Response of Systems to a Sinusoidal Input, as Viewed in the Frequency Domain

It is useful in electrical engineering to be able to anticipate how a system will react to a sinusoid. If a sinusoid is applied to the input of a system, what appears on the system's output? This question is addressed here for linear time-invariant (LTI) systems, some example nonlinear systems, and an example linear time-varying system. The interest here is primarily in anticipating the spectrum of the output signal.

Sinusoidal Signals

Why are electrical engineers so interested in sinusoids? There are at least four reasons. First, the voltage available at a wall outlet is a sinusoid. Second, wireless transmission works, in most cases, by sending a sinusoidal carrier through the air. Third, fiber-optic communication works by sending a sinusoidal light wave through a fiber-optic cable. Fourth, in the development stage of an electronics project, it is necessary to test the system; this requires a test signal, and the sinusoid is commonly used for this purpose. There could be objections to the above statements. Two potential objections are addressed below.

In communications, whether accomplished wirelessly or with an optical-fiber cable, the transmitted signal is never a pure sinusoid, because a pure sinusoid does not carry a message. In practice, the sinusoidal carrier is always modulated by a message, and the modulated carrier is therefore never a pure sinusoid. This observation is correct. However, when studying the passage of a modulated carrier through a system, it is often easier to begin the analysis by examining how the system responds to a pure sinusoid (an unmodulated carrier) that has the frequency of the actual carrier. This is easier to visualize and to calculate. Once the response of the system to a sinusoid is understood, it is often surprisingly easy to estimate the response of that same system to a modulated carrier.

It is fair to say that different kinds of test signals are used to test electronic systems. For example, it is common to test the response of a control system to a unit-step input. However, the most expedient method of characterizing a filter, which is a type of LTI system, is to consider a sinusoidal test signal. That is because a sinusoid with a given frequency on the input of an LTI system (such as a filter) produces on the output a sinusoid of the same frequency. It is then necessary to determine only the amplitude and phase of the output sinusoid. Filters are widely employed in control systems, in communications, in signal processing, and in power electronics. Therefore, measurement and instrumentation also involve sinusoidal signals.

Modulated Carrier

Two types of modulation are defined here: amplitude modulation (AM) and phase modulation (PM). These terms are defined here because the effect of different systems on modulated carriers are considered later.

Amplitude Modulation (AM)

An AM carrier is modeled like this:

$$[1 + m(t)] \cdot \sin(2\pi f_C t)$$

Here f_C is the carrier frequency in hertz, and $m(t)$ is the message (a dimensionless quantity), where $|m(t)| \leq 1$.

Phase Modulation (PM)

A PM carrier is modeled like this:

$$\sin(2\pi f_C t + \theta m(t))$$

Here θ is a constant in radians. f_C and $m(t)$ are the carrier frequency and message.

A phase-modulated carrier is an example of a constant-envelope carrier; this means that the amplitude of the sinusoid is constant. A frequency-modulated (FM) carrier is another example of a constant-envelope carrier. For some wireless applications, a constant-envelope carrier is preferred because such a carrier will pass through a nonlinear system without incurring distortion of the message. Here a PM carrier is used as an example of a constant-envelope carrier. Much of what is stated below about the effect of different systems on a PM carrier is equally applicable to the larger class of constant-envelope carriers (including FM carriers).

Linear Time-Invariant Systems

A sinusoid at the input of an LTI system, such as a filter, produces a sinusoid of the same frequency on the output. This statement applies to the steady-state response of an LTI system. Of course, if a sinusoid is suddenly applied to the input, there will be a non-sinusoidal transient component in the output.

The steady-state response of an LTI system to a sinusoid is characterized by a frequency response $H(f)$. This is a complex-valued function of frequency f. The magnitude $|H(f)|$ is the ratio of the output amplitude to the input amplitude. When $|H(f)| > 1$ for the given frequency f, there is amplification. Otherwise, the LTI system attenuates the sinusoid with frequency f. The angle $\angle H(f)$ of the frequency response is the phase change of a sinusoid passing through the LTI system.

$$\sqrt{2}\cos(2\pi f t) \longrightarrow \boxed{\text{Filter}} \longrightarrow \sqrt{2}|H(f)|\cos[2\pi f t + \arg H(f)]$$

A more general statement about the spectrum of an LTI system's output is possible. For this purpose, any periodic input signal is considered. A periodic signal possesses a Fourier series expansion and will therefore have content only at a discrete set of frequencies: the fundamental frequency and some harmonics. With a periodic signal at the input of an LTI system, the (steady-state) output will not have content at any frequency not present on the input. For example, a square-wave has spectral content only at the odd harmonics of its fundamental frequency (including the fundamental harmonic); therefore, the output will have spectral content only at odd harmonics (including the fundamental).

Nonlinear Systems

Every nonlinear system considered here is memoryless. Each of these nonlinear systems can be defined by a simple plot, the characteristic curve, of the output as a function of the input. Figure 1 shows the characteristic curves of the memoryless nonlinear systems considered here.

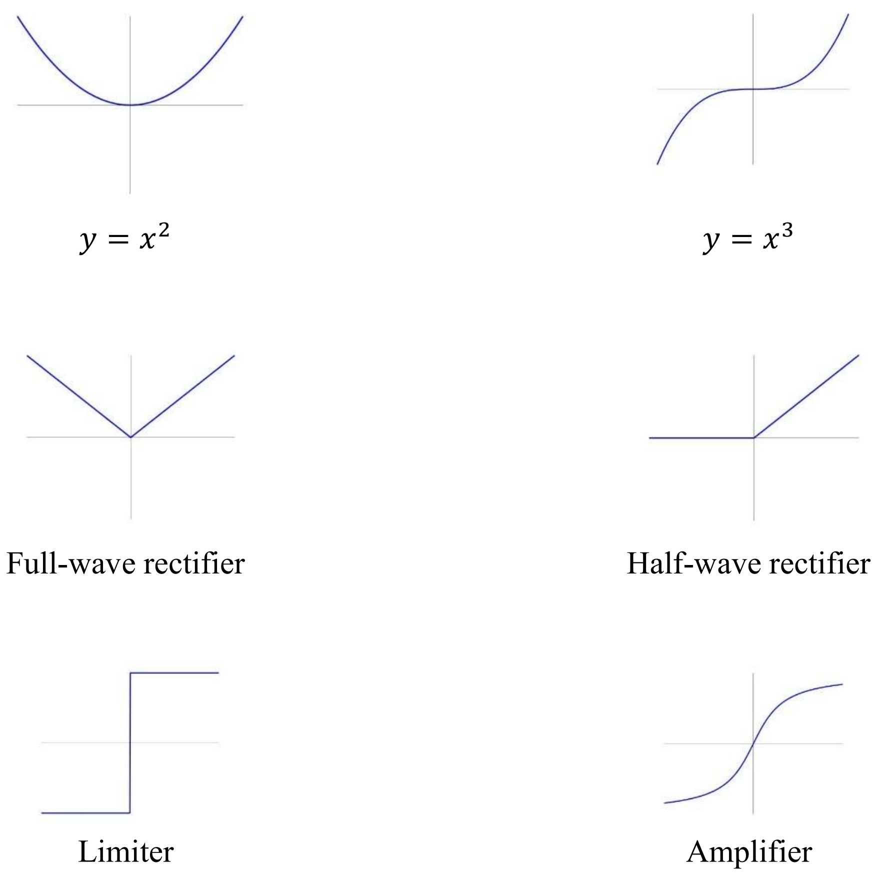


Figure 1. Characteristic curves of some memoryless, nonlinear systems

The first two nonlinear devices shown above, a square-law device ($y = x^2$) and a cube-law device ($y = x^3$), are treated mathematically below. This is done because the mathematics are straightforward in these cases, involving just algebra and trigonometry, and because these two devices illustrate the effect of having a memoryless, nonlinear system whose characteristic curve has either even or odd symmetry. All other nonlinear systems considered here are treated in an intuitive way, without the use of explicit mathematics.

Square-law device

A square-law device has even symmetry in its characteristic curve. The effect of a square-law device on a sinusoid can be understood from the following trigonometric identity.

$$2\sin^2(\alpha) = 1 - \cos(2\alpha)$$

For an unmodulated carrier, the input and output are:

$$\sqrt{2}\sin(2\pi f_C t) \longrightarrow \boxed{y = x^2} \longrightarrow 1 - \cos(2\pi \cdot 2f_C\, t)$$

Only two terms appear on the output: DC and the second harmonic. *In general, a sinusoid at the input of a nonlinear system with an even-symmetric characteristic curve produces an output with only even harmonics.* In the case of a square-law device, two even harmonics appear on the output. (DC may be regarded as the zeroth harmonic, and zero is regarded as an even integer.)

What happens when an AM carrier is applied to a square-law device? The above analysis for the unmodulated carrier can be modified as illustrated below.

$$\sqrt{2}\,[1 + m(t)]\sin(2\pi f_C t) \longrightarrow \boxed{y = x^2} \longrightarrow [1 + 2m(t) + m^2(t)] \cdot [1 - \cos(2\pi \cdot 2f_C\, t)\,]$$

This is a bad result. Added to the message signal is the square of the message signal. It will be difficult, perhaps impossible, to separate $m(t)$ from $m^2(t)$ in the signal processing. So the message has been forever muddled. *In general, an amplitude-modulated carrier should not be passed through a nonlinear device*. By the way, the constant 1 is not a problem. In most applications involving an analog message, that message does not have a significant DC component. So the signal processing will be designed to block DC.

What happens when a modulated carrier with a constant envelope is applied to a square-law device? The analysis for the unmodulated carrier can be modified for the case of PM as illustrated below.

$$\sqrt{2}\sin(2\pi f_C t + \theta m(t)) \longrightarrow \boxed{y = x^2} \longrightarrow 1 - \cos(2\pi \cdot 2f_C\, t + 2\theta m(t))$$

The second harmonic at the square-law output incorporates the phase modulation, but with a modulation index (a measure of the strength of the modulation) that is twice as large. This is quite acceptable, since the message can be extracted from the modulated carrier in the usual way.

Cube-law device

A cube-law device has odd symmetry in its characteristic curve. The effect of a cube-law device on a sinusoid can be understood from the following trigonometric identity.

$$4\sin^3(\alpha) = 3\sin(\alpha) - \sin(3\alpha)$$

For an unmodulated carrier, the input and output are:

$$\sqrt[3]{4}\sin(2\pi f_C t) \longrightarrow \boxed{y = x^3} \longrightarrow 3\sin(2\pi f_C t) - \sin(2\pi \cdot 3f_C\, t)$$

Only two terms appear on the output: the fundamental and third harmonics. *In general, a sinusoid at the input of a nonlinear system with an odd-symmetric characteristic curve produces an output with only odd harmonics.*

If an AM carrier is applied to the input of a cube-law device, the output will not be pretty (or useful).

The illustration below indicates the result when a PM carrier is applied to a cube-law device.

$$\sqrt[3]{4}\sin(2\pi f_C t + \theta m(t)) \longrightarrow \boxed{y = x^3} \longrightarrow 3\sin(2\pi f_C t + \theta m(t)) - \sin(2\pi \cdot 3f_C\, t + 3\theta m(t))$$

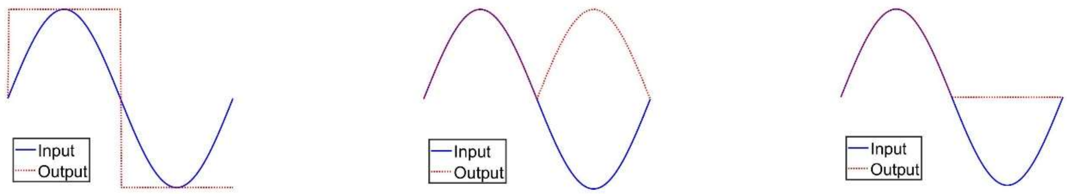
The fundamental with phase modulation is present on the output. Also, the third harmonic with phase modulation is present. For the third harmonic the modulation index is thrice as large.

Limiter

The sinusoidal input and corresponding output of a limiter are shown, in the time domain, in Figure 2. The limiter's characteristic curve has odd symmetry, as shown in Figure 1. Therefore, a sinusoid on the input of a limiter will produce, on the output, the odd harmonics. The magnitude of these odd harmonics may be found by calculating the Fourier series expansion of a square-wave. (The even harmonics of a square-wave have zero amplitude.)

Full-Wave Rectifier

The sinusoidal input and corresponding output of a full-wave rectifier are shown, in the time domain, in Figure 2. The full-wave rectifier's characteristic curve has even symmetry, as shown in Figure 1. Therefore, a sinusoid on the input of a full-wave rectifier will produce, on the output, the even harmonics. (This includes a DC component, since DC is the zeroth harmonic.)



Limiter output   Full-wave rectifier output   Half-wave rectifier output

Figure 2. The sinusoidal input and corresponding output for three nonlinear systems

If a carrier with PM is applied to a full-wave rectifier, the output will consist of even harmonics. Each even harmonic will have phase modulation with a modulation index that is appropriate to that harmonic number. For example, the second harmonic will have a modulation index that is twice that of the original modulated carrier. The fourth harmonic will have a modulation index that is four times that of the original modulated carrier, and so on.

<u>Half-Wave Rectifier</u>

The sinusoidal input and corresponding output of a half-wave rectifier are shown, in the time domain, in Figure 2. The half-wave rectifier's characteristic curve is shown in Figure 1. This characteristic curve is neither even symmetric nor odd symmetric.

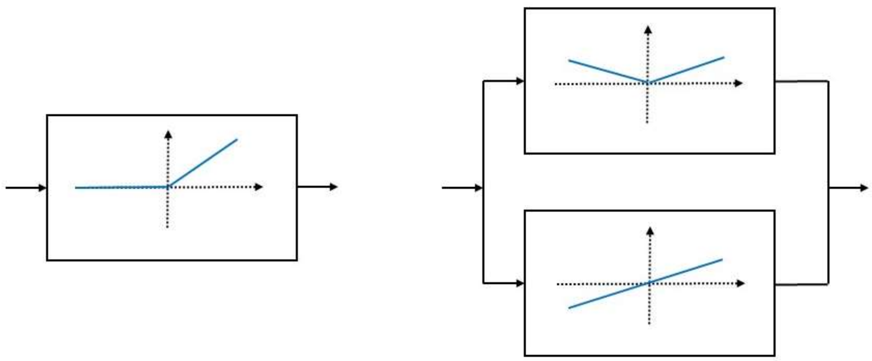
The nature of the half-wave rectifier's response to a sinusoid can be guessed using Figure 3. The system illustrated on the right side of Figure 3 is mathematically equivalent to a half-wave rectifier. The equivalent system is a parallel combination of two subsystems. The top subsystem is a full-wave rectifier, and its response to a sinusoidal input will be even harmonics (including a DC component). The bottom subsystem is LTI, so a sinusoid on the input will produce a sinusoid of the same frequency on the output. Putting this together, a half-wave rectifier should respond to a sinusoid with the fundamental harmonic plus even harmonics. Higher-order odd harmonics (beyond the fundamental) will be absent from the output.

<u>Amplifier</u>

The characteristic curve for an example amplifier is shown in Figure 1. This particular curve has odd symmetry, and that is typical for amplifiers.

When there is a relatively small signal on its input, an amplifier can be accurately modeled as an LTI system. The small signal only experiences the (approximately) linear portion of the curve. In such a case, the amplifier simply multiplies the input signal by a constant, and so only the fundamental appears on the output.

No practical amplifier can effect a linear characteristic for arbitrarily large signal levels on its input. When the input signal gets large enough, the nonlinearity of the amplifier's input/output characteristic will manifest itself. Under these circumstances, it becomes necessary to model the amplifier as a nonlinear device. Since an amplifier's characteristic curve is typically odd symmetric, the amplifier's response to a sinusoid will be odd harmonics. This will include the fundamental plus a third harmonic, a fifth harmonic, etc.



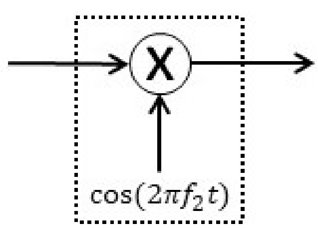
Ideal half-wave rectifier characteristic curve

Mathematically equivalent system that easily suggests the output spectrum

Figure 3. Viewing the half-wave rectifier as a parallel combination of a full-wave rectifier and a memoryless LTI system

Linear Time-Varying Systems

Only one example linear time-varying system is considered here. In the following illustration, this system lies within the dashed rectangle. In words, this system multiplies the input by a sinusoid.



When a sinusoid of frequency f_1 is applied to the input of this system, the output $y(t)$ is

$$y(t) = 2\cos(2\pi f_1 t) \cdot \cos(2\pi f_2 t) = \cos[2\pi(f_1 - f_2)t] + \cos[2\pi(f_1 + f_2)t]$$

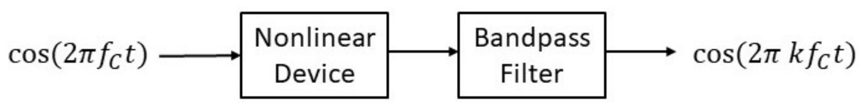
where f_2 is the frequency of the sinusoid that is internal to the system. The output equals the sum of two sinusoids: one having the difference frequency $f_1 - f_2$ and the other the sum frequency $f_1 + f_2$. *In general, when a sinusoid is applied to the input of a linear time-varying* *system, there will be signal content on the output at new frequencies that were not present on the input; however, these new frequencies will not, in general, be harmonics of the input frequency.*

Applications

Below are some applications within wireless communications for some of the systems discussed above.

<u>Frequency Synthesis</u>

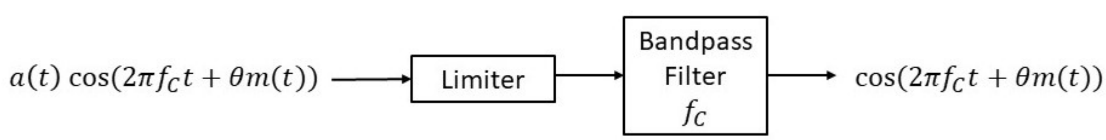
In wireless communications it is important that the frequency sources (oscillators) have good frequency stability. This is most commonly achieved with quartz crystal oscillators. However, these frequency-stable oscillators are relatively expensive and bulky. It is not acceptable to use a separate quartz crystal oscillator for every frequency source required by the wireless application. Instead, many different frequencies are generated from a single quartz crystal oscillator by using frequency synthesis. Sometimes the desired result is obtained by combining a nonlinear system with an LTI system. Here is one example of that:



A (frequency-stable) sinusoid is applied to the input of a nonlinear device, and harmonics are created. A bandpass filter is then used to pass the desired harmonic, blocking all others. The net result is a sinusoid having the desirable frequency stability of the original sinusoid but with a frequency that is k times as large, where k is the integer corresponding to the harmonic, the k-th harmonic, which was selected by the bandpass filter.

<u>Amplitude Leveling</u>

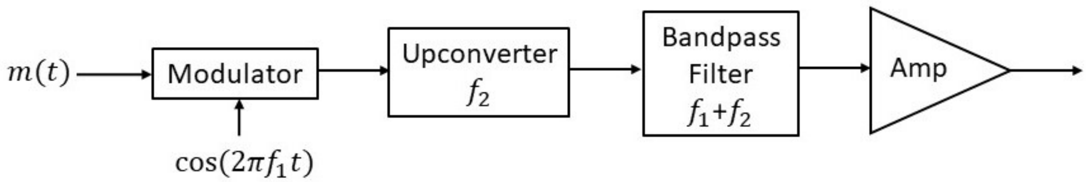
Sometimes the wireless channel introduces an unwanted amplitude variation on a carrier. The illustration below shows the modulated carrier within the receiver. The message $m(t)$ is phase-modulated onto the carrier. There is an unwanted amplitude variation $a(t)$, where $a(t) > 1$, that has been impressed on the received carrier by the wireless channel. If the amplitude variation is not removed within the receiver, that variation might adversely affect the demodulation of the carrier.



The configuration shown above wipes out the unwanted amplitude variation. This is accomplished by using a limiter which replaces the (phase-modulated) sinusoid with a (phase-modulated) square-wave, which has a level amplitude. The phase information in the original carrier is preserved in the zero-crossings of the square-wave, so the message in the phase modulation is not lost. Of course, the nonlinear limiter also creates harmonics. However, a carefully designed bandpass filter will pass the carrier frequency f_C and block the other harmonics.

Transmitter

A wireless transmitter may often be modeled like this:

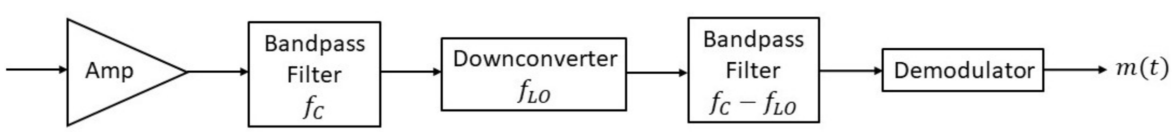


The message $m(t)$ modulates a sinusoid. This could be AM or a constant-envelope modulation, such as PM or FM. The frequency f_1 of the sinusoid in this stage of the electronics is much less than the carrier frequency that will be sent to the antenna. Modulators often can be made to work more precisely when they operate with a lower-frequency sinusoid. At the modulator output, it is necessary to shift the frequency f_1 to the desired carrier frequency. This is accomplished with a linear time-varying device, called an upconverter here, that produces the sum frequency $f_1 + f_2$, which is the desired carrier frequency. Typically, $f_2 \gg f_1$. A bandpass filter passes $f_1 + f_2$ and blocks the difference frequency (which comes out of the upconverter and which is unneeded). The output of the bandpass filter is the modulated carrier. (The upconverter does not affect the modulation, and the bandpass filter will have a wide enough passband that the modulation is not distorted in passing through this filter.) Finally, amplification is required to achieve an adequate transmitted power.

Since the signal is large in this final amplification of the transmitted signal, nonlinearity might come into play here. If the modulation is AM, it is important to avoid all nonlinearity in the signal path; so it becomes necessary to design the final amplifier to accommodate large signals that will not experience significant nonlinearity. If a constant-envelope modulation is used, the modulated carrier can tolerate some nonlinearity in the amplifier.

Receiver

A wireless receiver may often be modeled like this:



The weak signal arriving from the antenna is amplified. If AM is used, this amplifier should operate in the linear regime of its characteristic curve. This is usually not a problem, however, since the arriving signal is small.

Demodulation may be done more precisely if the frequency of the carrier is first reduced to a much smaller value. This is accomplished with a linear time-varying downconverter, which uses a local-oscillator frequency f_{LO} to multiply the received carrier, producing a difference-frequency $f_C - f_{LO}$. Typically, f_{LO} is selected to be close to f_C, so that $f_C - f_{LO} \ll f_C$. The sum-frequency term $f_C + f_{LO}$ is blocked with a bandpass filter. The signal sent to the demodulator is a modulated carrier, but with the carrier's frequency having been shifted to a much lower value.

Summary

It is helpful to think of filters as LTI systems with memory. This means that they cannot be characterized by a characteristic curve and ought instead to be characterized by a frequency response.

A memoryless nonlinear system can typically be understood in terms of its characteristic curve and in terms of the harmonics that it produces in response to a sinusoid.

A linear time-varying system, such as an upconverter or a downconverter, produces new frequency content not seen on the input. Unlike a nonlinear system, however, upconverters and downconverters do not generally produce harmonics.

In the above descriptions, modulators and demodulators have not been discussed in terms of system properties. Clearly, modulators and demodulators are not LTI. For example, the demodulator takes a modulated carrier on its input and produces on its output the baseband message; there is frequency content on the output that was not on the input. So are modulators and demodulators nonlinear or time-varying? It depends on the modulation type. Sometimes these systems are simultaneously nonlinear and time-varying. In any event, it is usually not helpful to think of modulators and demodulators in these terms.