding magnetic fields because it uses electromagnetic fields as the method to cross the isolation barrier. If external magnetic fields do interfere with the electromagnetic field produced by the transformer, this could affect the accuracy of the measurement.

c. Optical Isolation

Optical isolation uses an LED and a photodetector to transmit the signal information across the isolation barrier. The isolation barrier in optical isolation is typically an air gap and the signal is transmitted using light. The light intensity produced by the LED is proportional to the measured signal.

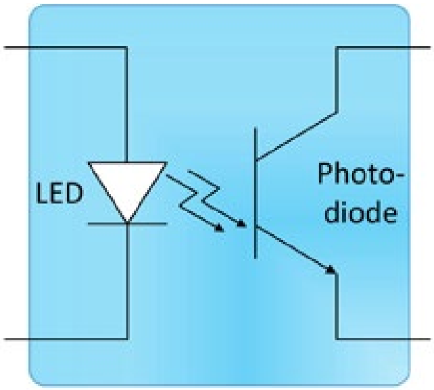


Figure 8. Optical isolation uses an LED and a photodetector to transmit the signal information across the isolation barrier.

Because optical isolation uses light as the energy to transfer the measured signal across the isolation barrier, it gains the advantage of immunity from electrical- and magnetic-field interference. This can make optical isolation an effective technique in industrial areas where strong electric or magnetic fields could be present. The advantages gained by using light are balanced by some disadvantages. Optical isolation typically has slower data transfer rates, which are limited to the LED switching speed. It also has relatively high power dissipation when compared to capacitive and inductive isolation.

Summary

- **Isolation** is a method of physically and electrically separating two distinct parts of an instrument.
- Breaking **ground loops** to measure the signal of interest more accurately is a main advantage of isolated measurement systems.
- Based on the application requirements, you can choose the isolation topology that best fits system needs.
  - **Channel-to-earth isolation** isolates the channels from the instrument ground.
  - **Bank (channel-to-bus) isolation** isolates groups (banks) of lines from other groups of lines as well as from the instrument ground.
  - **Channel-to-channel isolation** isolates every line from every other line present and from the instrument ground.

- **Analog isolation** circuitry protects the ADC from high voltages and transient voltages, but it can add gain, nonlinear, and offset errors to the signal before it reaches the ADC.
- **Digital isolation** circuitry does not protect the ADC, and its advantages over analog isolation include lower cost, higher data transmission speeds, and greater accuracy because the signal is less altered prior to reaching the ADC.
- Isolation type overview:

Isolation Type	Advantages	Disadvantages
Capacitive	• Fast data transmission rate • Magnetic field interference immunity	• Susceptible to electric field interference
Inductive	• Fast data transmission rate • Electric field interference immunity	• Susceptible to magnetic field interference
Optical	• Electric field interference immunity • Magnetic field interference immunity	• Slower data transmission rates • Relatively high power dissipation

Power Supply Fundamentals:

Modes of Operation, Remote Sense, Ripple, and Noise

Overview

Learn about programmable DC power supply basics, including constant voltage mode, constant current mode, remote sense, ripple, noise, isolation, rise time, settling time, and transient response. This tutorial is part of the Instrument Fundamentals series.

Contents

What Is a Programmable DC Power Supply?

Commonly used in research, design, development, and production applications, a DC power supply is an instrument that can source DC power to a connected device. A device connected to a power supply can be referred to as a load, device under test (DUT), or unit under test (UUT), depending on the context. To characterize a DUT or test if a DUT is working as expected, many DC power supplies have the ability to simultaneously source power and measure the voltage or current consumed by the DUT. Typically, power supplies provide a constant current or constant voltage and monitor the resulting voltage drop or current draw. A programmable DC power supply can be automated using a computer to communicate with the device. Some programmable DC power supplies can store output sequences or measurements in onboard memory, while others can handle only immediate actions.

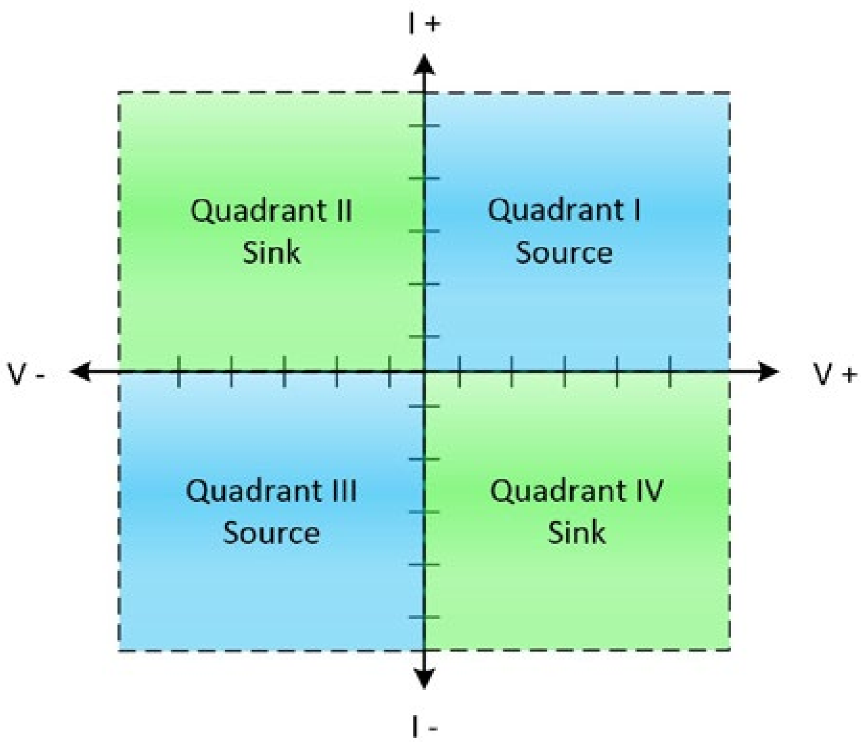


Figure 1. Most DC power supplies operate in quadrant I, providing positive voltage and positive current, or quadrant III, providing negative voltage and negative current.

Referring to the I-V diagram in Figure 1, most DC power supplies operate in Quadrant I, providing positive voltage and positive current, or Quadrant III, providing negative voltage and negative current. The formula to calculate DC power is $P = V \times I$. In Quadrant I, voltage and current are both positive; in Quadrant III, voltage and current are both negative. In both cases, plugging the numbers into the power formula results in a positive power output, which is called sourcing. Operating in Quadrant II and IV results in a negative power output, which is called sinking. When sourcing, the power is generated in the supply and dissipates in the DUT. When sinking, the power is generated in the DUT and dissipates in the supply.

Some devices called source measure units (SMUs) can operate in all four quadrants, sourcing and sinking power. You can think of an SMU as an ideal rechargeable battery. When you connect the battery to the charger, the battery draws, or sinks, power from the charger. Then when you disconnect the battery from the charger and use it to power a flashlight, the battery becomes a source that provides power for the light bulb. SMUs are commonly used to characterize batteries, solar cells, power supplies, DC-DC converters, or other power-generating devices.

Another differentiating factor between a DC power supply and an SMU is precision. Some applications are especially demanding and require higher precision than a typical power supply offers. It is common for SMUs to have high precision in the μV or pA range, which is why they are often preferred when the accuracies of sourced and measured values are important and the application requires sensitivity beyond that of a typical power supply. Precision is covered in greater detail in the Analog Sample Quality: Accuracy, Sensitivity, Precision, and Noise white paper, and you can learn more about SMUs by reading What Is a Source Measure Unit?

Constant Voltage and Constant Current Modes

In addition to understanding the differences between sourcing and sinking power, it is also important for you to understand the difference between constant voltage mode and constant current mode. Programmable DC power supplies can operate in either constant voltage mode or constant current mode, depending on your desired output levels and load conditions.

a. Constant Voltage Mode

In constant voltage mode, which is sometimes referred to as voltage-controlled mode, a power supply behaves like a voltage source, holding the voltage across the output terminals constant while the current output varies, depending on load conditions. If your load resistance changes, Ohm's law (V = I x R) dictates that the supplied current must also change proportionally to maintain the power supply output voltage level. If a DUT's resistance suddenly drops, then the power supply increases the current to hold the voltage constant.

When using a programmable DC power supply, you can set the desired current limit. If your load attempts to draw more current than the programmed current limit allows, then the power supply begins to operate in compliance, meaning that the power supply is unable reach the requested output voltage level without violating the user-programmed current limit in place. At this time, the power supply switches to constant current mode and the current is held at the current limit. This pivotal load resistance level is referred to as the compliance resistance, which can be calculated by dividing the voltage setpoint by the current limit. Other common names for compliance resistance are critical resistance and crossover resistance.

For example, suppose you want to supply a constant 5 V (VS = 5 V) to your DUT, which typically provides a 50 Ω load resistance (RL = 50 Ω). Additionally, you decide to limit the current output to 300 mA (IS = 0.3 A) to prevent damage to the DUT. Using the compliance resistance formula (RC = VS / IS), you calculate that 16.67 Ω is the minimum load resistance to keep the output operating in constant voltage mode. If your load resistance fluctuates, but stays above 16.67 Ω, then your power supply continues to provide a constant 5 V. If the DUT fails, dropping the load resistance below 16.67 Ω, then the power supply begins to operate in compliance, switching to constant current mode and outputting a steady 300 mA at a voltage level less than 5 V.

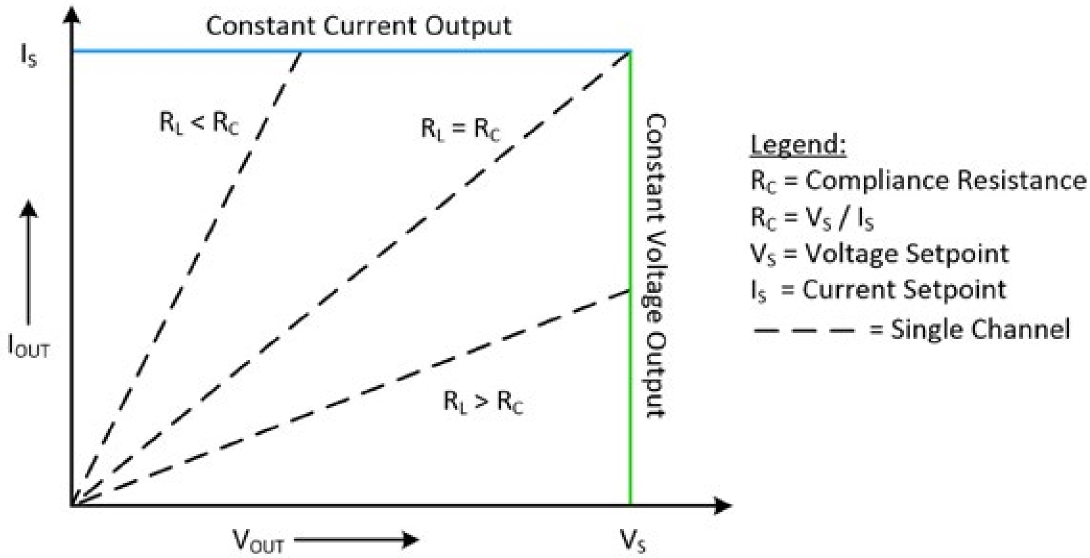


Figure 2. When outputting a constant voltage, you can set a current limit to protect the DUT.

b. Constant Current Mode

Constant current mode is essentially the opposite of constant voltage mode. In constant current mode, also known as current-controlled mode, the power supply behaves like a current source, holding the current flowing through the output terminals constant while the output voltage varies depending on load conditions. Referencing Ohm's law, if your load resistance changes, then the voltage must also change appropriately to maintain a constant current. If the DUT from the previous example fails and causes the load resistance to drop, then the power supply decreases the output voltage proportionally to hold the current constant. For example, constant current operation is desirable when controlling LEDs that can be damaged by high current.

Constant current mode is also limited by a configurable voltage limit, imposing a compliance resistance similar to constant voltage mode. You can use the same calculation used in the Constant Voltage Mode section to calculate your compliance resistance for constant current operations. However, for constant current mode, your load resistance must stay below the compliance resistance to maintain the desired constant current. Figure 2 illustrates the compliance resistance concept for both constant voltage and constant current mode.

A unique application that requires both constant voltage and constant current operation is charging a Lithium-Ion battery, which is a common type of rechargeable battery that is used in portable electronics because of its energy density, lack of memory effect, and slow loss of charge when not in use. To recharge a Lithium-Ion battery, the power supply should apply a constant current, monitoring the battery voltage level until the battery reaches its maximum voltage. After the Lithium-Ion battery is fully charged, the power supply should switch to constant voltage mode, which provides the minimum current required to hold the battery at its maximum voltage

Taking Measurements With a Programmable DC Power Supply

A key feature on most programmable DC power supplies is the ability to measure the generated current and voltage. This feature is essential for many applications such as I-V curve tracing, where current draw must be measured for multiple voltage setpoints. The measurement operation of a programmable DC power supply is similar to the measurement capabilities of a digital multimeter (DMM). As with any measurement device, there is a trade-off between the speed at which you perform measurements and the amount of noise in those measurements. Key measurement concepts include accuracy, aperture time, auto zero, guarding, remote sense, input ranges, resolution, and sensitivity. For more information on these topics, see the DMM Measurement Types and Common Terminology and Analog Sample Quality: Accuracy, Sensitivity, Precision, and Noise white papers included in the Instrument Fundamentals series.

Remote Sense

A challenge in accurately sourcing or measuring precise voltages is the effect that lead resistance has on the voltage that a DUT sees. Lead resistance is always present but can become problematic when using very long, small-gauge wires. Table 1 provides the typical resistances of different gauges of copper wire. Although typically no larger than a few ohms, these small resistances can have a large affect on the voltage a DUT receives, especially when the internal resistance of the DUT is small.

Typical Resistance of Copper Wire

AWG Rating	mΩ/m (mΩ/ft)
10	3.3 (1.0)
12	5.2 (1.6)
14	8.3 (2.5)
16	13.2 (4.0)
18	21.0 (6.4)
20	33.5 (10.2)
22	52.8 (16.1)
24	84.3 (25.7)
26	133.9 (40.8)
28	212.9 (64.9)

Table 1. Lead resistance from wire can have a large affect on the voltage a DUT receives.

Figure 3 shows a diagram of a generic circuit that consists of a power sourcing instrument, lead wires, and a DUT. In this case, the leads are 24-foot-long 26 AWG copper wires, with a resulting lead resistance of approximately 1 Ω for both the positive and negative lead wires connecting the power source to the DUT. The current coming out of the power supply causes a voltage drop across Rlead1 and Rlead2, resulting in the voltage across RDUT being less than V_{source}.

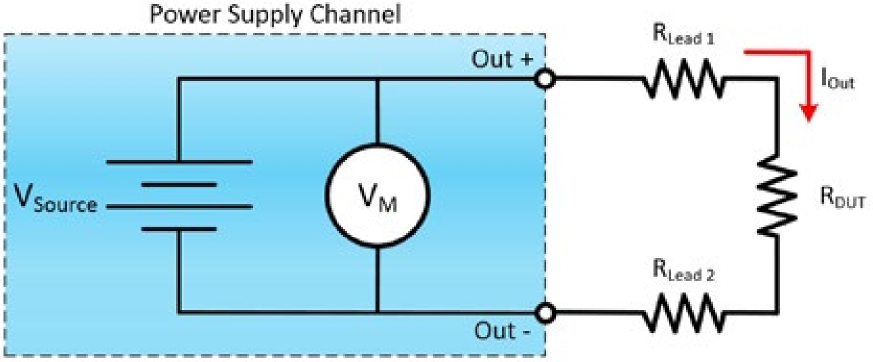


Figure 3. This shows an example connection diagram for a typical programmable DC power supply that can be used to calculate the voltage a DUT receives.

Assuming that the power source is set to an output of 5 V and the DUT has an impedance of 1 kΩ, you can calculate the actual voltage seen at the terminals of the DUT using the following equation.

$$V_{DUT} = V_{Supply} \times \frac{R_{DUT}}{R_{DUT} + 2R_{Lead}}$$

For the initial case, the voltage seen is actually just 4.99 V. For some devices, this small change is not an issue; however, for applications that require precise characterization based on operating voltage, this error can become critical. Furthermore, for devices that have lower input impedances and therefore draw a lot of current, the actual voltage at the DUT can be substantially lower than the voltage at the output of the power supply. Table 2 lists the values that the example DUT sees based on lower values of its input impedance.

DUT Impedance	DUT Voltage
1 kΩ	4.99 V
100 Ω	4.9 V
10 Ω	4.16 V

Table 2. For devices with lower input impedances, the voltage observed at the DUT can be substantially lower than the voltage at the output of the power supply because of lead resistance.

The solution to lead-resistance-induced voltage error is remote sensing, also known as 4-wire sensing. This technique accounts for the voltage drop across the lead resistance by measuring the voltage directly at the DUT and compensating accordingly. This method is similar to the way that DMMs perform 4-wire resistance measurements to remove the effect of lead resistance from resistance measurements. Most power supplies, SMUs, and DMMs feature two extra terminals on the output to allow for this 4-wire remote sensing technique, and these extra terminals are connected directly at the DUT, as shown in Figure 4. Although there is still lead resistance in the wires used for remote sensing, voltage measurements are high-impedance so no current flows through the sense wires and no voltage drop is seen.

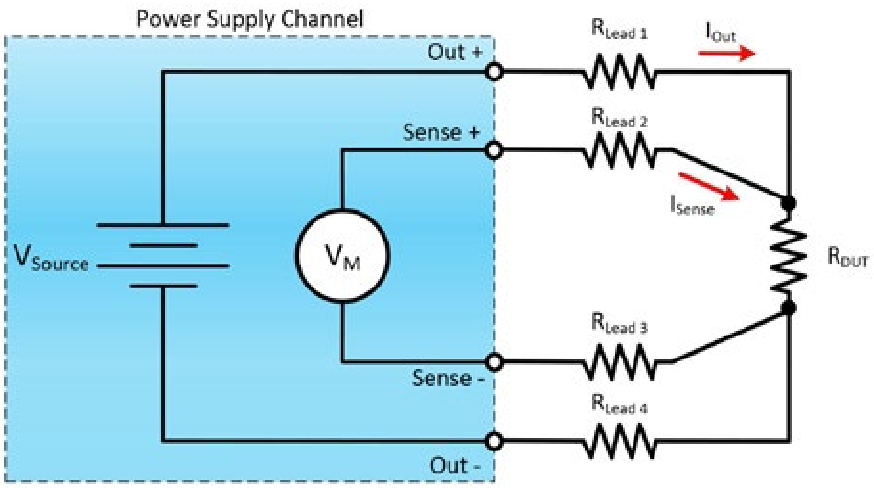


Figure 4. Remote sense is a 4-wire connection technique that can eliminate the effects of lead resistance.

Common Specifications of DC Power Supplies

a. Ripple and Noise

When considering which programmable DC power supply to use in your application, it is important to consider output ripple and noise, which is sometimes referred to as Periodic and Random Deviation (PARD). True noise is random and spread across all frequencies when viewed in the frequency domain, whereas ripple is typically periodic. Ripple is introduced by the AC-to-DC rectification required to convert the AC power from the wall outlet to the desired DC levels. Depending on the type of regulation used by a power supply, ripple has either one or two fundamental frequencies.

DC power supplies typically use either linear or switching regulation to convert the 50/60 Hz AC power source to a DC power signal. Linear regulation power supplies use an AC-to-DC transformer to convert the line voltage to a stable DC output. Therefore, the voltage output of a linear regulated power supply generally has a 50/60 Hz low-frequency ripple in addition to any extra noise present. Linear regulated power supplies typically have low ripple and noise, but they also have low efficiency, large size, and produce more heat. On the other hand, switching power supplies convert the 50/60 Hz current to a much higher frequency, resulting in some periodic, high-frequency ripple in addition to the 50/60 Hz low-frequency ripple. Switching power supplies are typically more compact, produce less heat, and are more efficient, but they are very susceptible to high-frequency noise. Figure 5 shows an illustration of a high-frequency ripple and random noise.

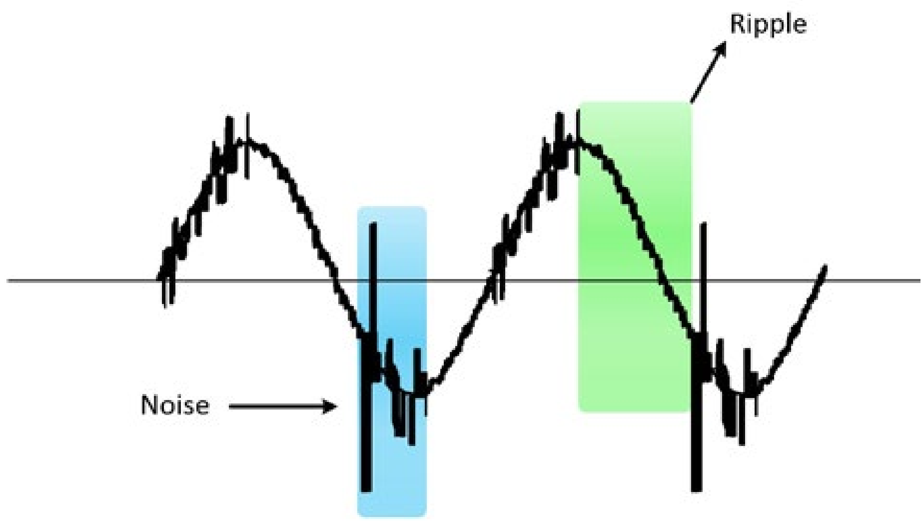


Figure 5. In power supplies, noise typically is random and spread across all frequencies, where as a ripple is periodic.

Additionally, transmissions from programmable DC power supplies can be affected by environmental noise, which adds to any inherent system noise. To reduce the effects of environmental noise, it is important to use shielded, twisted pair wires when possible.

b. Rise Time and Settling Time

Rise time and settling time are key indicators of a power supply's ability to reach the desired voltage level and stabilize. Specifically, rise time is the amount of time required for the output to transition from 10 percent to 90 percent of the configured output. Settling time describes the amount of time it takes for an output channel to stabilize within a specified percentage of its final value, including the rise time. Figure 6 illustrates both rise time and settling time for a power supply output changing from 0 V to 10 V.

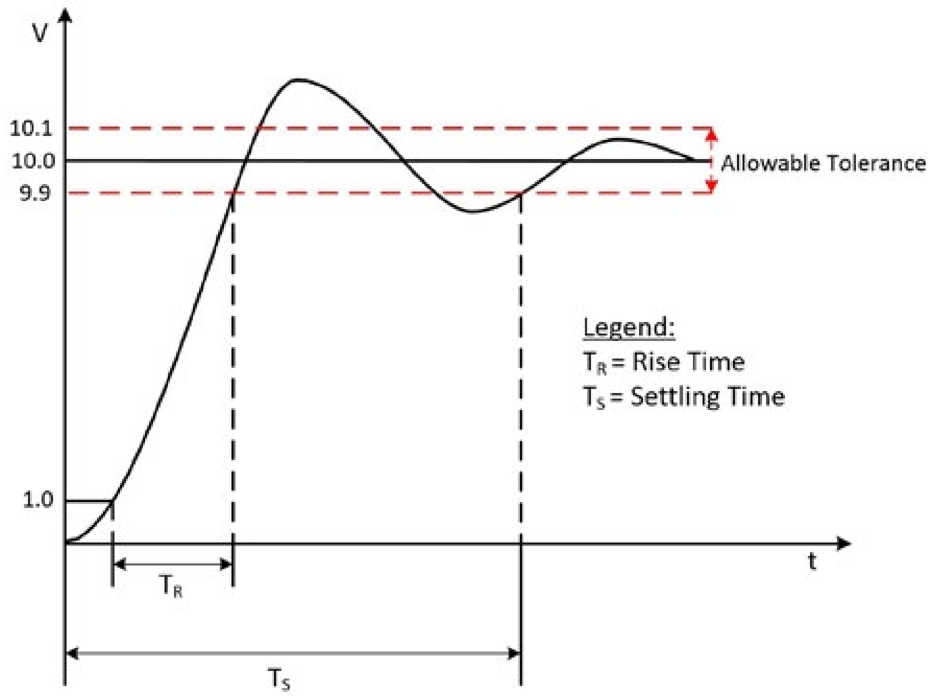


Figure 6. Rise time and settling time are key indicators of a power supply's ability to reach the desired voltage level and stabilize.

Rise time and settling time are important power supply specifications because they can directly affect measurement time, requiring extra time to wait for the circuit to recover from the transient before you can take the next measurement. Measurement time is especially important for situations, such as in automated test systems, where reducing measurement time can also reduce your overall cost.

c. Transient Response

A transient response typically describes the response of a system to a change from equilibrium. For a DC power supply, transient response describes how a power supply operating in constant voltage mode responds to a sudden change in load current. Changes in load current, such as a current pulse, can cause large voltage transients as shown in Figure 7. As the power supply's internal control circuit compensates for the change in load current, the voltage settles back at the desired level. The transient response of a power supply specifies how long it takes for the transients to recover within a certain percentage of the voltage setting. Typically, the transient response is specified as the amount of time required to recover to a percentage of the voltage setpoint after a 50 percent change in load current. For example, a device might be capable of recovering to 0.1 percent of the original voltage setpoint within 50 μs after a 50 percent change in load current.

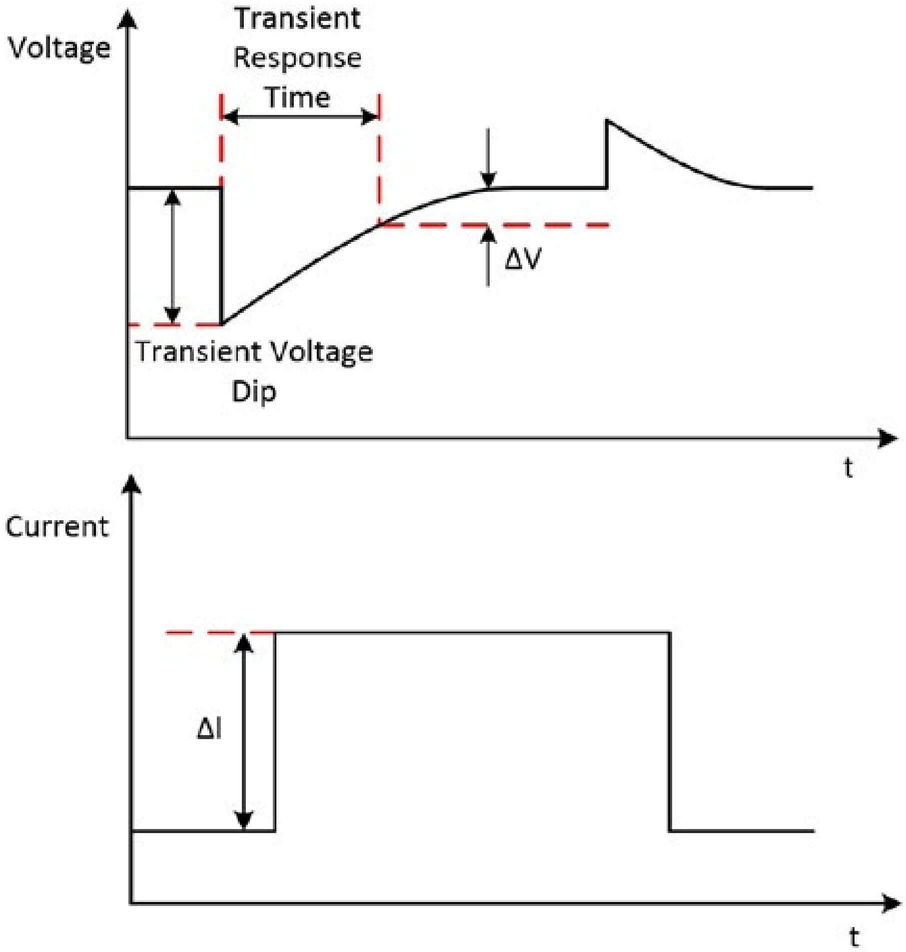


Figure 7. Transient Response to a Current Pulse

Considering an application, if your DUT resistance suddenly drops causing a current pulse to occur, then a transient voltage dip occurs before the power supply's internal control circuit can compensate for the change in load. Similar to rise time and settling time, the transient response specification of a power supply is important because it can impact measurement time.

For more information about transient response and load considerations, see the Power Supply Line and Load Regulation and Cascading Considerations white paper.

d. Isolation

Isolation is a means of physically and electrically separating two parts of a measurement or sourcing device. Electrical isolation pertains to eliminating ground paths between two electrical systems. By providing electrical isolation, you can break ground loops, increase the common-mode range of the power supply, and level shift the signal ground reference to a single system ground. Power supply isolation specifications are particularly important if you are considering cascading the outputs of a power supply to extend the voltage and current ranges, which is covered in greater detail in the [Power Supply Line and Load Regulation and Cascading Considerations] white paper.

The most robust isolation topology is channel-to-channel isolation. In this topology, each channel is individually isolated from one another and from other non-isolated system components. In addition, each channel has its own isolated power supply.

Summary

- Commonly used in research, design, development, and production applications, a **programmable DC power supply** is an instrument that can source power to a connected device.
- When **sourcing**, the power is generated in the supply and dissipates in the DUT. When **sinking**, the power is generated in the DUT and dissipates in the supply.
- DC power supplies operate in Quadrant I or III. SMUs work in all four quadrants.
- Programmable DC power supplies can operate in either constant voltage mode or constant current mode.
- In **constant voltage mode**, a power supply behaves like a voltage source, holding the voltage across the output terminals constant while the current output varies.
- In **constant current mode**, the power supply behaves like a current source, holding the current constant while the output voltage varies.
- If a load **exceeds a compliance resistance** and passes the current or voltage limit, then the power supply begins operating in **compliance**.
- **Remote sense** is a 4-wire connection technique that can eliminate the effects of lead resistance.
- **Ripple** is a type of periodic noise resulting from AC-to-DC rectification required to convert the AC power from the wall outlet to the desired DC levels.
- **Rise time** and **settling time** are key indicators of a power supply's ability to reach the desired voltage level and stabilize.
- **Transient response** describes how a power supply operating in constant voltage mode responds to a sudden change in load current.
- By providing **electrical isolation**, you can break ground loops, increase the common-mode range of the power supply, and level shift the signal ground reference to a single system ground.

Power Supply Line and Load Regulation and Cascading Considerations

Overview

Learn about specific programmable DC power supply applications, including load considerations, line and load regulation, cascading the outputs of a power supply, and switching power supply signals. This tutorial is part of the Instrument Fundamentals series.

Contents

Line and Load Regulation

For a power supply to maintain a stable output, it is important to make sure that it can maintain a given output regardless of changes in the input voltage, connected device, or load. Line regulation refers to a power supply's ability to maintain its output voltage given changes in input voltage. This is especially important in situations where the power supply's input source is unstable or if the power supply is not line regulated, which could cause large output swings.

DC power supplies convert an AC input power source to a desired DC output level. Some DC power supplies require additional power from an auxiliary power supply to achieve the desired output levels. Line regulation is commonly published for power supplies that require an auxiliary power supply, but is not specified otherwise. Therefore, you should not be alarmed if a particular power supply does not publish a line regulation specification.

Line regulation involves a power supply's stability in relation to its input voltage; load regulation is a power supply's ability to maintain a constant output level given changes in its load. For example, if a 10 W power supply is set to output 10 V in constant voltage mode, then it should remain at 10 V whether it is outputting 1 mA or 1 A of current. Load regulation is the measure of how much you can expect the output to change across the full output capability of the supply. Alternatively, in constant current mode, load regulation refers to the amount of change in output current in relation to the change in voltage drop.

Load Considerations

As mentioned before, different load conditions can have an impact on the ability of a programmable DC power supply to perform as expected. You should exercise caution when sourcing power to capacitive, inductive, and reverse-current loads. Improper use can cause ringing in your output signal or damage your power supply. Ringing in a power supply signal is an undesired oscillation of the output voltage as the power supply attempts to recover from a transient caused by a sudden change in current. Ringing affects a system's ability to stabilize, which increases measurement time and, if the oscillation peaks are high enough, can even damage connected circuitry. Below, you can find general guidelines for different load conditions; however, when in doubt, reference your power supply documentation for more information.

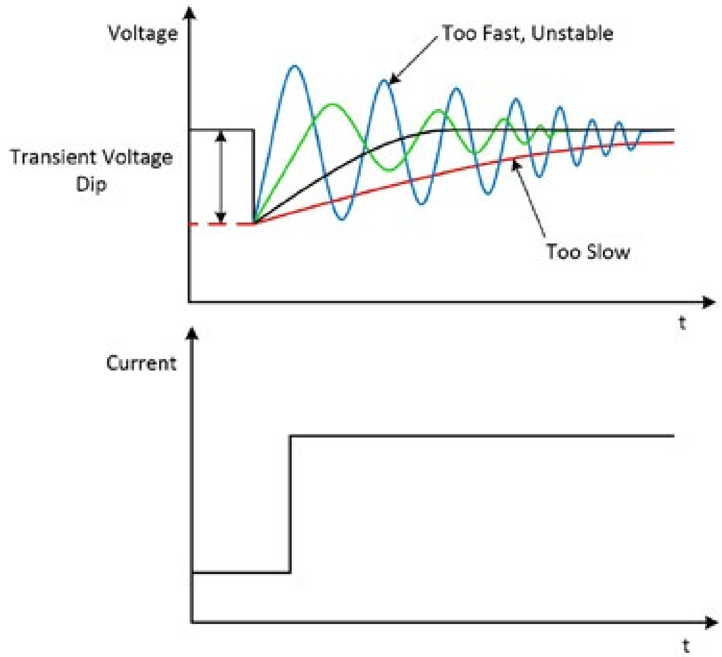


Figure 1. Transient response can affect measurement time and accuracy if unstable or too slow.

a. Capacitive Loads

Generally, a power supply remains stable when driving a capacitive load, but certain loads can cause ringing in the transient response of the device. The slew rate of a power supply is the maximum rate of change of the output voltage as a function of time, which is directly related to transient response. When using a power supply to drive a capacitor, the slew rate is limited to the output current limit divided by the total load capacitance, as shown below.

$$Slew\ Rate = \frac{\Delta V}{\Delta t} = \frac{I}{C}$$

Using the slew rate formula, you can see that the larger the load capacitance, the slower the change in output voltage. If the transient response is too slow, then measurement time could be negatively affected as you must wait for the system to stabilize before taking accurate measurements. However, if the slew rate is too high, then ringing can occur. Furthermore, capacitors are commonly used to dampen ringing in other load conditions.

b. Inductive Loads

A power supply typically remains stable when driving an inductive load in constant voltage mode. If an inductive load is driven by a power supply operating in constant current mode, specifically in higher current ranges, the power supply can become unstable. In these situations, increasing output capacitance may help improve the stability of the system.

Some power supplies have a user-programmable output capacitance option, which gives you the ability to choose a higher capacitance setting to reduce the chance of ringing. Alternatively, you can provide an external capacitance parallel to your load, which dampens ringing. Typical capacitor values used to reduce ringing when driving an inductive load are 0.1-10 µF. However, as described in the previous section, the larger the capacitance, the slower the output response. Therefore, you should use the minimum capacitance required to reduce the effects of ringing. Typically, you want the output voltage to recover from transients as fast as possible to limit the time that your circuit is receiving undesired voltage levels. The quicker your system returns to a stable output level, the quicker you can take your measurements, which results in a shorter overall test time. Refer to your power supply documentation for more information.

c. Reverse Current Loads

Occasionally, an active load may pass a reverse current to the power supply. Power supplies not designed for four-quadrant operation may become damaged if reverse currents are applied to their output terminals. Reverse currents can cause your power supply to move into an unregulated mode. To avoid reverse currents, you can use a bleed-off load to preload the output of the device. Ideally, a bleed-off load should draw the same amount of current from the device that an active load may pass to the power supply.

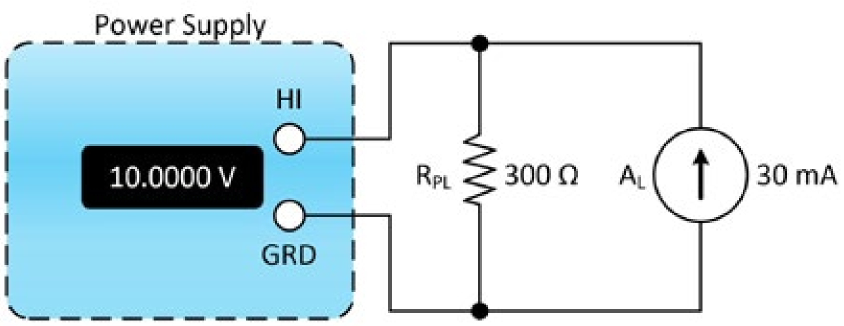


Figure 2. Use a bleed-off load to protect your power supply from the damage that reverse currents can cause.

For example, suppose your power supply is operating in constant voltage mode, supplying 10 V to an active load that can produce a 30 mA reverse current. In this case, a parallel resistor serves as a bleed-off load to preload the power supply output. The value of the bleed-off resistor should be such that the current flowing out of the power supply output is greater than, or equal to, the reverse current produced by your active load. Dividing 10 V by 30 mA suggests using a preload resistance of 333 Ω, effectively matching the reverse current and preventing damage to the power supply.

Cascading Power Supply Channels

Output voltage and current can be increased by cascading the outputs of a multichannel power supply or multiple power supplies. Sometimes an application requires more voltage or current than a single channel of a power supply can output. Cascading power supply channels can extend an output's voltage and current capabilities, but you should practice extreme caution when doing so, as this can easily damage the power supply or user if not done correctly.

a. Cascading Power Supply Channels to Increase Voltage Output

If your power supply offers isolated outputs, or if you have multiple isolated power supplies available, then you can easily extend the maximum voltage output range by cascading channels in series. To cascade multiple isolated channels from a single power supply or multiple power supplies, connect the channels in series as shown in Figure 3. The resulting voltage supplied to your load will be equal to the sum of individual channel voltages.

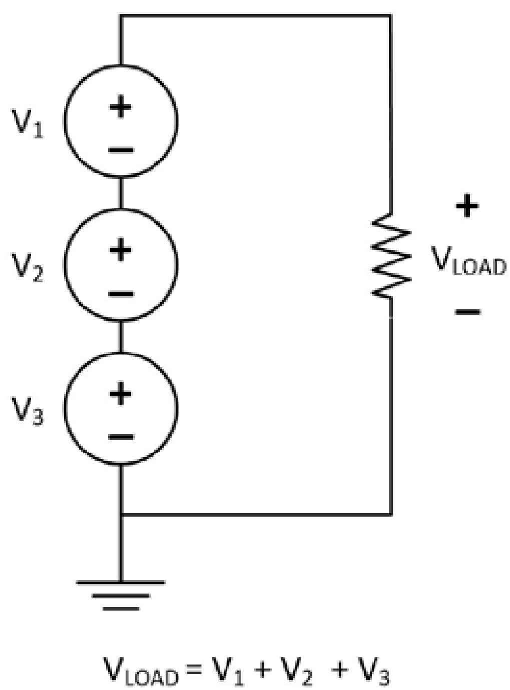


Figure 3. Increase output voltage by cascading isolated power supply channels.

Important: When cascading power supply channels, make sure that the voltage between every pin and device ground is less than the maximum specified isolation voltage. For example, if your device's isolation specification states that each channel is isolated up to 60 VDC from ground, then the voltage between each pin and ground should be less than 60 VDC. Failing to adhere to this specification can damage the device and/or harm the user.

b. Cascading Power Supply Channels to Increase Current Output

If your power supply offers isolated outputs, or if you have multiple isolated power supplies available, then you can easily extend the maximum current output range by connecting channels in parallel. To increase current output by cascading multiple isolated channels from a single power supply or multiple power supplies, connect the channels in parallel as shown in Figure 4. The resulting current supplied by the power supply to your load will be equal to the sum of individual channel currents.

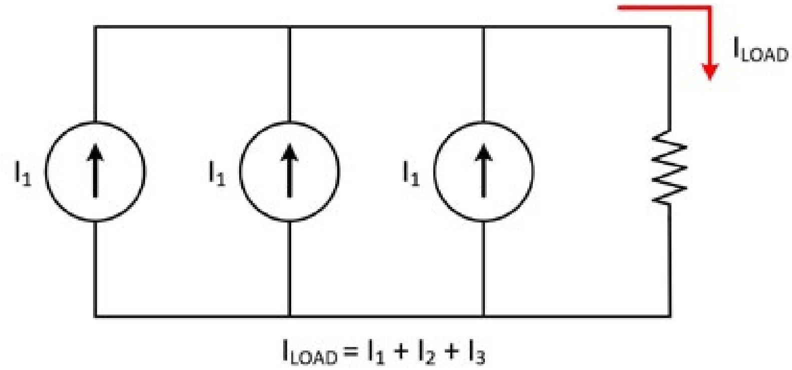


Figure 4. Increase output current by cascading isolated power supply channels.

Summary

- **Line regulation** is the ability of a power supply to maintain the output voltage given changes in the input line voltage.
- **Load regulation** is the ability of a power supply output to remain constant given changes in the load.
- You should e**xercise caution when sourcing power to capacitive, inductive, and reverse-current loads.** Improper use can cause ringing in your output signal or damage your power supply.
- **Ringing** in a power supply signal is an undesired oscillation of the output voltage as the power supply attempts to recover from a transient caused by a sudden change in current, which can increase test time or even damage connected circuitry with high-voltage spikes.
- The **slew rate** of a power supply is the maximum rate of change of the output voltage as a function of time.
- If an inductive load is driven by a power supply operating in constant current mode, increasing output capacitance may help improve the stability of the system.
- Reverse currents can cause your power supply to move into an unregulated mode and may cause damage. Avoid reverse currents by using a bleed-off load to preload the output of the device.
- Output voltage and current can be increased by cascading the outputs of a multichannel power supply or multiple power supplies.

- When cascading power channels, make sure that the voltage between every pin and device ground is less than the maximum specified isolation voltage.
- **Carry current** describes the amount of current that can pass through a previously closed relay without causing damage. On the other hand, **switching current** is the maximum rated current that can flow throw the switch as it makes or breaks contact without causing damage.

Understanding FFTs and Windowing

Overview

Learn about the time and frequency domain, fast Fourier transforms (FFTs), and windowing as well as how you can use them to improve your understanding of a signal. This tutorial is part of the Instrument Fundamentals series.

Contents

Understanding the Time Domain, Frequency Domain, and FFT

The Fourier transform can be powerful in understanding everyday signals and troubleshooting errors in signals. Although the Fourier transform is a complicated mathematical function, it isn't a complicated concept to understand and relate to your measured signals. **Essentially, it takes a signal and breaks it down into sine waves of different amplitudes and frequencies.** Let's take a deeper look at what this means and why it is useful.

a. All Signals Are the Sum of Sines

When looking at real-world signals, you usually view them as a voltage changing over time. This is referred to as the **time domain**. Fourier's theorem states that any waveform in the time domain can be represented by the weighted sum of sines and cosines. For example, take two sine waves, where one is three times as fast as the other–or the frequency is 1/3 the first signal. When you add them, you can see you get a different signal.

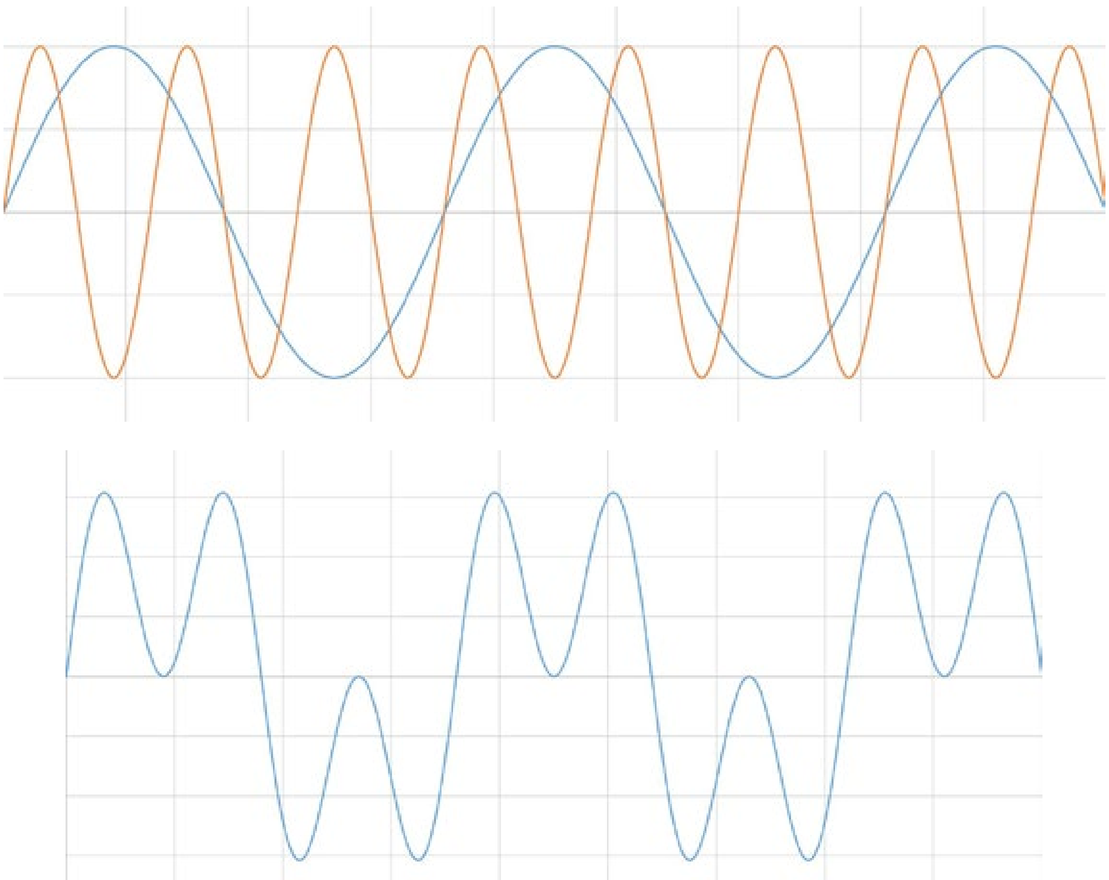


Figure 1. When you add two signals, you get a new signal.

Now imagine if that second wave was also 1/3 the amplitude. This time, just the peaks are affected.

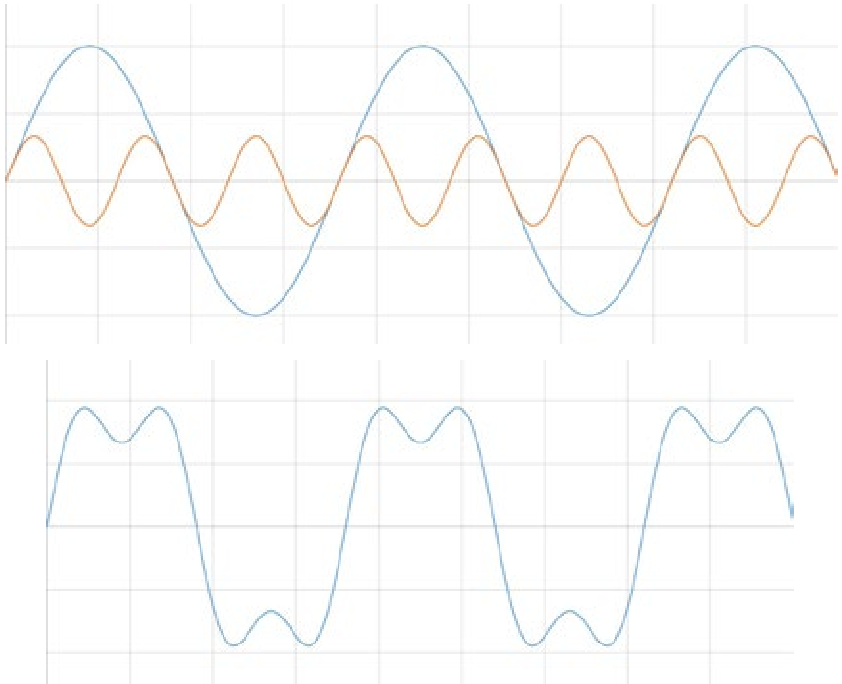


Figure 2. Adjusting the amplitude when adding signals affects the peaks.

Imagine you added a third signal that was 1/5 the amplitude and frequency of the original signal. If you continued in this fashion until you hit the noise floor, you might recognize the resulting waveform.

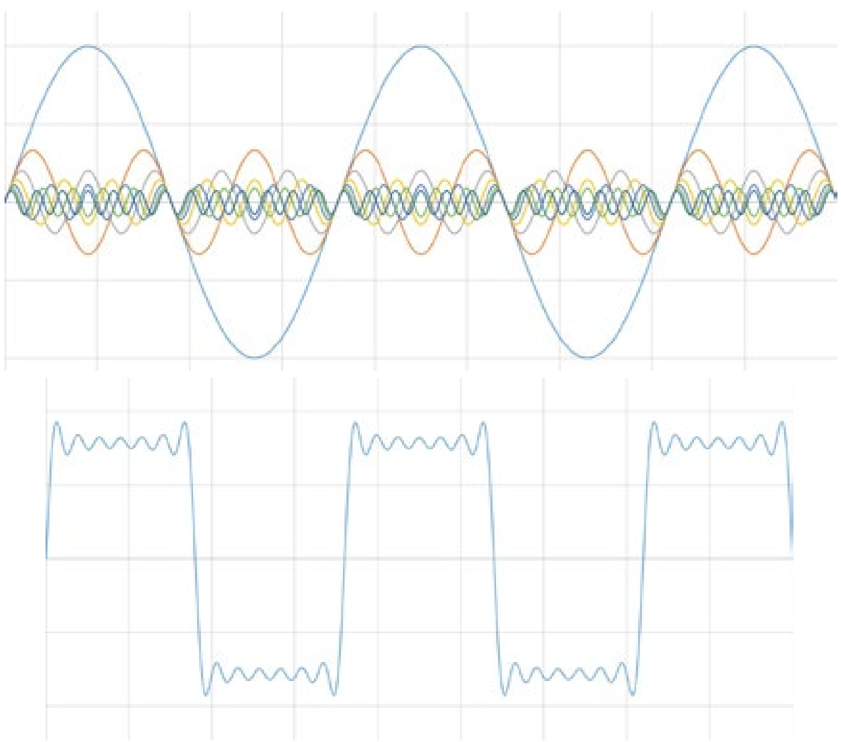


Figure 3. A square wave is the sum of sines.

You have now created a square wave. In this way, all signals in the time domain can be represented by a series of sines.

Although it is pretty neat that you can construct signals in this fashion, why do you actually care? Because if you can construct a signal using sines, you can also deconstruct signals into sines. Once a signal is deconstructed, you can then see and analyze the different frequencies that are present in the original signal. Take a look at a few examples where being able to deconstruct a signal has proven useful:

- If you deconstruct radio waves, you can choose which particular frequency–or station–you want to listen to.
- If you deconstruct audio waves into different frequencies such as bass and treble, you can alter the tones or frequencies to boost certain sounds to remove unwanted noise.
- If you deconstruct earthquake vibrations of varying speeds and strengths, you can optimize building designs to avoid the strongest vibrations.
- If you deconstruct computer data, you can ignore the least important frequencies and lead to more compact representations in memory, otherwise known as file compression.

b. Deconstructing Signals Using the FFT

The Fourier transform deconstructs a time domain representation of a signal into the frequency domain representation. The frequency domain shows the voltages present at varying frequencies. It is a different way to look at the same signal.

A digitizer samples a waveform and transforms it into discrete values. Because of this transformation, the Fourier transform will not work on this data. Instead, the discrete Fourier transform (DFT) is used, which produces as its result the frequency domain components in discrete values, or bins. The fast Fourier (FFT) is an optimized implementation of a DFT that takes less computation to perform but essentially just deconstructs a signal.

Take a look at the signal from Figure 1 above. There are two signals at two different frequencies; in this case, the signal has two spikes in the frequency domain–one at each of the two frequencies of the sines that composed the signal in the first place.

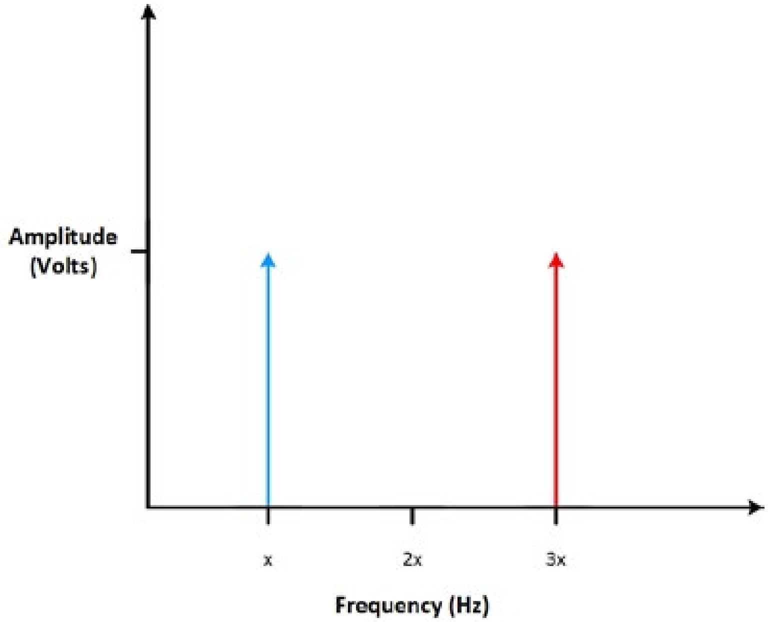


Figure 4. When two sine waves of equal amplitude are added, they result in two spikes in the frequency domain.

The amplitude of the original signal is represented on the vertical axis. If you look at the signal from Figure 2 above where there are two different signals at different amplitudes, you can see that the most prominent spike corresponds to the frequency of the highest voltage sine signal. Looking at a signal in the time domain, you can get a good idea of the original signal by knowing at what frequencies the largest voltage signals occur.

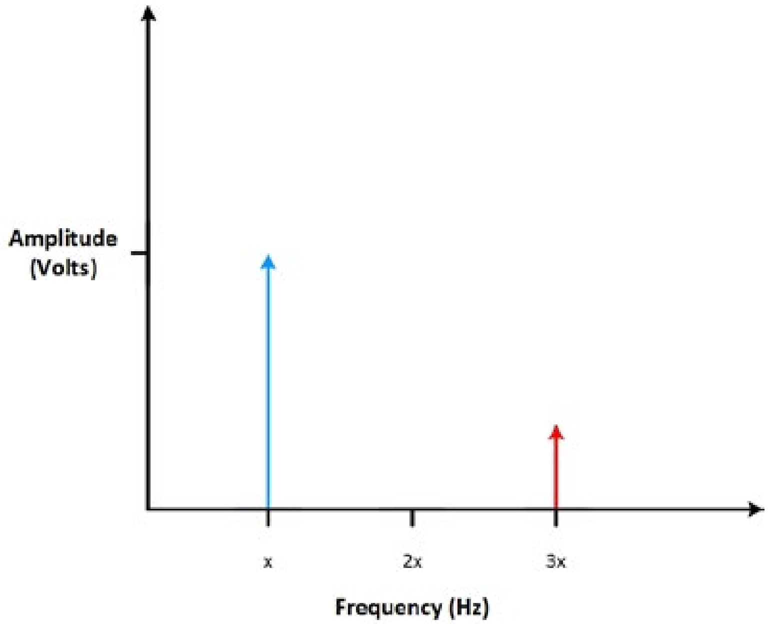


Figure 5. The highest spike is the frequency of the largest amplitude.

It can also be helpful to look at the shape of the signal in the frequency domain. For instance, let's take a look at the square wave in the frequency domain. We created the square wave using many sine waves at varying frequencies; as such, you would expect many spikes in the signal in the frequency domain—one for each signal added. If you see a nice ramp in the frequency domain, you know the original signal was a square wave.

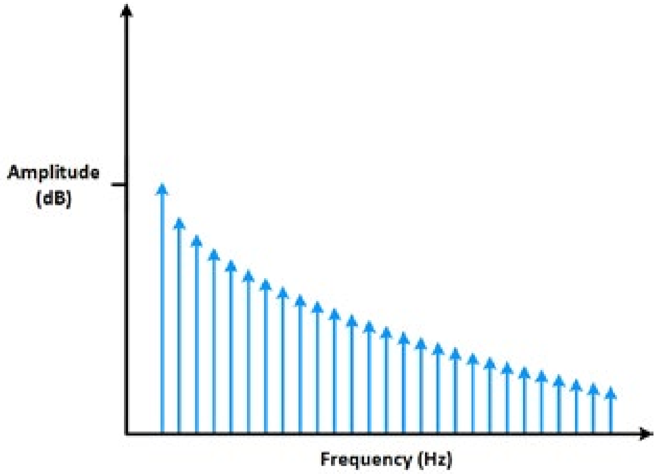
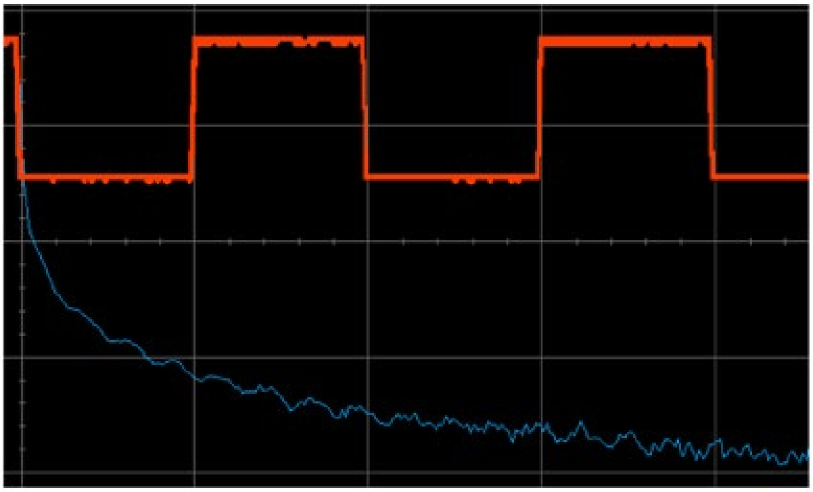


Figure 6. The frequency domain of a sine wave looks like a ramp.

So what does this look like in the real world? Many mixed-signal oscilloscopes (MSO) have an FFT function. Below, you can see what an FFT of a square wave looks like on a mixed-signal graph. If you zoom in, you can actually see the individual spikes in the frequency domain.

A



B

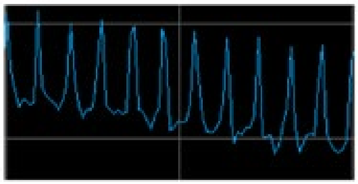


Figure 7. The original sine wave and its corresponding FFT are displayed in A, while B is a zoomed in portion of the FFT where you can see the individual spikes.

Looking at signals in the frequency domain can help when validating and troubleshooting signals. For instance, say you have a circuit that is supposed to output a sine wave. You can view the output signal on the oscilloscope in the time domain in Figure 8 below. It looks pretty good!

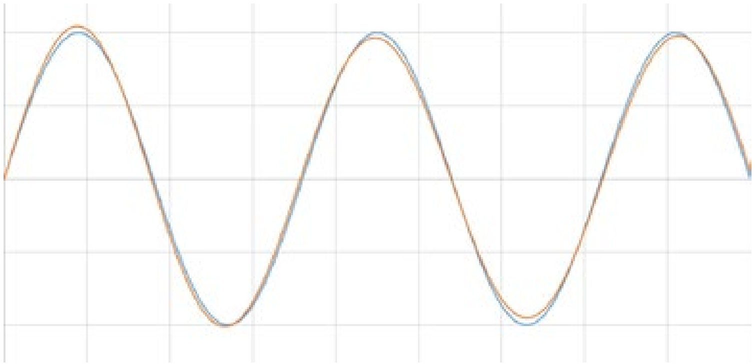


Figure 8. If these two waves were added, they would look like a perfect sine wave because they are so similar.

However, when you view the signal in the frequency domain, you expect only one spike because you are expecting to output a single sine wave at only one frequency. However, you can see that there is a smaller spike at a higher frequency; this is telling you that the sine wave isn't as good as you thought. You can work with the circuit to eliminate the cause of the noise added at that particular frequency. The frequency domain is great at showing if a clean signal in the time domain actually contains cross talk, noise, or jitter.

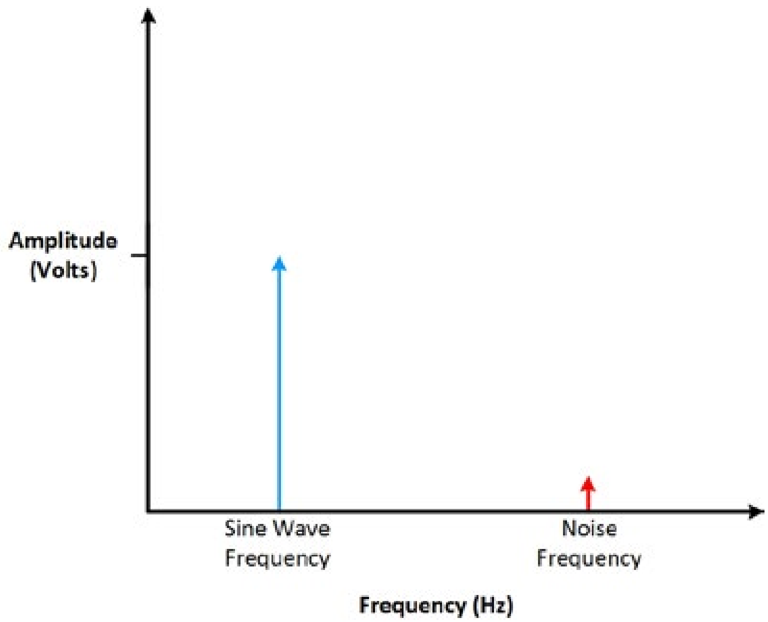


Figure 9. Looking at the seemingly perfect sine wave from Figure 8, you can see here that there is actually a glitch.

Windowing

Although performing an FFT on a signal can provide great insight, it is important to know the limitations of the FFT and how to improve the signal clarity using windowing.

a. What Is Windowing

When you use the FFT to measure the frequency component of a signal, you are basing the analysis on a finite set of data. The actual FFT transform assumes that it is a finite data set, a continuous spectrum that is one period of a periodic signal. For the FFT, both the time domain and the frequency domain are circular topologies, so the two endpoints of the time waveform are interpreted as though they were connected together. When the measured signal is periodic and an integer number of periods fill the acquisition time interval, the FFT turns out fine as it matches this assumption.

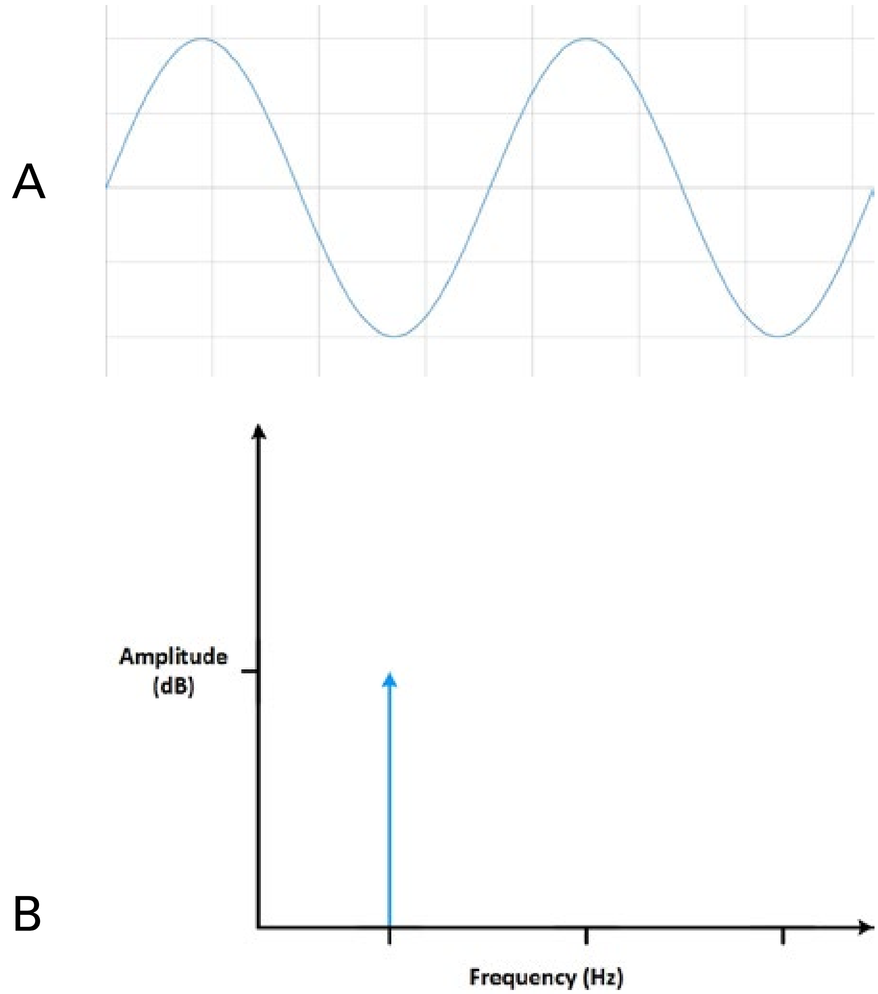
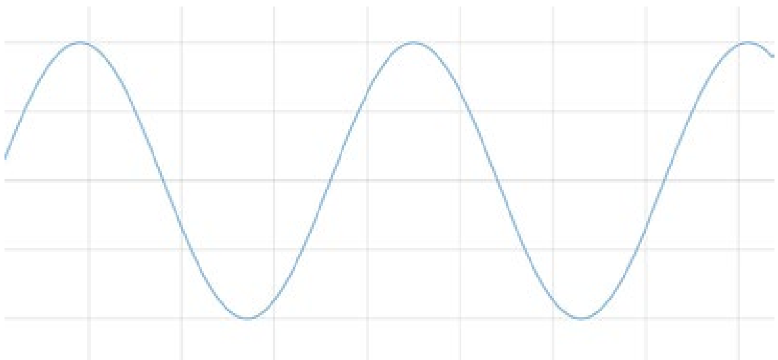


Figure 10. Measuring an integer number of periods (A) gives an ideal FFT (B).

However, many times, the measured signal isn't an integer number of periods. Therefore, the finiteness of the measured signal may result in a truncated waveform with different characteristics from the original continuous-time signal, and the finiteness can introduce sharp transition changes into the measured signal. The sharp transitions are discontinuities.

When the number of periods in the acquisition is not an integer, the endpoints are discontinuous. These artificial discontinuities show up in the FFT as high-frequency components not present in the original signal. These frequencies can be much higher than the Nyquist frequency and are aliased between 0 and half of your sampling rate. The spectrum you get by using a FFT, therefore, is not the actual spectrum of the original signal, but a smeared version. It appears as if energy at one frequency leaks into other frequencies. This phenomenon is known as spectral leakage, which causes the fine spectral lines to spread into wider signals.

A



B

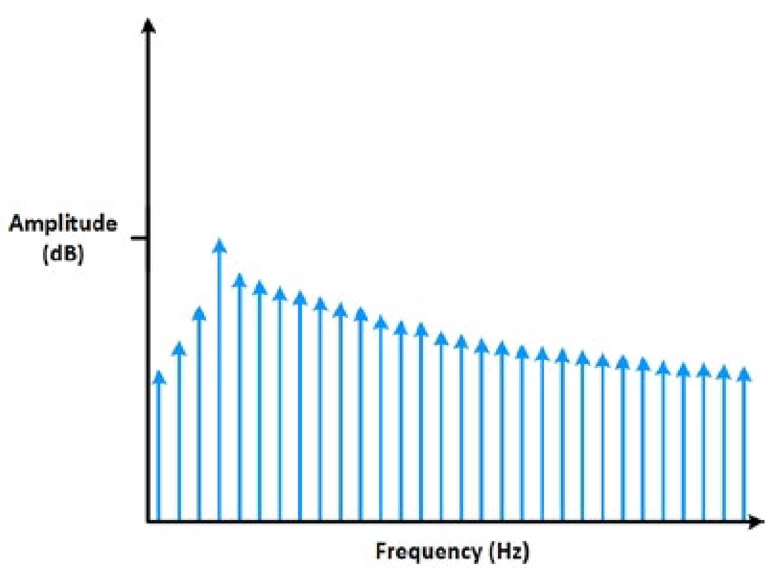


Figure 11. Measuring a noninteger number of periods (A) adds spectral leakage to the FFT (B).

You can minimize the effects of performing an FFT over a noninteger number of cycles by using a technique called windowing. Windowing reduces the amplitude of the discontinuities at the boundaries of each finite sequence acquired by the digitizer. Windowing consists of multiplying the time record by a finite-length window with an amplitude that varies smoothly and gradually toward zero at the edges. This makes the endpoints of the waveform meet and, therefore, results in a continuous waveform without sharp transitions. This technique is also referred to as *applying a window*.

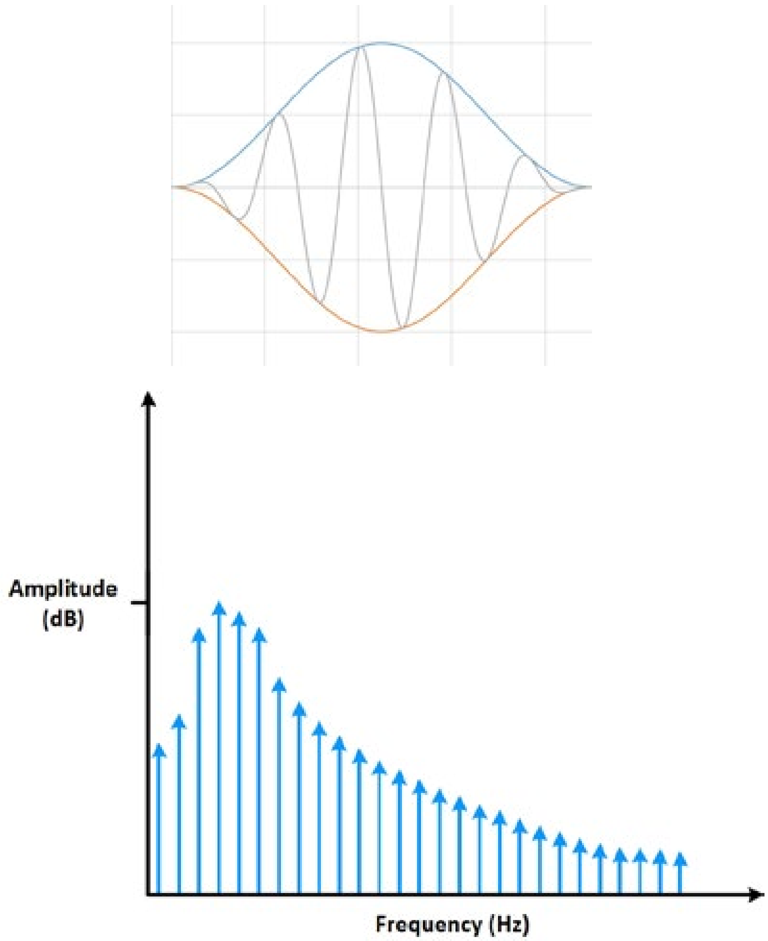


Figure 12. Applying a window minimizes the effect of spectral leakage.

b. Windowing Functions

There are several different types of window functions that you can apply depending on the signal. To understand how a given window affects the frequency spectrum, you need to understand more about the frequency characteristics of windows.

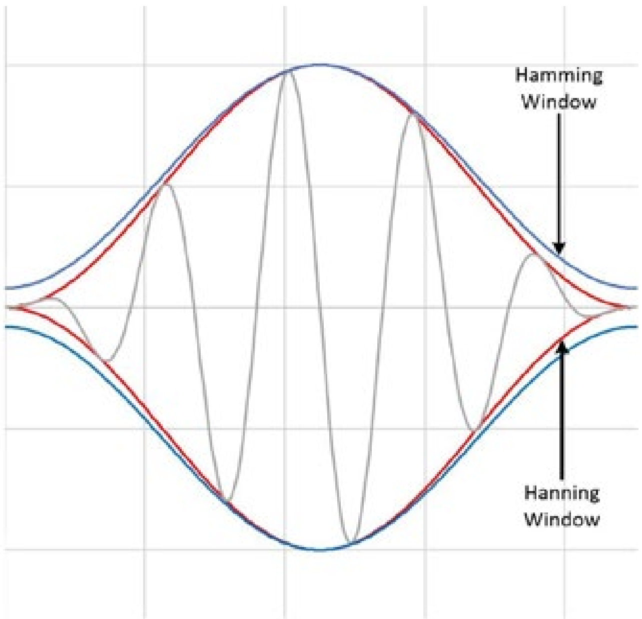
An actual plot of a window shows that the frequency characteristic of a window is a continuous spectrum with a main lobe and several side lobes. The main lobe is centered at each frequency component of the time-domain signal, and the side lobes approach zero. The height of the side lobes indicates the affect the windowing function has on frequencies around main lobes. The side lobe response of a strong sinusoidal signal can overpower the main lobe response of a nearby weak sinusoidal signal. Typically, lower side lobes reduce leakage in the measured FFT but increase the bandwidth of the major lobe. The side lobe roll-off rate is the asymptotic decay rate of the side lobe peaks. By increasing the side lobe roll-off rate, you can reduce spectral leakage.

Selecting a window function is not a simple task. Each window function has its own characteristics and suitability for different applications. To choose a window function, you must estimate the frequency content of the signal.

- If the signal contains strong interfering frequency components distant from the frequency of interest, choose a smoothing window with a high side lobe roll-off rate.
- If the signal contains strong interfering signals near the frequency of interest, choose a window function with a low maximum side lobe level.
- If the frequency of interest contains two or more signals very near to each other, spectral resolution is important. In this case, it is best to choose a smoothing window with a very narrow main lobe.
- If the amplitude accuracy of a single frequency component is more important than the exact location of the component in a given frequency bin, choose a window with a wide main lobe.
- If the signal spectrum is rather flat or broadband in frequency content, use the uniform window, or no window.
- In general, the Hanning (Hann) window is satisfactory in 95 percent of cases. It has good frequency resolution and reduced spectral leakage. If you do not know the nature of the signal but you want to apply a smoothing window, start with the Hann window.

Even if you use no window, the signal is convolved with a rectangular-shaped window of uniform height, by the nature of taking a snapshot in time of the input signal and working with a discrete signal. This convolution has a sine function characteristic spectrum. For this reason, no window is often called the uniform or rectangular window because there is still a windowing effect.

The Hamming and Hann window functions both have a sinusoidal shape. Both windows result in a wide peak but low side lobes. However, the Hann window touches zero at both ends eliminating all discontinuity. The Hamming window doesn't quite reach zero and thus still has a slight discontinuity in the signal. Because of this difference, the Hamming window does a better job of cancelling the nearest side lobe but a poorer job of canceling any others. These window functions are useful for noise measurements where better frequency resolution than some of the other windows is wanted but moderate side lobes do not present a problem.



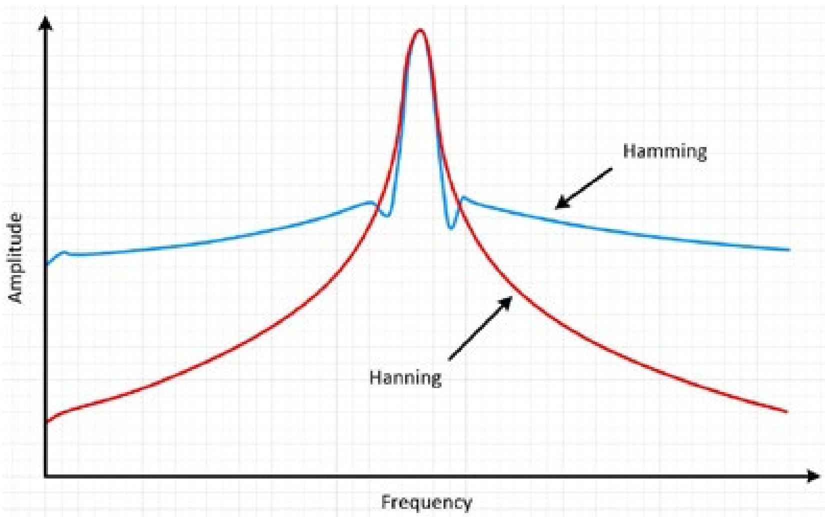
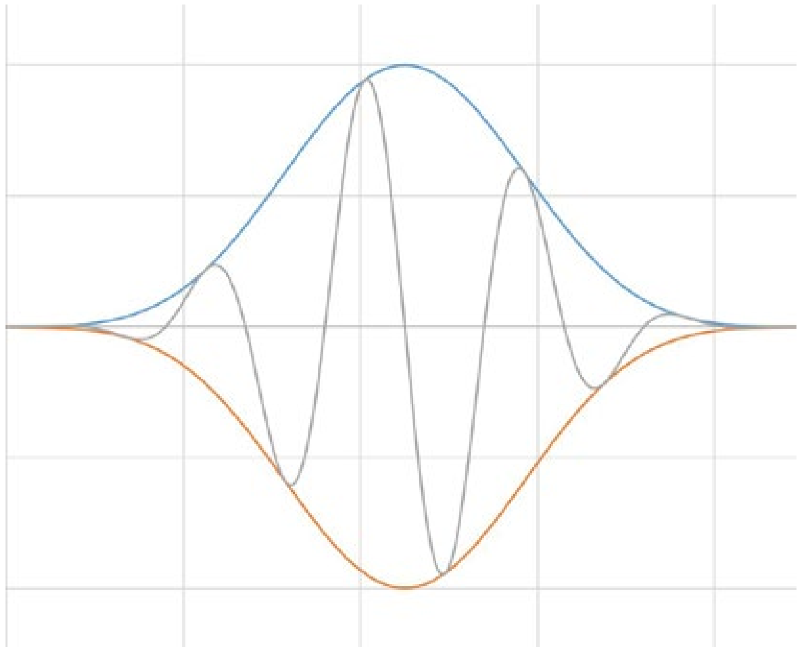


Figure 13. Hamming and Hann windowing result in a wide peak but nice low side lobes.

The Blackman-Harris window is similar to Hamming and Hann windows. The resulting spectrum has a wide peak, but good side lobe compression. There are two main types of this window. The 4-term Blackman-Harris is a good general-purpose window, having side lobe rejection in the high 90s dB and a moderately wide main lobe. The 7-term Blackman-Harris window function has all the dynamic range you should ever need, but it comes with a wide main lobe.



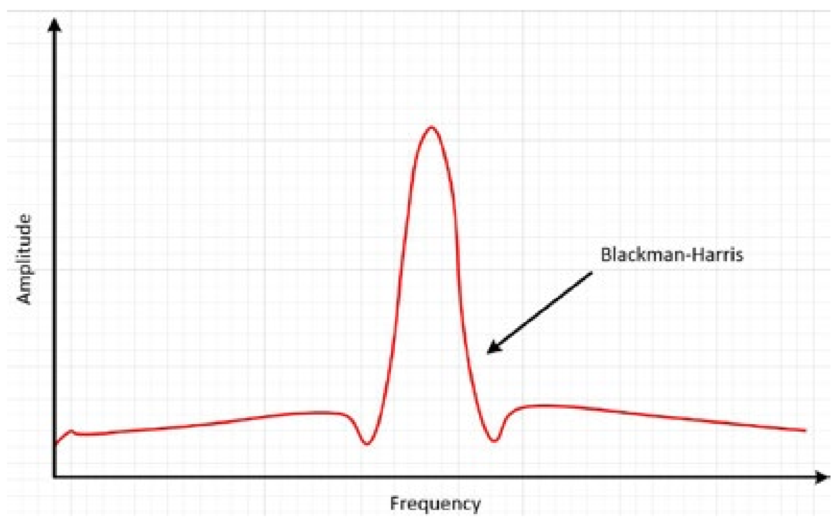
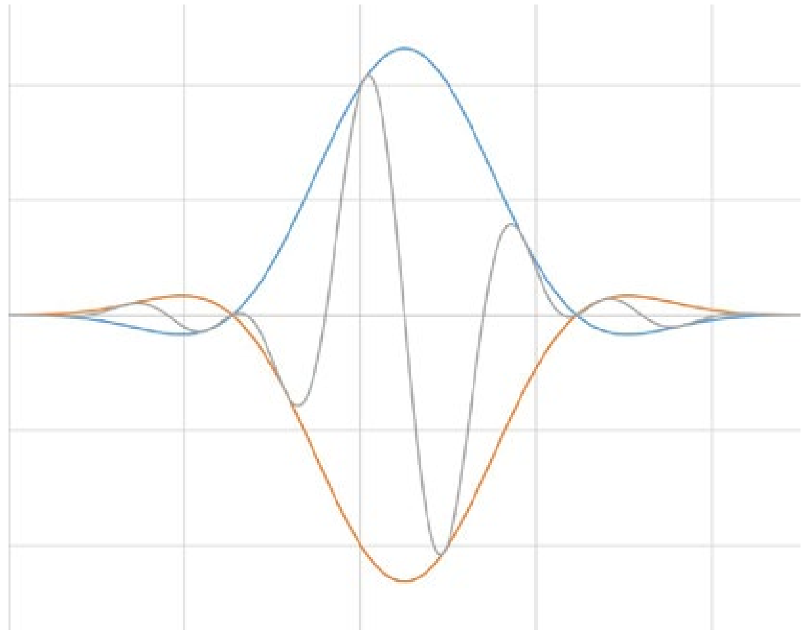


Figure 14. The Blackman-Harris results in a wide peak, but good side lobe compression.

A Kaiser-Bessel window strikes a balance among the various conflicting goals of amplitude accuracy, side lobe distance, and side lobe height. It compares roughly to the Blackman-Harris window functions, but for the same main lobe width, the near side lobes tend to be higher, but the further out side lobes are lower. Choosing this window often reveals signals close to the noise floor.

The flat top window is sinusoidal as well, but it actually crosses the zero line. This causes a much broader peak in the frequency domain, which is closer to the true amplitude of the signal than with other windows.



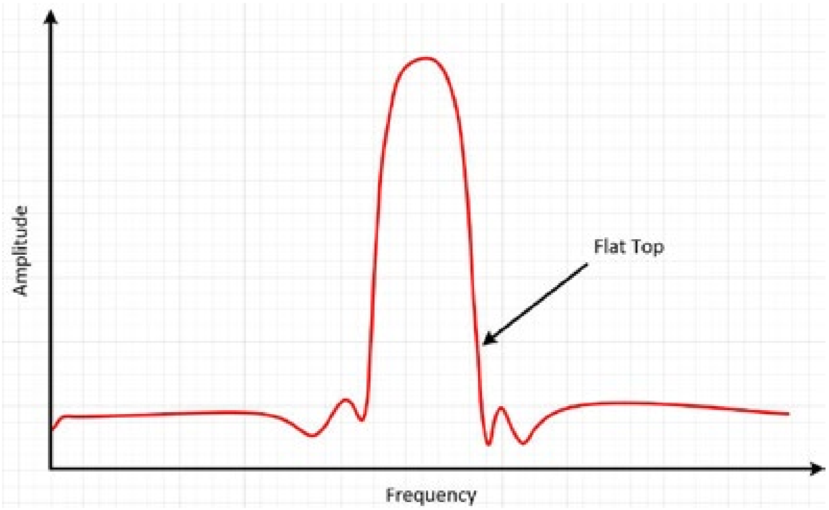


Figure 15. The flat top window results in more accurate amplitude information.

These are just a few of the possible window functions. There is no universal approach for selecting a window function. However, the table below can help you in your initial choice. Always compare the performance of different window functions to find the best one for the application.

Signal Content	Window
Sine wave or combination of sine waves	Hann
Sine wave (amplitude accuracy is important)	Flat Top
Narrowband random signal (vibration data)	Hann
Broadband random (white noise)	Uniform
Closely spaced sine waves	Uniform, Hamming
Excitation signals (hammer blow)	Force
Response signals	Exponential
Unknown content	Hann
Sine wave or combination of sine waves	Hann
Sine wave (amplitude accuracy is important)	Flat Top
Narrowband random signal (vibration data)	Hann
Broadband random (white noise)	Uniform
Two tones with frequencies close but amplitudes very different	Kaiser-Bessel
Two tones with frequencies close and almost equal amplitudes	Uniform
Accurate single tone amplitude measurements	Flat Top

Summary

- All signals in the time domain can be represented by a series of sines.
- An **FFT transform** deconstructs a time domain representation of a signal into the frequency domain representation to analyze the different frequencies in a signal.
- The **frequency domain** is great at showing you if a clean signal in the time domain actually contains cross talk, noise, or jitter.
- **Spectral leakage** is caused by discontinuities in the original, noninteger number of periods in a signal and can be improved using windowing.
- **Windowing** reduces the amplitude of the discontinuities at the boundaries of each finite sequence acquired by the digitizer.
- No window is often called the uniform or rectangular window because there is still a windowing effect.
- In general, the **Hanning window** is satisfactory in 95 percent of cases. It has good frequency resolution and reduced spectral leakage.
- You should compare the performance of different window functions to find the best one for the application.