RECORDERS

Recorders are used in instrumentation and measurement to record the variations in time of some physical quantity for later analysis. This is in distinction to many audio and video recordings for which the purpose is to play back the recorded material. Recording is essential to the instrumentation, measurement, and analysis processes, because many phenomena that must be measured occur much too rapidly to be analyzed in real time.

Modern high speed recorders share much in common with digital storage oscilloscopes. While these are often used as recorders, the emphasis in design is different. The primary design focus of the digital storage oscilloscope is convenience of capturing and displaying data to be instantly analyzed with the resolution of the eye. The primary focus in recorder design is to capture data for later analysis at potentially much higher resolution.

Important elements of any recording system are the transducers, to convert the physical quantities to be recorded into electrical signals, and the signal conditioners and instrumentation amplifiers to convert the signal to a form acceptable for recording. These elements are covered in other articles. Also covered in other articles are video recording, the medium that records the time history of two dimensional images, and data recording, the medium that records digital data. This article will be confined to a discussion of recording multiple channels of analog data that have already been converted to a voltage signal.

ANALOG-TO-DIGITAL CONVERSION IN MODERN RECORDERS

Modern recorders rely very heavily on analog to digital conversion. The signals to be recorded are converted to digital form by an analog-to-digital converter (ADC); the digital words are then recorded on media compatible with a personal computer (PC) such as magnetic disk or tape or magneto-optical or optical disk. This is generally the most economical approach, since digital data recording has become so inexpensive because of the large PC market. This approach is also very convenient, because a PC is usually the tool that will be used to analyze the recorded data. Often the recording will be done on a PC while the signal conditioning and analog to digital conversion will be on an add-in board to the PC.

IMPORTANT RECORDER CHARACTERISTICS

Sampling Rate and Bandwidth

The individual numbers stored by the recorder, which represent the amplitude of the recorded signal at a specific time, are called *samples*. The rate at which these samples are collected is called the *sampling rate* and is expressed in samples per second (Sa/s). Sampling rates of commercially available recorders vary from 1 Sa/s to 1 TSa/s (10^{12} Sa/s). Choosing an appropriate sampling rate for any particular application is important, because using too low a rate will result in loss of data while using too high a rate can add appreciably to the cost and complexity of the recorder.

The dynamic response of a recorder is usually modeled as a low pass filter followed by a perfect recorder. The frequency at which the response of this filter is 3 dB below its value at zero frequency is called the *bandwidth* of the recorder. Generally, bandwidth limitation is caused by nonidealness of all parts of the recorder between its input and (including) the ADC. Bandwidths of recorders commercially available records range from 1 Hz to 10 GHz (10^{10} Hz).

There is a relationship between the bandwidth required to record a signal and the sampling rate required. This comes from the sampling theorem (1,2) which states (roughly) that a signal can be reconstructed from samples taken at a rate of twice the signal's bandwidth but cannot be reconstructed from samples taken at a lower rate. In practice, it is often desirable to use a higher sampling rate. If the sampling rate is ten times the bandwidth, the signal can be reconstructed quite accurately by straight line segments between the sampled values. As the sampling rate is reduced to the theoretical limit of twice the bandwidth, much more complicated approaches to interpolation must be used to reconstruct the original signal. These are covered in Refs. (2,3).

Figures 1 and 2 show the different effects of inadequate sampling rate and inadequate bandwidth. Figure 1 is the signal

$$v(t) = \frac{1}{\sqrt{2\pi\tau_1^2}}\exp[-t^2/2\tau_1^2] + .005\frac{1}{\sqrt{2\pi\tau_2^2}}\exp[-(t+10)^2/2\tau_2^2] \quad (1)$$

where $\tau_1 = 10$ ms and $\tau_2 = 0.1$ ms. The cross marks are every 1 ms and show where the signal might be sampled with a sampling rate of 1 kSa/s. The spike at $t = -10$ ms is almost completely missed, but if the samples were offset by 0.5 ms there would be one point at the peak of the spike. Figure 2 shows the same signal after having passed through a 500 Hz bandwidth filter. The spike has become much smaller and much wider.

Resolution and Dynamic Range

The resolution of a recorder is a measure of the precision of the digital words that are used to represent the sample values, though this terminology is not universal. Typically, the ADCs in a recorder convert the amplitudes to binary words, and the resolution is expressed as the number of bits in one of these words. If the resolution is N bits, the number of different values representable is 2^N. The resolutions of commer-

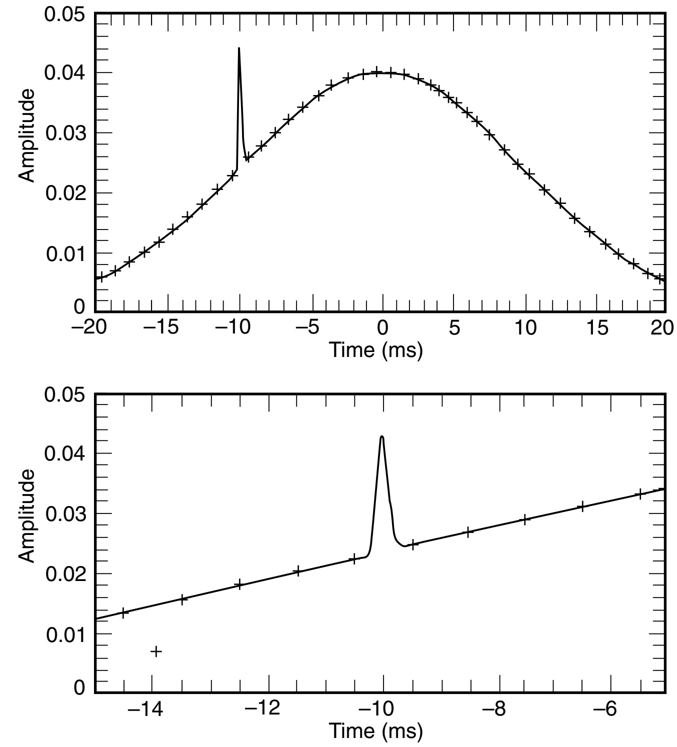

Figure 1. The signal of Eq. (1) sampled at 1 kSa/s (sample points indicated by + signs.) The short pulse at $t = -10$ ms is completely missed by the sampling. However, the Fourier transform of the signal (shown in Fig. 18) and the usual "rules-of-thumb" indicate that 1 kSa/s is an adequate sampling rate.

cially available recorders range from 6 bits to 24 bits. For some lower speed recorders the ADC converts directly to a decimal number. In this case the resolution will be expressed as the number of digits in the decimal number. If the resolution is N digits, then the number of possible values is 10^N. Frequently ADCs produce 4×10^N different values; these are said to have N-and-one-half digits of resolution.

The dynamic range of a recorder is the ratio of the largest to the smallest signal magnitude that can be measured in a single record. This term has significant uncertainty in its meaning due to the lack of precision of the phrase "smallest signal magnitude that can be measured"; this will be covered in more detail later in the article. In most cases an upper bound on the dynamic range of a recorder with a resolution of N bits is 2^N. The dynamic range is frequently lower than this because of noise signals generated in the recorder that interfere with the measurement of small signals. Dynamic range is usually expressed in decibels rather than as a pure ratio.

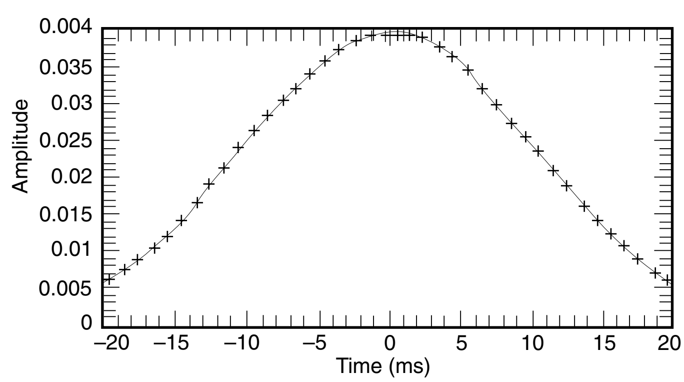

Figure 2. The signal of Fig. 1 filtered with a 500 Hz bandwidth low-pass filter. Only a minuscule hint of the pulse at $t = -10$ ms remains, but the 1 kSa/s sampling adequately represents the resulting signal.

Amplitude Accuracy

The accuracy of a recorder is a very complex subject that will be dealt with in more detail later in the article. It can depend very strongly on the signal being measured, and it can't be adequately quantified with a single number or even several numbers. However, for many recorders the accuracy is significantly poorer than the resolution and there are a few specifications that are often given which help quantify it. A fairly overall measure of the accuracy is the *effective number of bits,* N_e. This is specified for a specific frequency (or several of them) and means that the errors in recording a sinusoidal signal of that frequency are of the same size as would be made by an ideal ADC with N_e bits of resolution. Sometimes an accuracy specification is given as a percentage of full scale. This usually refers to the measurement of a direct current (dc) or very slowly varying signals.

Time Base Accuracy

For a digital recorder there are two primary measures of time base accuracy—the sampling rate accuracy and the aperture error. The sampling rate accuracy is the difference between the actual sampling rate and the nominal sampling rate. For time bases controlled by a crystal oscillator, this is usually in the range of 1 Hz/MHz to 100 Hz/MHz. The aperture error is either the maximum or the root-mean-square (rms) deviation between the time at which a sample is taken and the time the sample would have been taken if the sample spacing were perfectly uniform. The effect of aperture error on the accuracy of a sampled signal value is proportional to the rate of change of the signal. The error in amplitude units (e.g., volts) is the product of the aperture error in seconds and the slew rate of the signal in amplitude units per second.

Superior time base accuracy is one of the by-products of digital techniques of recording. When analog tape is used for recording the slippage of the tape in the rollers that propel it and variations in motor speed due to mechanical vibrations play a significant role in the accuracy of the time base. Similar problems occur in almost all analog recording mechanisms.

Output and Controls

A recorder should supply the recorded signals to the user in a form that is convenient for the analysis that is to be carried out. In most cases the analysis will be carried out (at least partially) on a PC, and means of transferring data from the recorder to the PC are essential. If the recorder is an attachment to the PC and records directly to the PCs hard disk, there is no problem. In other cases some additional means must be supplied. Some recorders have an integral floppy disk drive which allows the data transfer with no additional hardware or software. Often it is desirable to transfer data from the recorder to the PC without human intervention. If the recorder supports a common protocol, such as RS232 (a standard for data communications interfaces from the Electronics Industry Association) or the General Purpose Interface Bus (GPIB), this is facilitated. Sometimes it is necessary to transfer data faster than the aforementioned protocols allow, so another means, such as through a parallel port on the computer, must be available.

It is often necessary for the computer to be able to control the recorder. In this case the ability of the recorder to accept control commands through RS232 or GPIB is usually most practical. In complex instrumentation systems, recorders often interface to a standard bus, such as the Computer Automated Measurement and Control (CAMAC) bus or VXI (a high speed bus standard defined by the VXIbus Consortium), and transfer data and control commands through an intermediate controller on the bus.

Many recorders are able to output their data visually along with the outputs to a computer. This is typically through a cathode ray tube (CRT) or a liquid crystal display and/or on paper using a self-contained printing mechanism.

Other Characteristics

A few other characteristics that might be important in some recorder applications are listed here. The existence of differential inputs could be required and, if so, the common mode rejection ratio is of importance. If the recorder has multiple input channels, the maximum cross talk between channels is of interest. If the input signals are not adequately limited, the over voltage recovery time is important. If the recorder does not record at its maximum rate continuously, then there are a few important characteristics: (1) the range of record lengths that can be recorded, (2) the *cycle time,* which is the required time delay between the end of one record and the beginning of the next, and (3) the *throughput,* which is the rate at which data can be continuously transferred out of the recorder while recording is taking place.

RECORDER CONFIGURATIONS

In this section various common recorder configurations will be covered with a description of the limitations and problems usually found in each configuration. This is not all inclusive, and there are variations on each configuration. In particular, the fact that certain problems and limitations are stated to exist for a particular configuration doesn't imply that designs haven't been successfully carried out to circumvent the particular limitations. For the most part, the more complex configurations are used to overcome sampling rate, bandwidth, or amplitude resolution limitations of the simpler configurations.

ADC—Permanent Recorder Configuration

This is the simplest configuration, in which the input signal passes through a buffer amplifier to an ADC, and the output of the ADC is recorded (usually with intervening buffer memory) on a permanent recording device. The permanent recording device could be hard magnetic disk, digital magnetic tape, magneto-optical disk, floppy disk, optical disk (CD-ROM or audio CD) or, for the slowest data rates, printed paper. These media are listed in order of decreasing speed. In this configuration there may be numerous channels, each with its own ADC, feeding the same recording medium.

Data can be written to PC compatible hard disks at rates of up to 100 Mbit/s. However, if one wishes to record at near the maximum rate, extra care must be taken. One must take special care with the software that is storing data on the disk; going through the Windows or Mac operating system is not likely to work. Also hard disks designated for audio/video use, which don't have lengthy temperature calibration cycles, must be used.

To get a perspective, the 100 Mbit/s rate can record 100 channels of audio with a 50 kSa/s sampling rate and a resolution of 20 bits. Digital audio workstations typically use this configuration with up to 50 channels.

Multiplexer Before ADC

This is the same configuration as the previous section with an analog multiplexer preceding the ADC. This is used for the economy gained by sharing one ADC among many (typically 16 to 64) channels. This configuration leads to potential cross talk problems from two sources. First, the settling time requirement for the ADC or the sample-and-hold preceding it is much more stringent than for single channel operation. Even when input signals are constant, with a multiplexer the input to the ADC will abruptly switch from the level of one channel to the level of another just before the second channel is to be converted. This can cause the reading of the second channel to be influenced by the amplitude of the first. The second cross talk source is the bringing together on one chip of several analog channels.

ADC and Solid State Memory

When sampling rates are higher than can be recorded on a permanent medium, recorders typically store their data in solid state memory at a high sampling rate and transfer it to permanent memory later at a slower rate. This approach is necessary when sampling rates exceed 10 MSa/s. The sampling rates attainable with this approach are limited only by the ADC. Record lengths are limited by the amount of memory installed in the recorder. As memory densities increase and memory costs decrease, this becomes less of a limitation.

Frequently, for economic reasons, memories that have cycle times longer than that required by the sampling rate are used. The data from the ADC is sent thorough a demultiplexer and interleaved into several memories. For example, to interleave the data into three memories, the demultiplexer would send the first data value to memory number one, the second value to memory number two, the third to memory number three, and the fourth to memory number one, etc. Each memory in this example receives data at one-third the sampling rate.

This configuration is used at sampling rates up to 500 MSa/s. Currently at rates above 100 MSa/s technology limits the resolution attainable in this configuration to eight bits.

Interleaved ADCs

Several ADCs can be combined in one analog channel to increase the sampling rate. With n ADCs each sampling the same analog signal at a sampling rate of r_0 Sa/s, one adjusts the timing so that ADC $k + 1$ samples the signal at time $1/nr_0$ s after ADC k. When the data from the n ADCs are interleaved, one obtains the signal sampled at a rate of nr_0 Sa/s.

The technological problems involved in accomplishing this are larger than they appear on the surface. The sample and hold circuit preceding each ADC must have speed and bandwidth commensurate with the higher sampling rate of the interleaved system. The most difficult problem is caused by the requirement for very precisely matching the frequency response of the path from the recorder input to each ADC. This problem can best be illustrated with a hypothetical example. From results in later parts of this article one can determine that to maintain eight effective bits with an input signal of 500 MHz requires an rms uncertainty of less than 1 ps in the time any ADC samples the signal. A 1 ps delay error will result, at 500 MHz, from a 5×10^{-4} rad (0.18°) phase shift. This means that to meet the stated requirement (eight effective bits) the phase shift of each signal path must match to 5×10^{-4} rad.

Analog Memories

At sampling rates of 100 MSa/s and higher, current ADC technology limits resolution to eight bits. Figure 3 illustrates a configuration that has been used to obtain higher resolution at high sampling rates. The signal to be recorded is sampled repetitively by a sample-and-hold circuit, and these samples are recorded, temporarily, in an analog memory. After the entire record is recorded, the data is read from the analog memory at a much slower rate than at which it was recorded and stored in digital memory. Analog memories have substantial imperfections that must be corrected for before the data is delivered to the user. The advantage of these recorders can be increased signal resolution for small signals and the resulting increase in dynamic range.

The analog memories used are of two types—sequential access and random access. In both types the signal is recorded as the amount of charge on a capacitor. Random access analog memories are constructed much like Complementary Metal Oxide Semiconductor (CMOS) dynamic RAMs, except that the write circuitry stores a variable amount of charge, and the readout circuitry generates a voltage proportional to the amount of charge stored. Sequential access memories are charge-coupled-devices (CCDs). These contain several hundred charge storage elements coupled in a linear array. When a strobe signal is given to the device, the charge stored in each element is transferred to the next element in the array. The signal to be recorded is transferred to the first element, and the readout is taken from the last element. Analog memories that are adequate for recording applications are not off-the-shelf items; they are generally proprietary devices made by the recorder manufacturers.

This approach has difficult technical problems due to the lack of idealness of the analog memories. The relationship between the charge read in to the memory and the charge read out of the memory is nonlinear and typically different for each memory element. The charge read out may also depend on the charge stored in neighboring memory elements. Recorders in this configuration typically require extensive internal calibration to characterize the memories and extensive calculation to correct the data. These corrections are usually far from perfect. The errors remaining after the corrections have been made are roughly proportional to the instantaneous signal amplitude, and, for signals near the full scale of the recorder are often larger than would have been obtained with an ADC operating at the full sampling rate. The advantage is potentially lower noise at small signal levels.

Sampling Oscilloscope

If the signal to be recorded can be produced repetitively, it can be recorded with a sampling oscilloscope. Sampling oscilloscopes have bandwidths of up to 50 GHz and unlimited sampling rates. They take one sample on the signal for each occurrence of the signal. Figure 4 shows the functioning of a sampling oscilloscope simplified form. A trigger signal that is synchronized with the signal must be available. To capture the leading edge of the signal, it must occur before the signal. The trigger signal starts a variable delay circuit. At the end of the programmed delay time the sample-and-hold is strobed. The voltage stored by the sample-and-hold circuit is digitized and the value is stored as the amplitude value of the signal. The corresponding time value is that which was programmed into the variable delay. The time delay is incremented by the reciprocal of the desired sampling rate, and the next sample

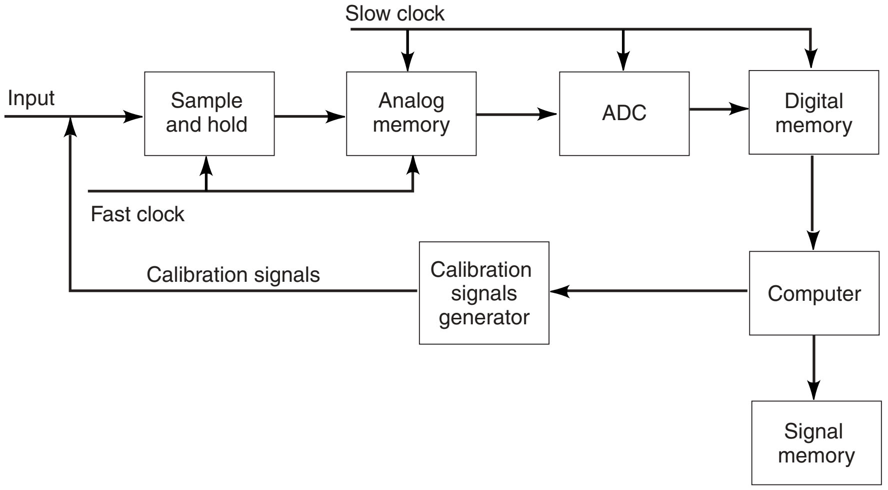

Figure 3. Representation of a recorder using analog memory. The fast clock operates at the sampling rate and stores samples into the memory. The slow clock begins after recording is completed and retrieves data from the memory at a rate at which the ADC can operate. The computer makes corrections to the data to produce a representation of the input signal and controls the calibrations which allow these corrections to be made.

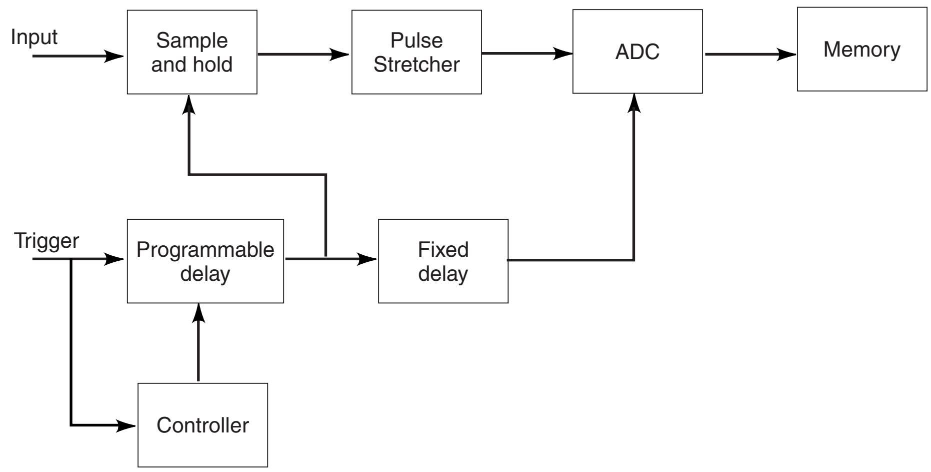

Figure 4. Representation of a sampling oscilloscope. The sample-and-hold produces a pulse whose amplitude is proportional to the signal at the time of the strobe signal. This pulse is stretched to be long enough for its amplitude to be measured by the ADC. One recorded data point is produced for each pair of synchronized input and trigger signals.

is taken. Programming the variable delay to N different delay values gives N samples of the signal at different times. This approach is referred to as *equivalent time sampling,* because the differences in the time values associated with successive samples are not the actual time differences between when the samples were taken but represent equivalent time differences relative to the signal.

The subsystems that make up a modern sampling oscilloscope are very specialized systems. The sample-and-hold circuit must be precisely constructed out of components specifically designed for the purpose. To obtain good linearity, the sample-and-hold usually measures the difference between the applied signal and a feedback signal. The feedback signal is an estimated value of the applied signal. To obtain high accuracy with reasonably long delays, the programmable variable delay is typically an elaborate subsystem.

Multiple Samplers

One can construct recorders that use the very high bandwidth sample-and-hold circuitry of the sampling oscilloscope and record single transients. This is illustrated in Fig. 5 and in Ref. 4. This approach requires one sample-and-hold for each sample of the signal; such recorders have been made which take from 20 to 100 samples. Both the signal to be measured and the strobe signals must be distributed to each of the sample-and-hold circuits. The strobe signal must be delayed to each sample-and-hold taking into account the time, relative to the trigger signal, that the sample is to be taken and the delay time of the signal to the particular sample-and-hold.

The construction of this type of recorder has all the problems covered earlier for interleaved ADCs, but on a larger scale. The signal must be routed to a large number of distinct circuits while maintaining the same frequency response on each path. Since this type of system is used for extremely high bandwidth, the tolerances on the matching of frequency responses (in particular phase shifts) are very tight.

Equivalent Time Sampling

The sampling oscilloscope is the simplest example of using equivalent time sampling to obtain higher sampling rates for repetitive signals. The same principal can be applied to increasing the effective sampling rate of recorders based on real time digitization. This works much the same as the sampling oscilloscope illustrated in Fig. 4, except that the trigger signal causes the recorder to sample the signal at the rate r_1 rather than to take only one sample. To increase the effective sampling rate by a factor of n, $n-1$ additional trigger signals cause the recorder to take records delayed, with respect to the repetitive signal, by amounts of k/r_1. This is illustrated, for $n = 4$, in Fig. 6. This approach is frequently used in digital storage oscilloscopes.

Although equivalent time sampling is applicable to any of the recording methods that have been covered in this article, it is not a trivial matter to add it to a recorder. It requires, with an exception described shortly, synchronization between the sampling times and the repetitions of the signal being measured. This is problematic if the sampling times are derived form a crystal oscillator. One means of solving this problem is called *random equivalent time sampling,* in which the phase relationship between the trigger signal and the sampling clock is made random, and the time is measured (rather than controlled) between the trigger signal and the first sampling time after the trigger. The time value associated with jth amplitude value of the ith record is then given by $t_{ij} = t_0 + j/r_1 - \tau_i$, where t_0 is a constant and τ_i is the measured time between the trigger and the first sampling point. This approach does not give uniform sampling.

Traveling Wave Cathode Ray Tube

The cathode ray oscilloscope has been one of the highest bandwidth recording devices for decades and remains so today. Modern recorders that use a cathode ray tube (CRT) as the recording device have bandwidths in excess of 5 GHz. An advantage of CRT recorders, besides the very high bandwidth, is their ability to withstand input signals hundreds of times their full scale signal range without damage. Historically, the cathode ray oscilloscope has been converted from a display device to a recording device by attaching a camera to record the display on the face of the CRT on film. Film recording was then replaced by an electronic camera whose output was digitized to form a digital recording of the two dimensional image of the CRT. In some modern CRT recorders the phosphorous face, which converts the electron beam to light, is omitted. Instead, the electron beam writes directly onto a CCD array. The CCD array is read out and the image digi-

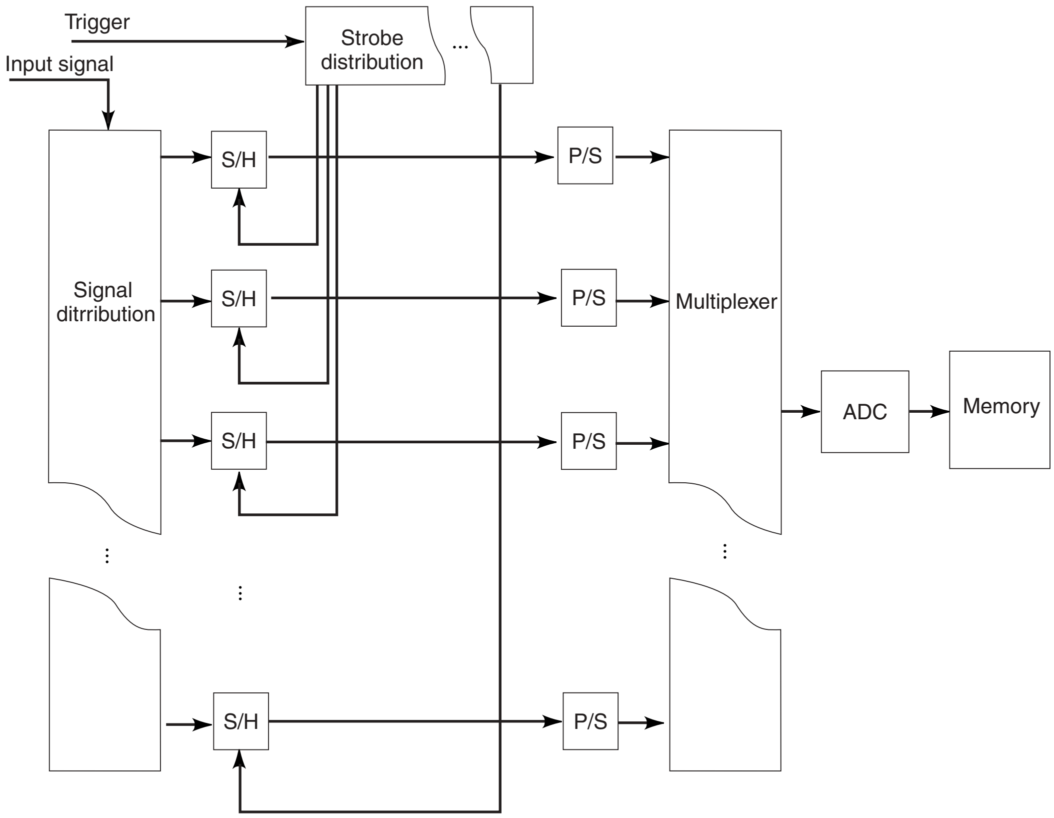

Figure 5. Representation of a recorder using many copies of sampling oscilloscope circuitry to record a single transient.

tized for analysis. The image is analyzed to produce a record of time-volts pairs.

A traveling wave deflection structure is used on the vertical axis to obtain the extremely high bandwidths that give the CRT an advantage over an ADC. In a traveling wave deflection structure the potential difference to be recorded travels as a wave along the deflection structure at the same speed and in the same direction as the electron beam to be deflected. The speed of the electron beam is typically around one-tenth the speed of light. Why use a traveling wave structure? The angle of deflection of the electron beam is proportional to the time an individual electron in the beam remains in the deflection structure. This time must be several nanoseconds to obtain adequate deflection. With a conventional deflection structure, the transition duration of a recorded step signal will be, approximately, the time it takes an electron to traverse the deflection plate. This limits the bandwidth to 100 MHz or so. With a traveling wave structure the transition duration can be reduced by a factor of 100.

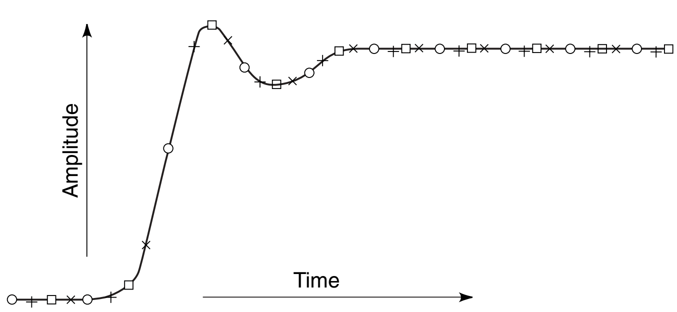

Figure 6. A step signal as recorded using interleaving. Points using the same symbol are recorded with the same ADC. This example shows four interleaved channels.

The errors produced by CRT recorders are generally much larger than one might predict from the published specifications. This is because of limitations inherent in the use of electron beams. Typical specifications might be: bandwidth, 5 GHz; vertical resolution, 11 bits; sweep lengths, form 1 ns to 1 ms; number of data points per record, 1000.

With a record length of 1000 samples, the sampling rate = 1000/(sweep length). The useful sampling rate is typically an order of magnitude or more lower than this, because of the nonzero spot size of the electron beam. The useful sampling rate is obtained by replacing the number 1000 with the number of electron beam radii that fit along the time axis of the writing area of the CRT. This is because signal samples taken at a higher rate than this will interfere with each other, as illustrated in Fig. 7. The figure understates the magnitude of the problem, because the density of sample points per beam radius is likely larger than shown. Figure 8 illustrates the distortion that occurs because of the nonzero beam size. The peak of the pulse is recorded lower than the true value, because points more than one beam radius below the peak are written on by the beam on the rising and falling edges of the pulse. In the illustration it is assumed that the recorded signal will be halfway between the lower and upper edge of the trace; in actual practice the situation is usually worse then

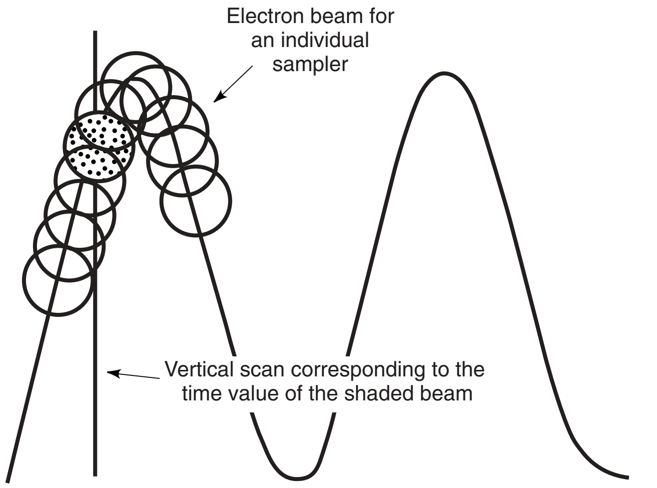

Figure 7. Each circle represents the electron beam striking the face of the CRT corresponding to one time point on the signal. The vertical line represents the scan line of the read out for the time corresponding to the shaded circle. It passes through several circles causing the y-coordinate of the output to be an average of the beam from different times. This results in an effective bandwidth reduction due to the beam size. This phenomenon is not taken into account when reporting the bandwidth of the recorder.

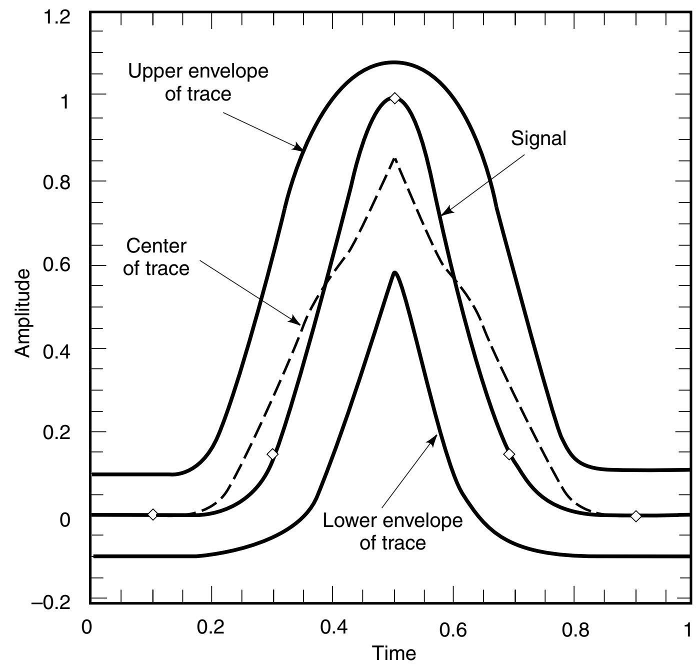

Figure 8. Illustration of the signal distortion which can be caused by the nonzero size of the electron beam in a CRT based recorder. The upper and lower envelopes are determined by having a perfectly circular beam traverse the signal. The read out is taken by scanning the beam vertically and calculating an estimate of the center of the trace. When the display has significant curvature, the center of the trace does not coincide with the signal.

this. The trace will be more intense below the peak than above it, because the beam hits points below the peak twice—once on its way up, and once on its way down. This causes the recorded signal to be even lower than in the figure.

Another type of distortion peculiar to CRT recorders is called *wide-narrow distortion,* which is illustrated in Fig. 9. The input signal is a pure sine-wave. The recorded signal at the left of the display is wider than it should be at the positive peaks and narrower than it should be at the negative peaks. This gradually shifts to the opposite, narrow at positive peaks and wide at negative peaks, on the right side. This problem is inherent in the traveling wave structure. When the signal increases rapidly with time, the voltage across the deflection structure decreases rapidly in the direction of beam travel. This decelerates the electron beam, which effects its deflection in the time direction.

The amplitude resolution of 11 or more bits typically advertised for CRT recorders is misleading. Amplitude values are typically obtained by analyzing the recorded beam intensity profile along a vertical slice (constant time) and calculating the centroid (or some closely related measure of the center of the trace) and relating this to a voltage value. The precision specified is usually the precision to which this calculation is rounded. The actual precision of the resulting data is close to seven bits in current commercially available recorders.

Streak Camera

A streak camera has several similarities to a CRT. Information is recorded by an electron beam striking a phosphor face. The image on the face is digitized and analyzed to determine the amplitude versus time of the recorded signal(s). The time information is obtained, as with a CRT, by sweeping the beam across the face of the *streak tube.* The amplitude information, however, is encoded in the intensity of the beam. The signals to be recorded must be converted to light whose intensity is proportional to the amplitude of the signal. The light strikes a photo-cathode that emits a number of electrons proportional to its intensity. The electron beam is amplified and swept across the face of the tube; see Fig. 10. About 20 channels can be recorded simultaneously on a streak camera. The light for each channel is placed at a different location along the vertical axis of the photo-cathode. The data for each channel is one horizontal streak across the face of the tube.

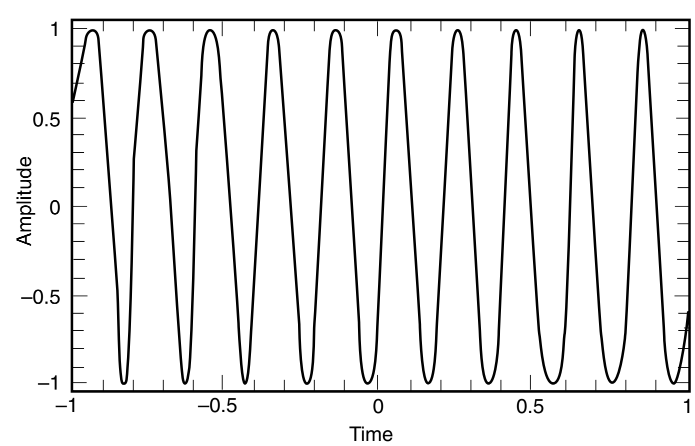

Figure 9. A display demonstrating "wide-narrow" distortion of traveling wave CRTs. The applied sine wave has wide positive peaks and narrow negative peaks on the left, wide negative peaks and narrow positive ones on the right and is relatively undistorted in the center. This phenomenon distorts the recorded width of narrow pulses.

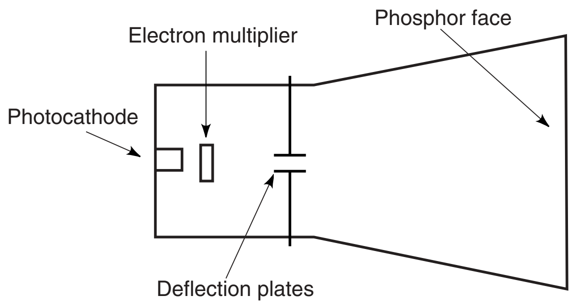

Figure 10. Top view illustration of a streak camera based recorder. Each channel is a dot of light on the photo cathode. The dots are arranged in straight line vertically. Electrons are emitted from each location on the photo cathode in a number proportional to the instantaneous light intensity. After multiplication the electron stream is deflected on the horizontal axis and recorded in the same manner as in a CRT.

The conversion of light to electrons at the photo-cathode is nearly instantaneous, yielding potential bandwidths in excess of 20 GHz. It is often the case in making very high bandwidth recordings that the signals to be recorded are converted to light to be transmitted to the recording station by fiber-optic cable, thus eliminating the high frequency skin effect loss of coaxial cable. In such situations the streak camera is an especially suitable recorder.

RECORDER ERROR MEASURES

One of the most important features of a recorder is the accuracy with which it records. This is a complex subject, because the errors are usually signal dependent. A number of measures of recorder accuracy and test methods for determining them have been established over the years and are described in this article. Most of those described are covered in Refs. 5 and 6; others are covered in Ref. 7.

The errors made by a recorder are often classified as either *static errors* or *dynamic errors.* Static errors are those that are evident in recording DC or very slowly varying signals; dynamic errors are those that are frequency dependent and typically disappear at low frequencies. Sometimes the classification is not clear and is made somewhat arbitrarily. Static errors will be covered first, followed by dynamic errors.

Sine Wave Tests

Sine wave tests are among the most useful tools for evaluating recorder errors. A sinusoidal signal of specified amplitude and frequency is recorded yielding a sequence (t_i, y_i) of time-amplitude pairs. One then performs a least squares fit to the data by determining values of A_0, B_0, C_0, and ω_0 that minimize

$$\sum_{i=1}^{M}[y_i - A_0\cos(\omega_0 t_i) - B_0\sin(\omega_0 t_i) - C_0]^2 \quad (2)$$

where M is the number of data values used in the analysis. Details on the calculation of the unknown parameters can be found in Ref. 5. Once the coefficients are determined the *residuals* are given by

$$r_i = y_i - A_0\cos(\omega_0 t_i) - B_0\sin(\omega_0 t_i) - C_0 \quad (3)$$

The residuals are a good approximation to a significant portion of the errors in the recording process. They can be analyzed in either the time domain or the frequency domain (Refs. (5,6,7,8) and the following sections of this article) to give useful information. Of particular interest is the rms value of the residuals, σ_r, given by

$$\sigma_r^2 = \frac{1}{M}\sum_{i=1}^{M} r_i^2 \quad (4)$$

Since the amplitude, phase, frequency, and dc offset of the input signal are estimated from the data, they don't have to be accurately known to perform sine wave tests. For the same reason, the frequency response of the cabling connecting the sinusoidal signal source to the recorder, and of the recorder itself, is not an important factor in sine wave tests.

If the frequency of the test signal is chosen so that an integer number of cycles occur in one record length, that is,

$$f = \frac{J}{M} f_s \quad (5)$$

where f is the frequency of the test signal, f_s is the sampling rate, and J is an integer, then the least squares Eq. (2) can be solved by taking the discrete Fourier transform of the data. The constant C_0 is the 0 frequency term in the DFT; A_0 and B_0 can be derived from the two terms at the frequency, f, and all other terms in the DFT represent the residuals.

There are two main reasons for the prevalence of the use of sine waves in evaluating recorder errors. The first is that sine waves can be produced with very high accuracy; their purity can be measured to even higher accuracy with a spectrum analyzer, and their purity can be improved on by the use of filters. Current technology doesn't permit the economic production of signals of other shapes that are known to the accuracy we expect of many recorders. The second reason for the prevalence of sine waves in this application is they allow one to ignore the frequency response effects of the recorder and the test setup.

Quantization Error

Quantization error is an attribute of analog to digital converters. It arises from the finite number of output values that can be produced by an ADC. Figure 11 is a plot of the output of an ADC versus the input voltage. Each interval of input voltages over which the output code remains constant is called a *code bin.* The difference between the voltage at the right edge of a code bin and the voltage at the left edge is called the *code bin width.* Ideally, the code bin width is a constant, Q. The voltage value assigned to a particular output code is assumed to be the value at the center of the code bin. The difference between the input voltage and the assigned voltage is called the *quantization error.* If the input voltage is equally likely to be any value in the range of the ADC, then the quantization error is equally likely to be any value between $-Q/2$ and $+Q/2$. In this case the average quantization error is zero and the rms quantization error is given by

$$\text{rms quantization error} = \sqrt{\int_{-Q/2}^{+Q/2} \frac{x^2\,dx}{Q}} = \frac{Q}{\sqrt{12}} \quad (6)$$

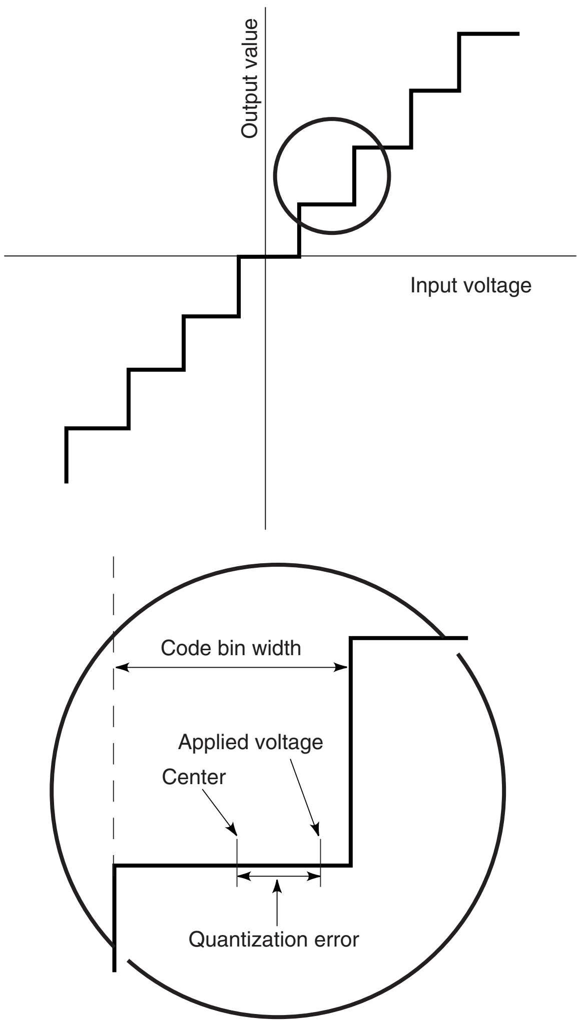

Figure 11. The stair step transfer function of an ADC.

Although the quantization error is a deterministic nonlinearity in the transfer function, it is frequently referred to as *quantization noise.* In many cases the sequence of values of the quantization error is most practically analyzed as white noise with an rms value given by (6).

Gain and Offset Errors

If one thinks of a recorder as an amplifier followed by an ideal recorder, then the gain and offset errors are the deviations of the amplifier gain and offset from their nominal values. There are different means of measuring these errors that yield slightly different results. For a sequence of ordered pairs (v_i, v_r), where v_i is an input voltage and v_r is the corresponding recorded voltage, the gain and offset are defined by the following relation,

$$v_i = Gv_r + v_0 + \epsilon \tag{7}$$

where G is the gain, v_0 is the offset error, and ϵ is the error (different for each value of v_i). The values for v_r in this equation are already corrected for the nominal gain and offset of the recorder; thus, if the recorder were ideal, we would have $G = 1$ and $v_0 = 0$. Values of G and v_0 are determined by fitting a straight line to the (v_i, v_r) pairs. Different means of fitting the line will yield different values. Values for v_i should be selected so that quantization error has negligible effect on the results. Values of v_r should be obtained by averaging so that random noise has negligible effect on the results. The *gain error* is $G - 1$ and is usually expressed as a percentage. The offset error is expressed in input units (e.g., volts).

The gain and offset errors are treated separately from other errors, because they are often larger than other errors. If necessary, the gain and offset errors can be corrected for by using the results of recording two known voltages.

Dynamic Range and Noise

The use of the terms dynamic range and noise is not totally consistent in the industry, reference (9). The most commonly used definition for dynamic range for recorders is

$$\text{dynamic range} = 20\log\frac{\text{maximum rms signal}}{\text{rms noise}}\ \text{dB} \tag{8}$$

Both the numerator and denominator in Eq. (8) are subject to multiple interpretations. The numerator is usually taken to mean the maximum rms sine wave which, for a recorder that covers the voltage range of from $-A$ to A, is $A/\sqrt{2}$. The denominator is usually taken to be the sum (in quadrature) of the noise present with no signal and the quantization noise. This quadrature combination is accurately approximated by the rms residuals from a sine wave test with a low amplitude ($5\,Q$ to $20\,Q$) input signal. For audio recorders the rms noise is often calculated with a frequency domain weighting (A weighting) based on the sensitivity of human hearing to low level sounds (11). For an ideal N bit digital recorder, one in which the only source of noise is quantization noise, the dynamic range is given by

$$\text{maximum dynamic range} = 6.02N + 1.76\ \text{dB} \tag{9}$$

Table 1 gives the dynamic range from Eq. (9) for typical values of N.

Integral Nonlinearity

The concept of integral nonlinearity (INL), as it has been used for decades, is illustrated in Fig. 12. The curve represents the voltage reported by the recorder as a function of the applied voltage. The straight line represents a best fit to this curve.

Table 1. Maximum Attainable Dynamic Range as a Function of Number of Bits

Number of Bits	Maximum Dynamic Range (dB)
6	38
8	50
10	62
12	74
14	86
16	98
18	110
20	122
22	134
24	146

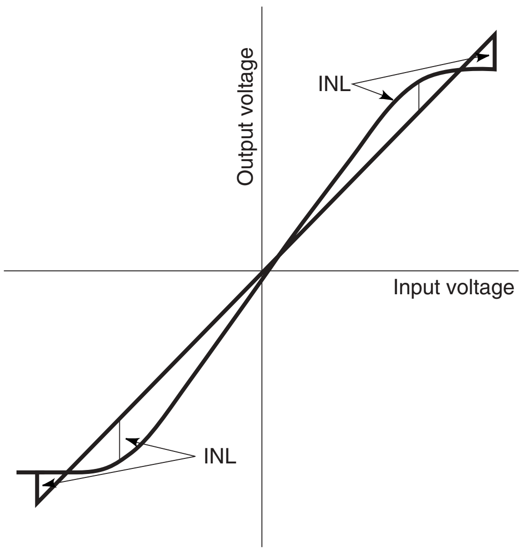

Figure 12. Illustration of INL as defined for analog instruments. The curved line is the dc transfer function of the instrument. Its maximum deviation from a straight line (occurring at four different places in this figure) in the integral nonlinearity.

The INL is the maximum difference between the curve and the straight line. The difference is typically expressed as a percentage of full scale for analog instruments.

For recorders that use an ADC the situation is a little more complicated. As seen in Fig. 11, the ideal curve for an ADC is a stair step rather than a straight line. The stair step necessarily differs from the straight line by $\pm Q/2$. This difference is not considered part of the integral non linearity. For an ADC only the values of the transition levels, the values of input signal for which the output jumps from one code to the next, are considered. The INL is the maximum difference between any transition level and its ideal value. Determining the INL for an ADC based recorder requires determining the transition levels. This is a lengthy process covered in Refs. (5,6,7). If the recorder has a noise level significantly larger than the quantization noise, the INL can be estimated without determining the transition levels. Points on the curve of Fig. 12 can be measured by recording the average outputs to known arbitrary inputs. This works, because the noise straightens out the stair step in the average measurements. Table 2 gives the amount of apparent INL caused by the stair step as a function of the rms noise divided by Q.

Table 2. Improvement in Analog INL Due to Noise

σ/Q	INL
0	.5
.01	.48
.02	.47
.05	.38
.1	.29
.2	.14
.3	.054
.5	2.2×10^{-3}
.6	2.5×10^{-4}
.7	2.0×10^{-5}

Differential Nonlinearity

The concept of differential nonlinearity (DNL), like integral nonlinearity, has been used for decades in analog instrumentation. It is defined in terms of the derivative of the recorded voltage with respect to the input voltage, and is a measure of how much this derivative deviates from a constant. Letting

$$S_M = \max\left[\frac{dv_r}{dv_i}\right], \quad \text{and } S_m = \min\left[\frac{dv_r}{dv_i}\right]$$

DNL is defined as

$$\text{DNL} = 200\frac{S_M - S_m}{S_M + S_m}\% \qquad (10)$$

This is illustrated in Fig. 13, which is the derivative of the transfer function in Fig. 12.

For ADC based recorders the derivative of the stair step like ideal transfer function jumps between zero and infinity, and Eq. (10) is not appropriate. A perfectly linear ADC is one for which all code bin widths are constant. DNL for an ADC is the following measure of the maximum deviation of any code bin width from a constant

$$\text{DNL} = \max\left|\frac{W(k) - Q'}{Q'}\right| \qquad (11)$$

where $W(k)$ is the width of the kth code bin, and Q' is the average code bin width. The unit typically supplied with this measure is lsb (for least significant bit), while "code bin width" would be a more grammatically correct unit. For ADCs whose noise level is larger than their code bin width, DNL is not a very useful specification.

Frequency Response

The frequency response characterizes the error that is usually the dominant dynamic error of a recorder. Use of a frequency response to characterize a system is only valid if the system is linear and time invariant (2). This is a good approximation for most high quality recorders. One can imagine the recorder as a linear time invariant filter followed by blocks that pro-

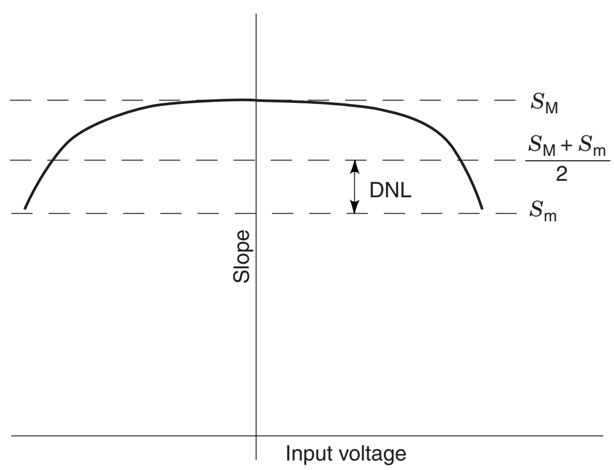

Figure 13. Illustration of DNL for an analog instrument. The curve is the derivative of that in Fig. 12.

duce all of the other sources of error of the recorder. For signals for which the errors are small (<10%), the order of the blocks has a second order effect on the overall error. The error produced by the equivalent filter at the input of the recorder is the *frequency response error.* Keep in mind that the frequency response error may actually be caused by limitations in several parts of the recorder; placing the problem in a single filter is just a mathematical convenience.

The quantity of most interest is actually the *impulse response,* $h(t)$, of the recorder. If $x(t)$ is the input signal to be recorded, the recorded signal (accounting only for frequency response errors) is given by

$$r(t) = \int_{-\infty}^{\infty} x(t - t')h(t')\,dt' \tag{12}$$

The integral in Eq. (12) is called the *convolution integral.* It allows one to calculate the recorded signal for any input signal, and thus estimate the frequency response error—the difference between $r(t)$ and $x(t)$. This error depends on the signal being recorded.

The frequency response is the Fourier transform of the impulse response,

$$H(f) = \int_{-\infty}^{\infty} h(t)\exp(-j2\pi ft)\,dt \tag{13}$$

The complex frequency response is usually expressed as a gain, $G(f)$, and a phase shift, $\phi(f)$, both of which are real valued and are related to $H(f)$ by

$$H(f) = G(f)\exp[j\phi(f)] \tag{14}$$

These are the gain and phase shift observed when recording a sinusoidal signal of frequency, f. In both Eqs. (13) and (14) j is the imaginary unit. For an ideal recorder (one with no frequency response error) $G(f)$ will be unity for all f, and $\phi(f)$ will be a linear function of f.

Typically the specifications for a recorder will give a lower and upper frequency between which $G(f)$ is within a certain tolerance of unity. This tolerance is usually 3 dB, but sometimes 1 dB is used. One should bear in mind that with a tolerance of 3 dB (1 dB) this specification only gives a range of frequencies for which the gain error is less then 30% (11%), and tells nothing about phase shift errors. If one needs to record signals to better accuracy than these tolerances, more detail concerning the frequency response error is required. A convenient approach (5) is to determine the gain error and non-linear phase error as a function of frequency. The gain error is $|1 - G(f)|$, and is expressed as a percentage. The non-linear phase error is the phase shift with a straight line subtracted, usually the line that passes through the origin and passes through the phase shift curve at a phase of $\pi/2$ rad (45°).

Measuring the Frequency Response. The most common method for measuring the frequency response of a recorder is to record the step response, numerically differentiate it (using a simple first difference) to determine the impulse response, and calculate the frequency response by taking the discrete Fourier transform of the result. The step response is recorded rather than the impulse response for a number of reasons. It gives more accurate results at low frequencies at which accuracy is most desired; it is less subject to aliasing errors than the impulse response (12), and the amplitude of the recorded step response is more easily predictable. Other types of signals are used to determine the frequency responses of systems (13), but they have not found wide acceptance in recorders.

If one only wants crude information about the frequency response, the results of sine wave tests can be used. This can determine if the gain error at certain frequencies is less some tolerance. The ratio of the fitted amplitude of the sine wave to the known input amplitude is the gain at the frequency of the test signal.

Step Response

Frequently, characteristics of the step response of a recorder are used as quality measures. These include the settling time, the amount of overshoot and the 10 to 90% transition duration (or rise time). The *settling time* is the amount of time it takes the step response to stabilize to within a certain tolerance of its final value. For example, a specification that the settling time is 5 μs to 0.1% means that for any time later than 5 μs after the onset of the step, the recorded value stays within 0.1% of its final value. Settling time is frequently the only dynamic specification that relates to precision measurements. The amount of overshoot is zero unless the value of the step response exceeds its final value at some time. Otherwise, the overshoot is the difference between the peak value of the step response and the final value, expressed as a percentage of the final value. Sometimes an overshoot specification will be contingent on the 10 to 90% transition duration of the input signal being longer than some specified value. The settling time and overshoot are illustrated in Fig. 14.

The 10 to 90% transition duration is a crude measure, as is the 3 dB bandwidth, of the response time of the recorder. Two step responses can have the same rise time and have much different distorting effects on a signal. To seriously estimate dynamic errors one should have a complete step response record and use the convolution integral to estimate distortion. Software to do the calculation is readily available.

Noise and Distortion

The sine wave tests described earlier can be used to measure the combined effect of many error sources. The error sources

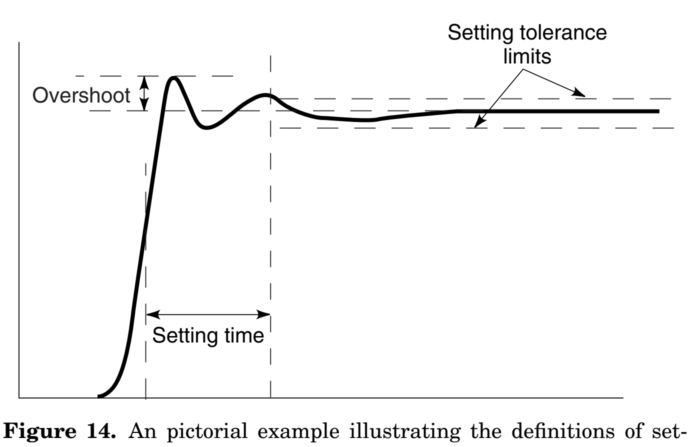

Figure 14. An pictorial example illustrating the definitions of settling time and overshoot.

left out are gain and offset errors, frequency response errors, and errors associated with uncertainties in trigger delay. Most other sources of error will show up in sine wave tests. In fact, one of the advantages of these tests is that they expose sources of error that aren't even known to the tester.

Signal to Noise Ratio, Effective Bits and Dynamic Range. Of greatest utility is the rms value of the residuals, given by Eq. (4). This is referred to as the total *noise and distortion,* and will usually depend of both the amplitude and frequency of the applied signal. The noise and distortion is often referred to as simply the noise. A related quantity is the *signal to noise ratio* (SNR), given by

$$\mathrm{SNR} = 20\log\left(\frac{A}{2\sigma_r}\right)\ \mathrm{dB} \tag{15}$$

where A is the rms amplitude of the applied signal, and σ_r is the rms value of the residuals. The amplitude is selected so that the signal covers nearly the entire range of the recorder. Such signals are called *large amplitude* signals. The SNR looks very much like the dynamic range, given by Eq. (8). They are both the ratios (in decibels) of the rms amplitude of a large sine wave divided by the rms value of the noise. There is, however, a significant difference. For SNR, the noise used in the ratio is the noise that occurs in the presence of the large signal. For dynamic range it is the noise in the presence of a small signal.

Another commonly used related quantity is the *effective bits, E*, given by

$$E = \log_2\left(\frac{V}{\sqrt{12}\sigma_r}\right)\ \mathrm{bits} \tag{16}$$

where V is the full scale range of the recorder (maximum recordable voltage minus minimum recordable voltage). If the only source of noise were quantization noise, then $E = N$, the number of bits. The number of effective bits is reduced by one for each doubling of the rms noise. The SNR and the effective bits are related by,

$$SNR = 6.02E + 1.76 + 20\log\left(\frac{2A}{V}\right)\ \mathrm{dB} \tag{17}$$

For large amplitude signals (>90 % of full scale) the last term is between -0.9 and 0. If E_0 is the effective bits measured for a very small input signal, the dynamic range is given by (see Eq. 9)

$$\text{dynamic range} = 6.02E_0 + 1.76\ \mathrm{dB} \tag{18}$$

Time Errors Versus Amplitude Errors

One can separate errors on the time axis from errors on the voltage axis by examining the variation in effective bits with frequency. If one assumes errors τ_i on the time axis and e_i on the voltage axis that are independent of each other and independent of the signal, then the recorded signal for sinusoidal input is

$$\begin{aligned} y(t_i) &= A\sin(2\pi f t_i + \tau_i) + e_i \\ &\cong A\sin(2\pi f t_i) + e_i + 2\pi A f \tau_i \cos(2\pi f t_i) \end{aligned} \tag{19}$$

where τ_i is the ith discrete sampling time, and the error induced on the voltage axis by that on the time axis is estimated using a first order Taylor expansion of the sin function. The second and third terms are a very good approximation to the residuals. This gives for the rms residuals

$$\sigma_r^2 = \sigma_e^2 + 2\pi^2 f^2 A^2 \sigma_\tau^2 \tag{20}$$

where σ_e and σ_τ are the rms values of e_i and t_i. The first term dominates at low frequencies, the second term at high frequencies. Figure 15 shows this converted to effective bits and plotted against frequency on a logarithmic scale. At low frequencies this is constant; at high frequencies it has a slope of -1 bit per octave. The time component of the error can be determined from the frequency f_b at which the two lines intersect by the relation

$$\sigma_\tau = \frac{\sigma_e}{\sqrt{2}\pi f_b A} \tag{21}$$

where σ_e is the rms value of the residuals at low frequency.

Harmonic Distortion. If the recorded signal resulting from an applied sine wave contains frequencies that are multiples of the applied frequency, the recorder is said to have harmonic distortion. Harmonic distortion is a consequence of nonlinearity. For a static nonlinearity for which the recorded signal, $v_r(t)$, is given by $v_r(t) = f(v_i(t))$, there is a direct relation between the function, f, and the harmonic distortion. If f is a polynomial of degree n, than there will be harmonic distortion up to order n. If a signal of the form $v_i(t) = A\cos(\omega t)$ is applied, and we let $f_A(x) = f(x/A)$ and expand $f_A(x)$ as

$$f_A(x) = \sum_{k=0}^{n} a_k T_k(x) \tag{22}$$

where the $T_k(x)$ are the Chebychev polynomials (13,14), then the recorded signal has the form

$$v_r(t) = A\sum_{k=0}^{n} a_k \cos(k\omega t) \tag{23}$$

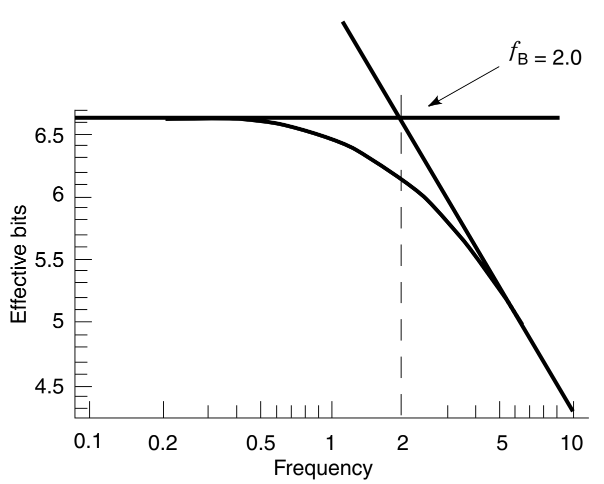

Figure 15. Effective bits versus frequency for a combination of amplitude errors and time errors. The time errors cause the 1 bit per octave roll off at high frequency. The frequency, f_B of the intersection of the two straight lines can be used in equation 21 to estimate the rms value of the time errors.

Harmonic distortion is specified as a ratio of the combined rms value of the harmonics to the rms value at the signal frequency. The ratio is expressed either as a percentage or in dB. The units used when expressing it in dB are dBc, meaning decibels relative to the carrier. For the example of Eq. (23) we have

$$\text{harmonic distortion} = 10\log\left[\frac{\sum_{k=2}^{n} a_k^2}{a_1^2}\right]\ \text{dBc} \qquad (24)$$

Harmonic distortion typically increases with amplitude and is usually measured with large amplitude signals.

The source of harmonic distortion described previously is closely related to INL, a static nonlinearity, and is, therefore, independent of the frequency of the applied signal. Harmonic distortion can result from dynamic nonlinearity and be frequency dependent. A typical example of a dynamic nonlinearity is a time delay that is proportional to the slew rate of the signal, that is, $v_r(t) = v_i(t - \tau)$, where $\tau = \epsilon(\mathrm{d}v_i/\mathrm{d}t)$. Estimating the time delay effect by first order expansion as for Eq. (19) gives

$$v_r(t) = v_i(t) - \epsilon(\mathrm{d}v_i/\mathrm{d}t)^2 \qquad (25)$$

Substituting a sinusoidal signal for $v_i(t)$, one obtains second harmonic distortion which is proportional to the signal's amplitude squared and to its frequency squared.

Harmonic distortion is most conveniently measured by performing a sine wave test and calculating the DFT of the recorded signal. The record length must be an integer number of cycles of the applied signal (see Eq. 5) and a large signal should be used. The DFT values are converted to decibels and the dB value for the applied frequency is subtracted from all of the values. This gives harmonic distortion readings directly in dBc. Figure 16 shows a simulated example result. The recorder sampling rate is 2 units, and the record length for the DFT is 1024 points. The input signal was full scale in amplitude with a frequency of 0.2012. The recorded signal has harmonic distortion given in Fig. 16 and had eight bit quantization noise applied.

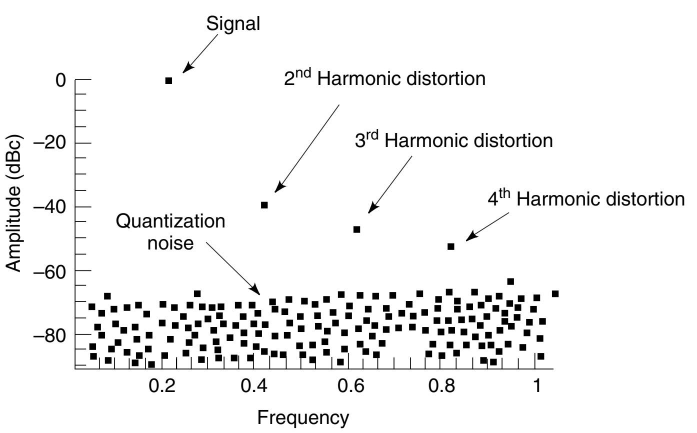

Figure 16. DFT of sine wave test results showing 1% second 0.5% third and 0.25% fourth harmonic distortion. A record length of 1024 points was used. The points below −60 dBc are the result of 8 bit quantization noise. This noise floor can be made as low as desired by using a sufficiently long record length.

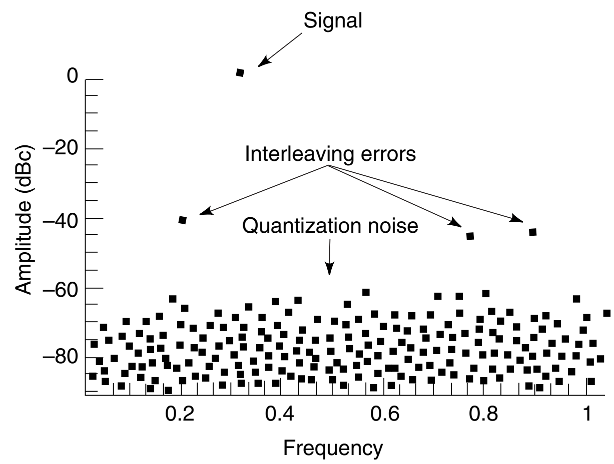

Figure 17. DFT of sine wave test results showing the results of the interleaving errors for the example given in the text. The interleave error frequencies given by Eq. (26) are 0.203, 0.703, and 0.797.

Interleaving Errors and Noise Spectrum. The methods used to measure harmonic distortion can be used to quantify interleaving errors. When a number of recording channels is interleaved to increase the sampling rate (see section titled "Interleaved ADCs"), errors result from the various channels not being perfectly matched. These are called interleaving errors. The interleaving factor, n, is the number of recorders interleaved. The sampling rate for each recorder is f_s/n, where f_s is the sampling rate of the recorder. When a sine wave of frequency f is recorded, interleaving errors produce additional signals at frequencies

$$\text{interleave error frequencies} = kf_s/n \pm f \qquad (26)$$

for integer values of k. When these frequencies are larger than the Nyquist frequency, $f_s/2$, they get aliased down to frequencies in the Nyquist band. However, at most $n - 1$ distinct frequencies below the Nyquist frequency are produced.

The measurements and data reduction for examining interleaving errors are identical to those for harmonic distortion. There is, however, an additional restriction on the input frequency, f. The frequency should satisfy

$$f = \frac{Kn}{M} f_s \qquad (27)$$

where K is an integer, and M is the record length. This causes all of the frequencies created by the interleaving errors to be DFT frequency values. Figure 17 shows results with $f_s = 2$, $n = 4$ and $f = 0.297$. All of parameters are the same as for Fig. 16 except that the harmonic distortion has been replaced by distortion due to the four interleaved channels having different gain errors. The gain errors for the channels are (in succession) 3%, 1%, −1% and −3%.

Spurious Signals—Spurious Free Dynamic Range. Spurious signals are any frequencies, other than harmonics of the applied signal, that occur in the recorded signal. The frequencies generated by interleaving errors are just one example. The spectral lines related to the spurious signals are called *spurs.* The test methods used to determine the magnitudes of the spurs are identical to that of the last two sections.

The negative of magnitude of the largest spur (in dBc) is called the *spurious free dynamic range* (SFDR) of the recorder. This specification is much more common for ADCs themselves than recorders in which they are used. Note that the SFDR can be much larger than the previously defined dynamic range, which was based on the noise level in the time domain. The observed noise floor (due to random noise, quantization noise, and various other sources) is much smaller in the frequency domain than in the time domain, and it can be made as small as desired by increasing the record length. Therefore, the SFDR is unaffected by this noise. SFDR is only meaningful if the recorder is going to be used in narrow band applications, such as communications or spectrum analysis.

Intermodulation Distortion. When two sine waves of different frequencies, f_1 and f_2, are passed through a nonlinear system, new signals are produced with frequencies $k_1f_1 \pm k_2f_2$, for integer values of k_1 and k_2. This is referred to as *intermodulation distortion* (IMD). The phenomena in the recorder that are responsible for IMD are the same as those responsible for harmonic distortion. However, there are situations in which harmonic distortion measurements will lead one to believe that the recorder is more accurate than it actually is. When input frequencies are above one-half of the bandwidth of the recorder, harmonics can be attenuated. This can hide some of the high frequency errors. If two high frequency signals are summed, the nonlinearities will cause a number of spectral lines that are well within the bandwidth of the recorder to be produced, making the errors easier to detect and quantify.

There is another situation in which IMD measurements have an advantage over harmonic distortion measurements. One can use test signals with harmonic distortion larger than the distortions one wishes to measure, because the harmonics will be at different frequencies than the IMD products.

Noise Spectrum for Cathode Ray Tube Based Recorders. Sine wave tests and frequency domain analysis of the residuals are also useful for recorders that use a CRT as the recording device. However, because the error phenomena are different some changes in approach have to be made. Two potentially substantial sources of error in a CRT, distortion of the image due to misalignment and time base nonlinearity, produce errors at the same frequency as the applied sinusoidal signal. The spectrum of the errors is spread out over a frequency range on the order of the reciprocal of the record length. This causes two difficulties when one attempts to use data displays such as those of Figs. 16 and 17. First, since the errors are at the same frequency as the signal, they become masked by the signal. Second, since the errors are spread out over a frequency band, their rms value isn't related to the peak value in the spectrum but to an integral. The first problem can be eliminated by performing the DFT on the residuals rather than on the recorded signal, the second by displaying the integral of the power spectrum (8).

The types of distortion shown in Figs. 8 and 9 also have particular signature in the frequency domain. The distortion at the signal peaks due to the spot size exhibits itself as odd order harmonic distortion, which is independent of amplitude and increases rapidly with frequency. The wide-narrow distortion has the same form as Eq. (25) with ϵ a linear function of t that passes through 0 near the mid point in time of the record. This produces second harmonic distortion and a zero frequency distortion proportional to $(Af)^2$, where A is the amplitude and f is the frequency of the input signal. The distortions at these two frequencies are also spread out over a frequency ranges on the order of the reciprocal of the record length.

The random noise and the quantization noise of ADC based recorders are typically white, having nearly constant power spectrum amplitude up to the Nyquist frequency. This is not true for CRT based recorders. The random noise for CRT recorders begins rolling off with increasing frequency at a cut-off frequency related to the spot size of the CRT beam. This limits the SNR enhancement that can be attained through low pass filtering.

Time Base Errors

The two types of time base errors are long term stability and aperture errors. The long term stability is a measure of the long term drift of the sampling rate from its nominal value. This was covered in the earlier *Time base accuracy* section. This error shows up in sine wave tests as a deviation between the fitted frequency and the actual frequency of the input signal. This error doesn't show up in the residuals.

Aperture errors do show up in the residuals of sine wave tests. In the *Noise and distortion* section various sources of aperture error were related to distortions that appear in sine wave tests and whose magnitudes increase with the frequency of the input signal.

OTHER RECORDER QUALITY ATTRIBUTES

There are some other recorder attributes that frequently appear in specifications that don't quite fit into the previous categories.

Cross Talk

Cross talk is the phenomenon, in multichannel recorders, of a small part of the signal into one channel appearing in other channels. The cross talk for any pair of channels, numbered i and j, is defined as

$$c_{ij} = \frac{\text{rms signal recorded on channel } i}{\text{rms signal applied to channel } j} \tag{28}$$

This is measured with all channels terminated in their usual source impedance and with a sinusoidal input signal applied only to channel j. The cross talk typically increases with frequency. The cross talk in practice is the sum of the values from all channels.

Overvoltage Recovery Time

The situation in which the input signal to a recorder exceeds the recording range for a period of time but doesn't exceed the maximum safe input voltage of the recorder is called an overvoltage. The *overvoltage recovery time* is the length of time, after the overvoltage has been removed, for recording at the specified accuracy to resume. The recovery time may depend on the magnitude and duration of the overvoltage.

Trigger Delay and Trigger Jitter

Most high speed (above audio frequency) recorders begin recording upon the receipt of a trigger signal. The *trigger delay* is the length of time between the trigger signal and the first time in the recorded signal. This time can be negative for recorders with pretrigger capability. The trigger jitter is (with the exception given shortly) the standard deviation of the trigger delay. For recorders with a continuously running clock signal, there is an inherent jitter of $1/f_s\sqrt{12}$, where f_s is the sampling rate. Some recorders provide a measurement of the time between the trigger signal and the first clock signal. For these recorders the trigger jitter is the standard deviation of the error in this measurement.

Common Mode Rejection Ratio

Common mode rejection ratio (CMMR) applies to recorders with differential amplifiers at the input. The meaning of CMMR for recorders is the same for amplifiers (15), the ratio of the output for differential signals to that for common mode signals of the same amplitude. The CMMR is especially easy to measure for recorders, because a DFT can be performed on the recorded data to more accurately determine the magnitude of a small common mode signal that is below the noise level. The CMMR is usually frequency dependent, generally increasing with increasing frequency.

RECORDER APPLICATION COMMENTS

Determining Recorder Sampling Rate and Bandwidth Requirements

The cost of recorders can increase drastically as sampling rate and bandwidth increase. Furthermore, the amount of data that must be stored increases linearly with sampling rate for a fixed record length (measured in time units). Hence, overestimating the bandwidth and sampling rate requirement for a recording system can drastically increase its cost, while underestimating them can lead to complete loss of the information that one desires to record. Thus, proper estimation of these requirements is an important part of recorder selection. The author has frequently seen errors of a factor of ten in estimated sampling rate requirements and has found errors by a factor of one hundred not to be uncommon.

The Sampling Theorem and Its Misuses. The sampling theorem (1,2) states that a signal of bandwidth f_B Hz can be exactly reconstructed from samples taken at a rate of $2f_B$ Sa/S. There is a common folklore that says that, to be safe, one should record a signal at a rate of ten times its bandwidth. These concepts lead people to ask "What is the bandwidth of the signal to be recorded?" as the first step in determining sampling rate requirements. This is, in most practical situations, a fruitless approach. The bandwidth in the sampling theorem refers to a frequency above which the Fourier transform of the input signal is exactly zero; real signals never have this property for any frequency upper limit. So for real signals the "bandwidth" is not a well-defined quantity. Often people will use the -3 dB bandwidth of a signal as an approximation. This, coupled with the five times over sampling by sampling at ten times the bandwidth rather than two times, seems reasonable. The error in this will be illustrated with a brief example.

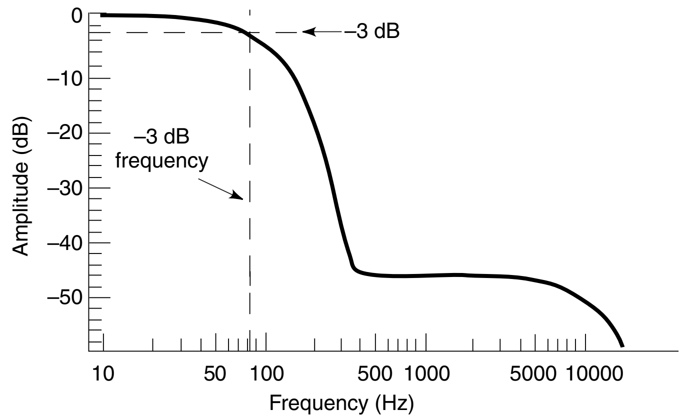

Figure 18. The magnitude of the Fourier transform for the signal in Fig. 1. The shelf at -46 dB is from the narrow pulse at -10 ms.

Consider the signal of Eq. (1) which is shown in Fig. 1. Figure 18 shows the magnitude of the Fourier transform of this signal; the -3 dB point is at 83 Hz. However, from Fig. 1 one can see that sampling the signal at 1 kHz (more than ten times the 3 dB bandwidth) can completely miss the peak that occurs early in the signal. Figure 2 shows that passing the signal through a system with a 500 Hz bandwidth removes the peak. Thus, if one wishes to record the first peak, the common rules of thumb completely fail. However, if the first peak (and any other narrow peaks) were noise, then it would be desirable to use a lower bandwidth and sampling rate to reduce the cost of the recording system.

Use of the Convolution Integral. The best way to determine the bandwidth requirement of a recorder is through simulation. One obtains sample input signals, simulates passing them through filters of various bandwidths, examines the result at each bandwidth and determines the smallest bandwidth that produces acceptable distortion. Simulations are done by evaluating the convolution integral, Eq. (12). To do this one must have sample input signals, $x(t)$, and estimates of an impulse response, $h_B(t)$, parameterized by the bandwidth. The exact form of $h_B(t)$ is seldom critical. A particularly convenient form to use in simulations is

$$h_B(t) = (t/\tau^2)\exp(-t/\tau), \text{ with } \tau = 0.102/B \qquad (29)$$

where B is the -3 dB bandwidth of the filter. With this impulse response the convolution can be rapidly carried out using an IIR filter (1,2,3). These filters are available in many software packages. Another useful impulse response model is the Gaussian,

$$g_B(t) = (1/\sqrt{2\pi}\sigma)\exp(-t^2/2\sigma^2), \text{ with } \sigma = 0.133/B \qquad (30)$$

Which of these one uses makes little difference. Of course, if one is interested in the adequacy of a particular recorder for which the impulse response has been measured, the known impulse response could be used.

There are situations in which estimates can be made with very simple calculations. Often the requirement is that pulses with widths exceeding a certain value have their peak values recorded to a certain tolerance. Assuming both the pulse to be recorded and the impulse response of the recorder to be Gaussian, one can use the relation

$$e = \frac{4.9}{(BW)^2}\% \tag{31}$$

where B is the -3 dB bandwidth of the recorder, W is the width (full width at half maximum), and e is the error in the peak amplitude of the pulse. This relation is valid for $e \leq 20\%$. For the example used previously of Eq. 1 and Fig. 1, the width of the first peak is $W = .24$; to obtain an error in the peak of less than 4.9% requires that $B \geq 1/W = 4.2$ kHz.

Use of Antialiasing Filters

An antialiasing filter is a low-pass filter applied to a signal before it reaches the recorder. The cut-off frequency of the filter is at (or slightly higher than) half the sampling rate (the Nyquist frequency) to be used in recording. When a recorder's bandwidth greatly exceeds half of its sampling rate, antialiasing filters should be employed. If not, both noise and signal components at higher than the Nyquist frequency will be aliased down to lower frequencies where they can't be filtered out. This problem is greatest when recorders with variable sampling rate are used at much lower than their maximum rate, because the bandwidth of the recorder will typically be suitable for the highest sampling rate.

If the noise in the signal source is white, the antialiasing filter will reduce the noise by a factor of $10\log(B_f/B_r)$ dB, where B_f and B_r are the bandwidths of the filter and the recorder. The bandwidths in the preceding are the *equivalent noise bandwidths,* rather than the -3 dB bandwidths, but the difference in these is less than 11% for filters of second order or higher. Note that the antialiasing filter does not reduce noise that is produced within the recorder (often a major source) only the external noise.

Noise Reduction through Oversampling

Noise can be reduced and accuracy and dynamic range improved by sampling the signal at a higher sampling rate than required followed by digital filtering of the recorded signal. This process, called *oversampling,* reduces the effect of noise produced within the recorder. If B is the bandwidth required to record the signal, then recording it at a sampling rate of $nB/2$ and filtering the recorded signal to a bandwidth of B reduces the noise level by $10\log(n)$ dB for the case of white noise. The factor, n, is called the oversampling ratio. To gain a precision of k bits requires an oversampling ratio of $n = 4^k$. The value of n required for a given value of k is shown in Table 3.

Table 3. Extra Bits of Precision Attainable by Oversampling

Oversampling Ratio	Extra Bits of Precision
4	1
16	2
64	3
256	4
1024	5
4096	6
16384	7
65536	8

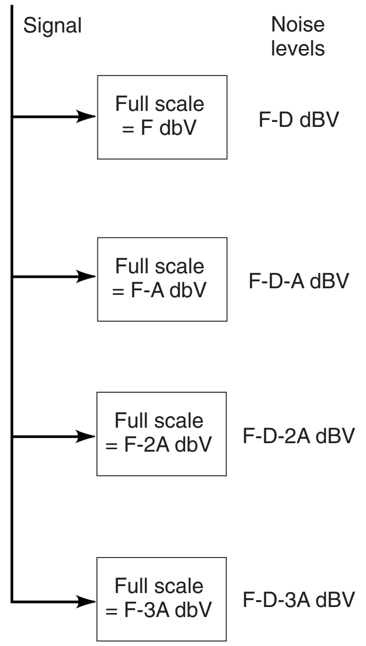

Figure 19. Four recorders used to record 1 signal. Each is A dB more sensitive than the one above it. The signals recorded by the four units are assembled together as described in the text to construct one record.

Buying a recorder with much higher sampling rate than required in order to apply this technique is usually not cost effective. A 256 MSa/s 8 bit recorder can emulate the accuracy of a 1 MSa/s 12 bit recorder with this technique. The 1 MSa/s 12 bit recorder would probably be cheaper to buy and would require much less (a factor of 170) storage capacity. The oversampling approach is, however, cost effective when the higher sampling rate recorder is already available.

The success of oversampling to reduce noise requires that the noise not be concentrated at frequencies lower than B, the bandwidth of the digital filter. This low frequency concentration of the noise can occur when the primary error source is quantization noise (when the number of effective bits is approximately equal to the number of bits) and the signal is nearly constant. In this situation performance can be significantly improved by adding noise or dither to the signal.

Increasing Dynamic Range with Multiple Recorders

In many applications the average signal level varies with time, and the precision required during a particular time interval is a fixed percentage of the average signal level in the interval. In some cases the dynamic range of expected signal levels exceeds that which can be accurately recorded with a single recorder. In these situations the dynamic range can be increased by using multiple recorders as in Fig. 19. The signal is sent to several recorders, and each recorder is set for its full scale voltage to be reduced by A dB with respect to the previous recorder. The data from the three recorders is assembled into a single record after recording is completed by using, for each data point, the largest value from the various recorders that isn't off scale. One must make gain and offset corrections for each recorder to avoid discontinuities in the data.

The optimum value for A and the required number, N, of recorders can be calculated in a straightforward manner. Let V_{max} be the largest signal to be recorded, D be the dynamic range of the recorder (Eq. 8) in dB and S be the required signal to noise ratio for the signal recording in dB. To use the recorders for this application at all we must have $D > S$. If V_{min} is the smallest signal level that must be recorded with signal to noise ratio, S, we define the signal dynamic range, $R = 20\log(V_{max}/V_{min})$. The number of recorders required is then $N = R/S$, and the ratio in dB between the ranges of successive recorders is $D - S$.

FUTURE DIRECTIONS

The art and technology of recording have changed drastically in the last decade or so as digital recording has virtually supplanted analog recording. There are several factors that have driven this. One is the rapid decrease in cost and increase in speed and capacity of digital recording media. Another is the advancement of analog to digital converter technology. A third factor is the use of computers within the recorder to control the generation of calibration signals and analyze the recorded results. Future progress is expected to occur along the same lines and is probably more controlled by demand than by technology.

One would expect to see an increase in the bandwidth and sampling rate of ADCs. As the bandwidth moves from its present limit of 1 to 2 GHz to the 5 to 10 GHz range, the more limited technologies of traveling wave CRTs and streak cameras will probably disappear. One could expect over the next decade to see ADC based recorders reach the same 20 to 30 GHz bandwidth range that sampling oscilloscopes now have.

Increased precision is another area for which there is steady progress. ADCs can be roughly grouped into the 8, 12, 16, and 20 bit classes. Each class has an upper sampling rate limit that can be economically accomplished. These upper frequency limits steadily increase. The increased sampling frequency at a fixed precision of ADCs translates directly into the same improvements for recorders. However, recorders improve at an even faster rate. As the cost, size, and power consumption of ADCs go down it becomes practical to interleave more of them into a single recorder channel—increasing the sampling rate more than that of the ADCs.

An area ripe for improvement at lower sampling rates (10 MSa/s and less) is increased capacity through real time data compression. This involves analyzing and compressing the data between the time it is digitized and the time it is written to disk, so that the disk can hold much longer records. Very specialized computing hardware and algorithms are already in use to accomplish this for audio and video recording. This could be extended to more general situations.

Many users would like to be able to predict a recorder's performance in hypothetical applications based on various recorder specifications. Some work has been done in this area (5) but this is far from straightforward. More research into modeling recorders with sufficient precision to predict performance in a wide range of practical applications is desirable.

A related situation is the current lack of precisely known test signals other than sine waves. One can't test a recorder's performance in a hypothetical situation by generating the hypothetical test signal, recording it, and examining the errors, because the art of accurately generating arbitrary test signals is lagging behind the art of recording them.

BIBLIOGRAPHY

1. A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing,* Englewood Cliffs, NJ: Prentice-Hall, 1975.
2. S. S. Stearns and D. R. Hush (eds.), *Signal Processing Algorithms,* Englewood Cliffs, NJ: Prentice-Hall, 1990.
3. S. S. Stearns and R. A. David (eds.), *Digital Signal Analysis,* 2nd ed., Englewood Cliffs, NJ: Prentice-Hall, 1988.
4. T. E. McEwan and J. D. Kilkenny, A 32 gigasample-per-second single-shot transient digitizer, *R&D Mag.,* **69**: 108–116, 1993.
5. *IEEE Standard for Digitizing Waveform Recorders (IEEE Std 1057-1994),* New York: Institute Electrical Electronics Engineers, 1994.
6. *User Guide for the Waveform Recorder Standard 1057,* New York: Institute Electrical Electronics Engineers, in press.
7. *IEEE Standard for Analog to Digital Converters (IEEE Std 1241),* New York: Institute Electrical Electronics Engineers, in press.
8. J. J. Blair, A method for characterizing waveform recorder errors using the power spectral distribution, *IEEE Trans. Instrum. Meas.,* **41**: 604–610, 1992.
9. *The new IEEE Standard Dictionary of Electrical and Electronics Terms,* 5th ed., New York: Institute Electrical Electronics Engineers, 1992.
10. F. H. Irons and D. M. Hummels, The modulo time plot—a useful data acquisition diagnostic tool. *IEEE Trans. Instrum. Meas.,* **45**: 734–738, 1996.
11. F. Rumsey and T. McCormick (eds.), *Sound and Recording: an Introduction,* Oxford: Focal Press, 1992.
12. J. J. Blair, Error estimates for frequency responses calculated from time domain measurements, *IEEE Trans. Instrum. Meas.,* in press.
13. R. W. Landee et al. (eds.), *Electronics Designers' Handbook,* 2nd ed., New York: McGraw-Hill, 1977.
14. W. H. Press et al. (eds.), *Numerical Recipes in C,* 2nd ed., Cambridge, UK: Cambridge Univ. Press, 1992.
15. J. C. Whitaker (ed.), *The Electronics Handbook,* Boca Raton, FL: CRC Press, 1996.

JEROME J. BLAIR
Bechtel Nevada

RECORDING, MAGNETIC. See MAGNETIC STORAGE MEDIA; MAGNETIC TAPE RECORDING.

RECORDING, VIDEO. See VIDEO RECORDING.

RECTIFIER DIODES. See DIODES FOR POWER ELECTRONICS.

RECTIFIERS. See AC-DC POWER CONVERTERS; POWER FACTOR CORRECTION.

RECTIFIERS, ACTIVE. See POWER FACTOR CORRECTION.

RECTIFIERS, SWITCHING. See POWER FACTOR CORRECTION.