17

Distortion Measurement

Michael F. Toner
Northern Telecom Ltd.

Gordon W. Roberts
McGill University

17.1 Mathematical Background 17-2
17.2 Intercept Points (IP) 17-7
17.3 Measurement of the THD 17-8
Classical Method • Spectrum Analyzer Method • DSP Method
17.4 Conclusions 17-10

A sine-wave signal will have only a single-frequency component in its spectrum; that is, the frequency of the tone. However, if the sine wave is transmitted through a system (such as an amplifier) having some nonlinearity, then the signal emerging from the output of the system will no longer be a pure sine wave. That is, the output signal will be a distorted representation of the input signal. Since only a pure sine wave can have a single component in its frequency spectrum, this situation implies that the output must have other frequencies in its spectral composition. In the case of *harmonic distortion,* the frequency spectrum of the distorted signal will consist of the fundamental (which is the same frequency as the input sine wave) plus harmonic frequency components that are at integer multiples of the fundamental frequency. Taken together, these will form a Fourier representation of the distorted output signal. This phenomenon can be described mathematically. Refer to Figure 17.1, which depicts a sine-wave input signal $x(t)$ at frequency f_1 applied to the input of a system $A(x)$, which has an output $y(t)$. Assume that system $A(x)$ has some nonlinearity. If the nonlinearity is severe enough, then the output $y(t)$ might have excessive harmonic distortion such that its shape no longer resembles the input sine wave. Consider the example where the system $A(x)$ is an audio amplifier and $x(t)$ is a voice signal. Severe distortion can result in a situation where the output signal $y(t)$ does not represent intelligible speech. The *total harmonic distortion* (THD) is a figure of merit that is indicative of the quality with which the system $A(x)$ can reproduce an input signal $x(t)$. The output signal $y(t)$ can be expressed as:

$$y(t) = a_0 + \sum_{k=1}^{N} a_k \cos(2\pi k f_1 t + \theta_k) \tag{17.1}$$

where the a_k, $k = 0, 1, \ldots, N$ are the magnitudes of the Fourier coefficients, and θ_k, $k = 0, 1, \ldots, N$ are the corresponding phases. The THD is defined as the percentage ratio of the rms voltage of all harmonics components above the fundamental frequency to the rms voltage of the fundamental. Mathematically, the definition is written:

$$\mathrm{THD} = \frac{\sqrt{\sum_{k=2}^{N} a_k^2}}{a_1} \times 100\% \tag{17.2}$$

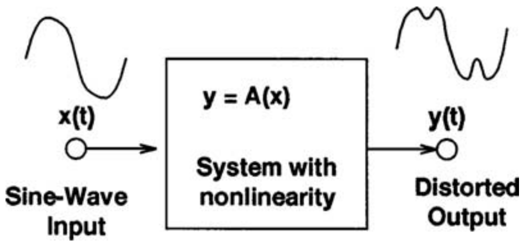

FIGURE 17.1 Any system with a nonlinearity gives rise to distortion.

If the system has good linearity (which implies low distortion), then the THD will be a smaller number than that for a system having poorer linearity (higher distortion). To provide the reader with some feeling for the order of magnitude of a realistic THD, a reasonable audio amplifier for an intercom system might have a THD of about 2% or less, while a high-quality sound system might have a THD of 0.01% or less. For the THD to be meaningful, the bandwidth of the system must be such that the fundamental and the harmonics will lie within the passband. Therefore, the THD is usually used in relation to low-pass systems, or bandpass systems with a wide bandwidth. For example, an audio amplifier might have a 20 Hz to 20 kHz bandwidth, which means that a 1 kHz input sine wave could give rise to distortion up to the 20^{th} harmonic (i.e., 20 kHz), which can lie within the passband of the amplifier. On the other hand, a sine wave applied to the input of a narrow-band system such as a radio frequency amplifier will give rise to harmonic frequencies that are outside the bandwidth of the amplifier. These kinds of narrow-band systems are best measured using *intermodulation distortion,* which is treated elsewhere in this Handbook. For the rest of the discussion at hand, consider the example system illustrated in Figure 17.1 which shows an amplifier system $A(x)$ that is intended to be linear but has some undesired nonlinearities. Obviously, if a linear amplifier is the design objective, then the THD should be minimized.

17.1 Mathematical Background

Let $y = A(x)$ represent the input-output transfer characteristic of the system $A(x)$ in Figure 17.1 containing the nonlinearity. Expanding into a power series yields

$$A(x) = \sum_{k=0}^{\infty} c_k x^k = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + \ldots \tag{17.3}$$

Let the input to the system be $x = \cos(2\pi f_0 t)$. Then the output will be

$$y = A(x) = c_0 + c_1 A_0 \cos(2\pi f_0 t) + c_2 A_0^2 \cos^2(2\pi f_0 t) + c_3 A_0^3 \cos^3(2\pi f_0 t) + \ldots \tag{17.4}$$

This can be simplified using the trigonometric relationships:

$$\cos^2(\theta) = \frac{1}{2} - \frac{1}{2}\cos(2\theta) \tag{17.5}$$

$$\cos^3(\theta) = \frac{3}{4}\cos(\theta) + \frac{1}{4}\cos(3\theta) \tag{17.6}$$

$$\cos^4(\theta) = \frac{1}{8} - \frac{1}{2}\cos(2\theta) + \frac{1}{8}\cos(4\theta) \tag{17.7}$$

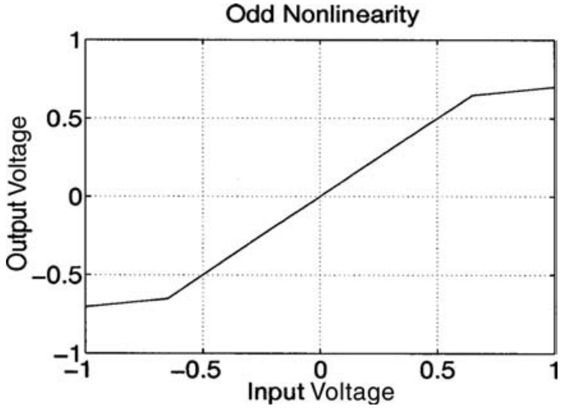

FIGURE 17.2 An odd nonlinearity with $f(-x) = -f(x)$.

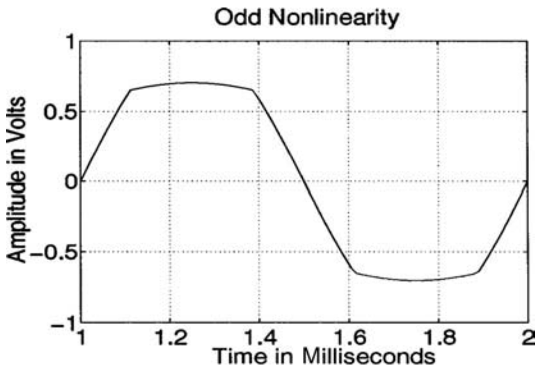

FIGURE 17.3 Distortion due to symmetrical two-sided clipping.

$$\cos^5(\theta) = \frac{5}{8}\cos(\theta) + \frac{5}{16}\cos(3\theta) + \frac{1}{16}\cos(5\theta) \tag{17.8}$$

and so on. Performing the appropriate substitutions and collecting terms results in an expression for the distorted signal $y(t)$ that is of the form shown in Equation 17.1. The THD can then be computed from Equation 17.2.

Closer inspection of Equations 17.6 and 17.8 reveal that a cosine wave raised to an odd power gives rise to only the fundamental and odd harmonics, with the highest harmonic corresponding to the highest power. A similar phenomenon is observed for a cosine raised to even powers; however, the result is only a dc component and even harmonics without any fundamental component. In fact, any nonlinear system that possesses an odd input-output transfer characteristic $A(x)$ (i.e., the function $A(x)$ is such that $-A(x) = A(-x)$) will give rise to odd harmonics only. Consider Figure 17.2, which illustrates an example of two-sided symmetrical clipping. It is an odd function. The application of a sine wave to its input will result in a waveform similar to that shown in Figure 17.3, which has only odd harmonics as shown in Figure 17.4. The majority of physical systems are neither odd nor even. (An even function is one that has the property $A(x) = A(-x)$; for example, a full-wave rectifier.) Consider the enhancement NMOS transistor illustrated in Figure 17.5, which has the square-law characteristic shown. Assume that the voltage V_{GS} consists of a dc bias plus a sine wave such that V_{GS} is always more positive than V_T (the threshold voltage). Then the current flowing in the drain of this NMOS transistor could have the appearance shown in Figure 17.6. It is observed that the drain current is distorted, since the positive-going side has a greater swing than the negative-going side. The equation for the drain current can be

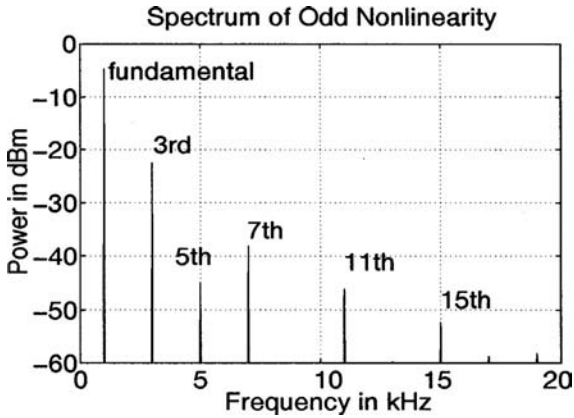

FIGURE 17.4 Frequency spectrum of signal distorted by symmetrical two-sided clipping.

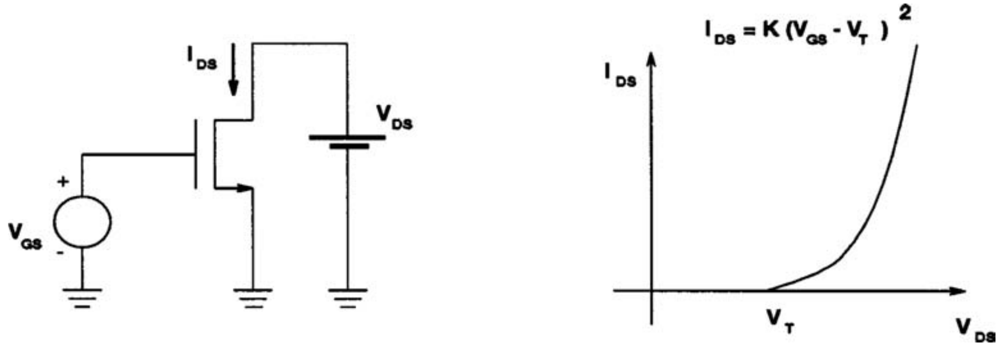

FIGURE 17.5 NMOS enhancement transistor is actually a nonlinear device. It is neither odd nor even in the strict sense.

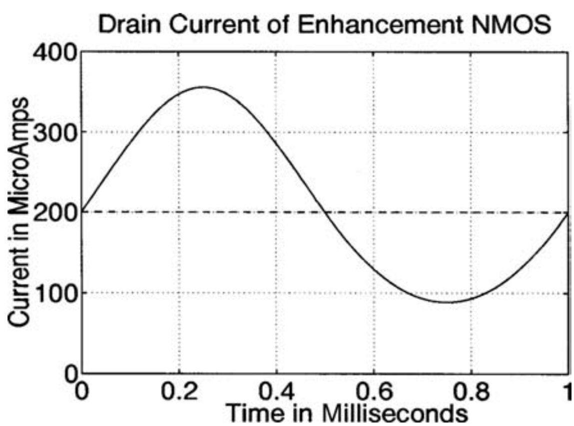

FIGURE 17.6 Showing how the drain current of the enhancement NMOS device is distorted.

derived mathematically as follows. A MOS transistor operating in its saturation region can be approximated as a square-law device:

$$I_{DS} = \frac{\mu C_{ox}}{2} \frac{W}{L} \left(V_{GS} - V_T\right)^2 \tag{17.9}$$

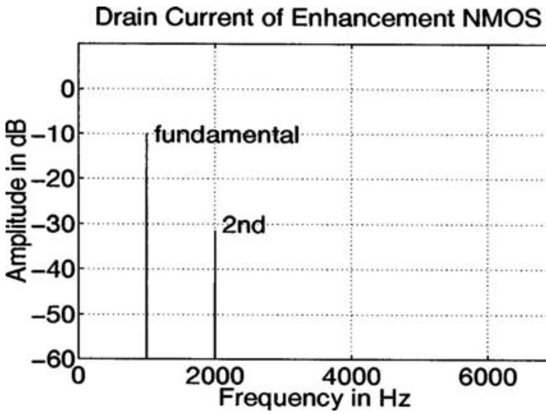

FIGURE 17.7 The drain current contains a dc bias, the fundamental, and the second harmonic only, for an ideal device.

If the gate of the *n*-channel enhancement MOSFET is driven by a voltage source consisting of a sine-wave generator in series with a dc bias, i.e.:

$$V_{GS} = V_B + A_0 \sin(2\pi f_0 t) \tag{17.10}$$

then the current in the drain can be written as:

$$I_{DS} = \frac{\mu C_{ox}}{2} \frac{W}{L} \left\{ \left[V_B + A_0 \sin(2\pi f_0 t) \right] - V_T \right\}^2 \tag{17.11}$$

Expanding and using the trigonometric relationship:

$$\sin^2(\theta) = \frac{1}{2} - \frac{1}{2} \sin\left(2\theta + \frac{\pi}{2}\right) \tag{17.12}$$

Equation 17.11 can be rewritten as:

$$I_{DS} = \frac{\mu C_{ox}}{2} \frac{W}{L} \left[(V_B - V_T)^2 + \frac{A_0^2}{2} + 2(V_B - V_T) A_0 \sin(2\pi f_0 t) - \frac{A_0^2}{2} \sin\left(4\pi f_0 t\right) + \frac{\pi}{2} \right] \tag{17.13}$$

which clearly shows the dc bias, the fundamental, and the second harmonic that are visible in the spectrum of the drain current I_{DS} in Figure 17.7. There is one odd harmonic (i.e., the fundamental) and two even harmonics (strictly counting the dc component and the second harmonic). This particular transfer characteristic is neither odd nor even. Finally, for an ideal square-law characteristic, the second harmonic is the highest frequency component generated in response to a sine-wave input. Another example of a transfer characteristic that is neither even nor odd is single-sided clipping as shown in Figure 17.8, which gives rise to the distortion of Figure 17.9. One last example of an odd input-output transfer characteristic is symmetrical cross-over distortion as depicted in Figure 17.10. The distorted output in response to a 1 kHz sine-wave input is shown in Figure 17.11. The spectrum of the output signal is shown in Figure 17.12. Note that only odd harmonics have been generated.

To round out the discussion, consider a mathematical example wherein the harmonics are derived algebraically. Consider an input-output transfer function $f(x) = c_1\, x + c_3\, x^3 + c_5 x^5$ that has only odd powers of *x*. If the input is a cosine $x = A_0 \cos(2\pi f_0 t)$, then the output will be of the form:

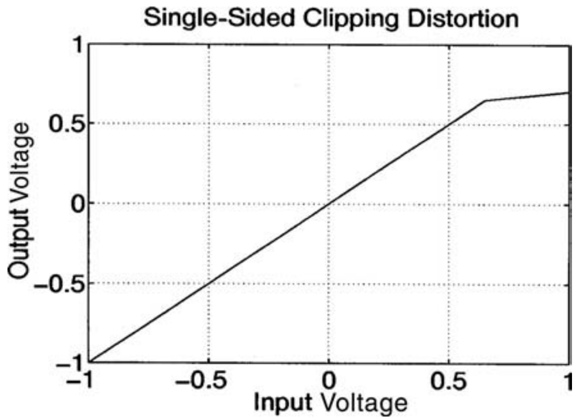

FIGURE 17.8 Single-sided clipping is neither even nor odd.

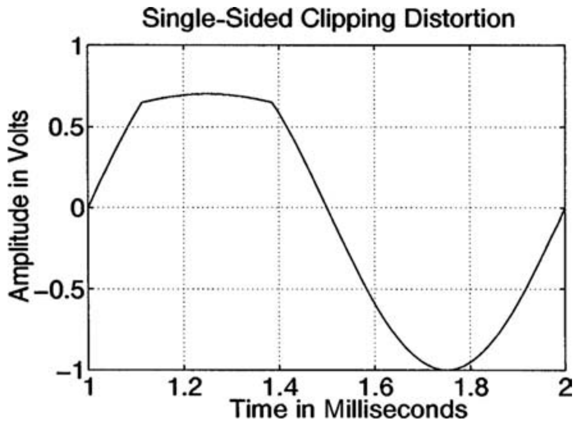

FIGURE 17.9 Distortion due to single-sided clipping.

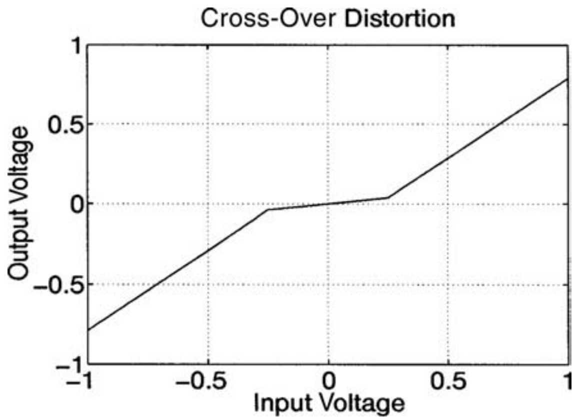

FIGURE 17.10 Symmetrical cross-over distortion is odd.

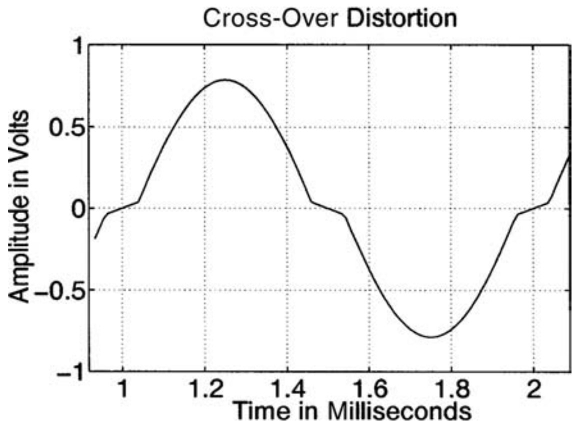

FIGURE 17.11 An example of cross-over distortion.

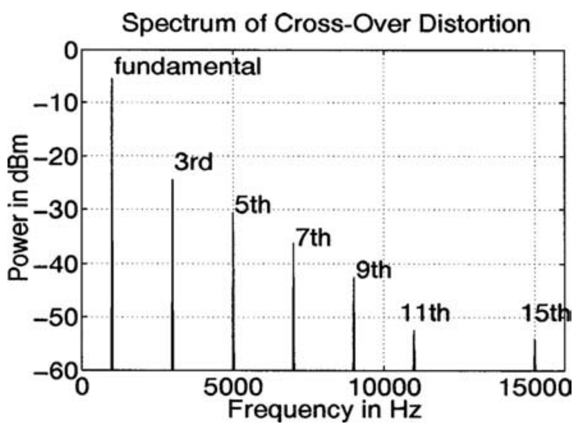

FIGURE 17.12 Symmetrical cross-over distortion gives rise to odd harmonics.

$$y(t) = f(x) = c_1 A_0 \cos(2\pi f_0 t) + c_3 A_0^3 \cos^3(2\pi f_0 t) + c_5 A_0^5 \cos^5(2\pi f_0 t) \tag{17.14}$$

This can be simplified using the trigonometric relationships given in Equations 17.5 through 17.8 with the following result:

$$y(t) = \left(c_1 A_0 + \frac{3c_3 A_0^3}{4} + \frac{5c_5 A_0^5}{8}\right)\cos(2\pi f_0 t) + \left(\frac{c_3 A_0^3}{4} + \frac{5c_5 A_0^5}{16}\right)\cos(2\pi 3 f_0 t) + c_5 A_0^5 \cos^5(2\pi 5 f_0 t) \tag{17.15}$$

Clearly, only the fundamental plus the third and fifth harmonics are present. Should the exercise be repeated for an input-output transfer function consisting of only even powers of x, then only a dc offset plus even harmonics (not the fundamental) would be present in the output.

17.2 Intercept Points (IP)

It is often desirable to visualize how the various harmonics increase or decrease as the amplitude of the input sine wave $x(t)$ is changed. Consider the example of a signal applied to a nonlinear system $A(x)$ having single-sided clipping distortion as shown in Figure 17.8. The clipping becomes more severe as the

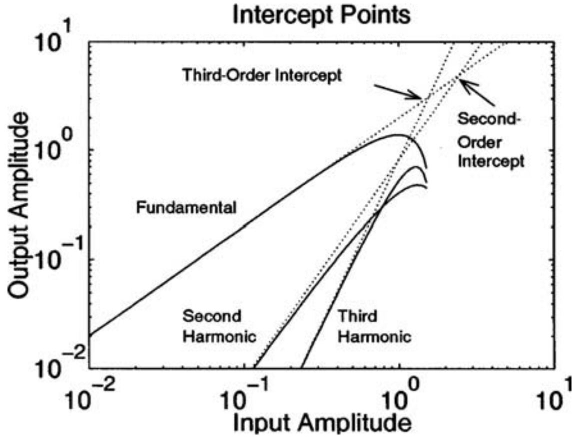

FIGURE 17.13 An example showing the second- and third-order intercept points for a hypothetical system. Both axes are plotted on a logarithmic scale.

amplitude of the input signal $x(t)$ increases in amplitude, so the distortion of the output signal $y(t)$ becomes worse. The *intercept point* (IP) is used to provide a figure of merit to quantify this phenomenon. Consider Figure 17.13, which shows an example of the power levels of the first three harmonics of the distorted output $y(t)$ of a hypothetical system $A(x)$ in response to a sine-wave input $x(t)$. It is convenient to plot both axes on a log scale. It can be seen that the power in the harmonics increases more quickly than the power in the fundamental. This is consistent with the observation of how clipping becomes worse as the amplitude increases. It is also consistent with the observation that, in the equations above, the higher harmonics will rapidly become more prominent because they are proportional to higher exponential powers of the input signal amplitude. The intercept point for a particular harmonic is the power level where the extrapolated line for that harmonic intersects with the extrapolated line for the fundamental. The second-order intercept is often abbreviated IP2, the third-order intercept abbreviated IP3, etc.

17.3 Measurement of the THD

Classical Method

The traditional method of measuring THD is shown in Figure 17.14. A sine-wave test stimulus $x(t)$ is applied to the input of the system $A(x)$ under test. The system output $y(t)$ is fed through a bandpass filter tuned to the frequency of the input stimulus to extract the signal. Its power p_1 can be measured with a power meter. The bandpass filter is then tuned to each of the desired harmonics in turn and the measurement is repeated to determine the required p_i. The THD is then calculated from:

$$\mathrm{THD} = \sqrt{\frac{\sum_{k=2}^{N} p_k}{p_1}} \times 100\% \tag{17.16}$$

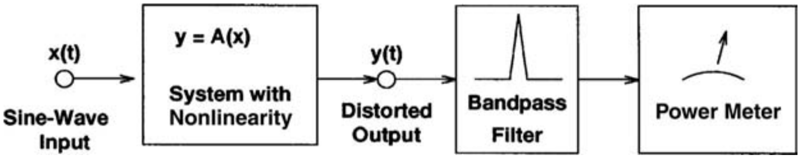

FIGURE 17.14 Illustrating the classical method of measuring THD.

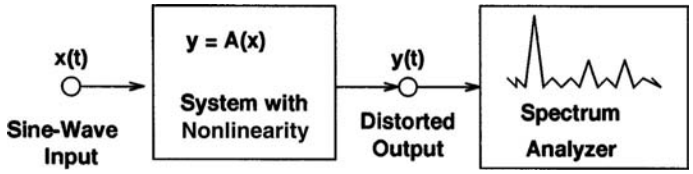

FIGURE 17.15 Illustrating measurement of THD using a spectrum analyzer.

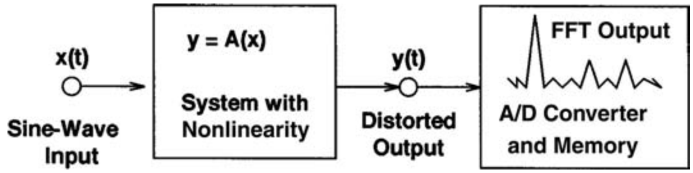

FIGURE 17.16 Illustrating the measurement of THD using FFT.

In the case of an ordinary audio amplifier, nearly all of the power in the distorted output signal is contained in the first 10 or 11 harmonics. However, in more specialized applications, a much larger number of harmonics might need to be considered.

Spectrum Analyzer Method

THD measurements are often made with a spectrum analyzer using the setup shown in Figure 17.15. The readings for the power levels of each of the desired harmonic components in the frequency spectrum of the distorted signal $y(t)$ are collected from the spectrum analyzer, usually in units of dB. They are converted to linear units by means of the relationship:

$$a_i = 10^{r_i/20} \tag{17.17}$$

where r_i is the reading for the i^{th} component in dB. The THD is then computed from Equation 17.2. The spectrum analyzer method can be considered as an extension of the classical method described above, except that the spectrum analyzer itself is replacing both the bandpass filter and the power meter.

DSP Method

Digital signal processing (DSP) techniques have recently become popular for use in THD measurement. In this method, the distorted output $y(t)$ is digitized by a precision A/D converter and the samples are stored in the computer's memory as shown in Figure 17.16. One assumes that the samples have been collected with a uniform sample period T_s and that appropriate precautions have been taken with regard to the Nyquist criterion and aliasing. Let $y(n)$ refer to the n^{th} stored sample. A fast Fourier transform (FFT) is executed on the stored data using the relationship:

$$Y(k) = \sum_{n=0}^{N-1} y(n) e^{-j(2\pi/N)kn} \tag{17.18}$$

where N is the number of samples that have been captured. The frequency of the input test stimulus is chosen such that the sampling is coherent. Coherency in this context means that if N samples have been captured, then the input test stimulus is made to be a harmonic of the primitive frequency f_P, which is defined as:

$$f_p = \frac{f_s}{N} = \frac{1}{T_s N} \tag{17.19}$$

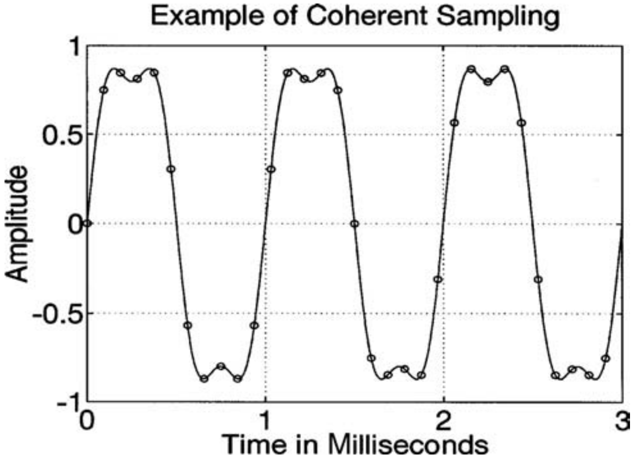

FIGURE 17.17 With coherent sampling, each sample occurs on a unique point of the signal.

One can view the primitive frequency f_P as the frequency of a sinusoidal signal whose period is exactly equal to the time interval formed by the N-points. Thus, the frequency of the test stimulus can be written as:

$$f_0 = M \times f_p = M \times \frac{f_s}{N} = \frac{M}{N} \times f_s \tag{17.20}$$

where M and N are integers. To maximize the information content collected by a set of N-points, M and N are selected so that they have no common factors, i.e., relatively prime. This ensures that every sample is taken at a different point on the periodic waveform. An example is provided in Figure 17.17, where $M = 3$ and $N = 32$. The FFT is executed on the distorted signal as per Equation 17.18, and then the THD is computed from:

$$\mathrm{THD} = \frac{\sqrt{\sum_{k=2}^{N} \left|Y(k \times M)\right|^2}}{\left|Y(M)\right|} \times 100\% \tag{17.21}$$

17.4 Conclusions

The *total harmonic distortion* (THD) is a figure of merit for the quality of the transmission of a signal through a system having some nonlinearity. Its causes and some methods of measuring it have been discussed. Some simple mathematical examples have been presented. However, in real-world systems, it is generally quite difficult to extract all of the parameters c_k in the transfer characteristic. The examples were intended merely to assist the reader's understanding of the relationship between even and odd functions and the harmonics that arise in response to them.

Defining Terms

Total harmonic distortion (THD): A numerical figure of merit of the quality of transmission of a signal, defined as the ratio of the power in all the harmonics to the power in the fundamental.
Fundamental: The lowest frequency component of a signal other than zero frequency.
Harmonic: Any frequency component of a signal that is an integer multiple of the fundamental frequency.
Distortion: The effect of corrupting a signal with undesired frequency components.
Nonlinearity: The deviation from the ideal of the transfer function, resulting in such effects as clipping or saturation of the signal.

Further Information

D.O. Pederson and K. Mayaram, *Analog Integrated Circuits for Communications,* New York: Kluwer Academic Press, 1991.

M. Mahoney, *DSP-Based Testing of Analog and Mixed-Signal Circuits,* Los Almos, CA: IEEE Computer Society Press, 1987.