

Peter O'Shea. "Phase Measurement."

Copyright 2000 CRC Press LLC. <<http://www.engnetbase.com>>.

Phase Measurement

- 41.1 Amplitude, Frequency, and Phase of a Sinusoidal Signal
- 41.2 The Phase of a Periodic Nonsinusoidal Signal
- 41.3 Phase Measurement Techniques
 - Direct Oscilloscope Methods • Lissajous Figures • Zero-Crossing Methods • The Three-Voltmeter Method • The Crossed-Coil Method • Synchrosopes and Phasing Lamps • Vector Voltmeters and Vector Impedance Methods • Phase Standard Instruments • The Fast Fourier Transform Method • Phase-Locked Loops
- 41.4 Phase-Sensitive Demodulation
 - The Phase-Locked Loop for Carrier Phase Tracking • Hidden Markov Model-Based Carrier Phase Tracker
- 41.5 Power Factor
- 41.6 Instrumentation and Components

Peter O'Shea

Royal Melbourne Institute of
Technology

The notion of “phase” is usually associated with *periodic* or repeating signals. With these signals, the waveshape perfectly repeats itself every time the *period* of repetition elapses. For periodic signals one can think of the phase at a given time as the fractional portion of the period that has been completed. This is commonly expressed in degrees or radians, with full cycle completion corresponding to 360° or 2π radians. Thus, when the cycle is just beginning, the phase is zero. When the cycle is half completed, the phase is half of 360° , or 180° (See [Figure 41.1](#)). It is important to note that if phase is defined as the portion of a cycle that is completed, the phase depends on where the beginning of the cycle is taken to be. There is no universal agreement on how to specify this beginning. For a sinusoidal signal, probably the two most common assumptions are that the start of the cycle is (1) the point at which the maximum value is achieved, and (2) the point at which the negative to positive zero-crossing occurs. Assumption (1) is common in many theoretical treatments of phase, and for that reason is adopted in this chapter. It should be noted, however, that assumption (2) has some benefits from a measurement perspective, because the zero-crossing position is easier to measure than the maximum.

The measurement of phase is important in almost all applications where sinusoids proliferate. Many means have therefore been devised for this measurement. One of the most obvious measurement techniques is to directly measure the fractional part of the period that has been completed on a cathode-ray oscilloscope (CRO). Another approach, which is particularly useful when a significant amount of noise is present, is to take the Fourier transform of the signal. According to Fourier theory, for a sinusoidal signal, the energy in the Fourier transform is concentrated at the frequency of the signal; the initial phase of the signal (i.e., the phase at time, $t = 0$) is the phase of the Fourier transform at the point of this energy concentration. The measurements of initial phase and frequency obtained from the Fourier transform can then be used to deduce the phase of the signal for any value of time.

Start of cycle
(phase is 0)

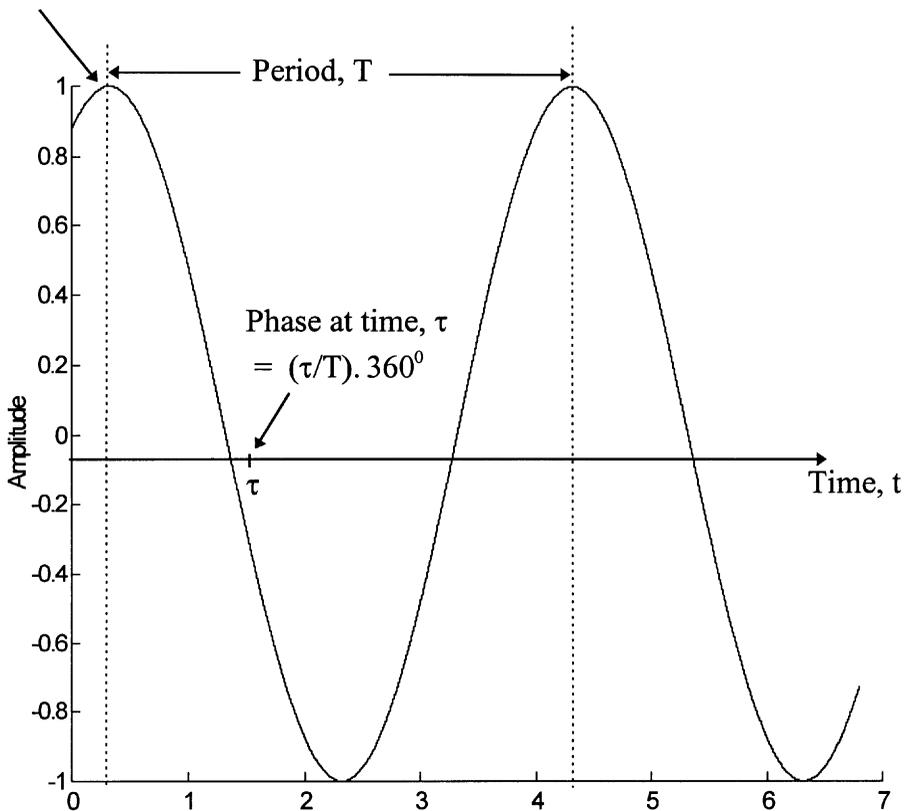


FIGURE 41.1 The phase of a periodic sinusoidal signal. The time scale is arbitrary.

Frequently what is needed in practice is a measurement of the *phase difference* between two signals of the same frequency; that is, it is necessary to measure the *relative phase* between two signals rather than the *absolute phase* of either one (see [Figure 41.2](#)). Often, in the measurement of the relative phase between two signals, both signals are derived from the same source. These signals might, for example, be the current and voltage of a power system; the relative phase, ϕ , between the current and voltage would then be useful for monitoring power usage, since the latter is proportional to the cosine of ϕ .

Several techniques are available for the measurement of “relative phase.” One crude method involves forming “Lissajous figures” on an oscilloscope. In this method, the first of the two signals of interest is fed into the vertical input of a CRO and the other is fed into the horizontal input. The result on the oscilloscope screen is an ellipse, the intercept and maximum height of which can be used to determine the relative phase. Other methods for determining relative phase include the crossed-coil meter (based on electromagnetic induction principles), the zero-crossing phase meter (based on switching circuitry for determining the fractional portion of the period completed), the three-voltmeter method (based on the use of three signals and trigonometric relationships), and digital methods (based on analog-to-digital conversion and digital processing).

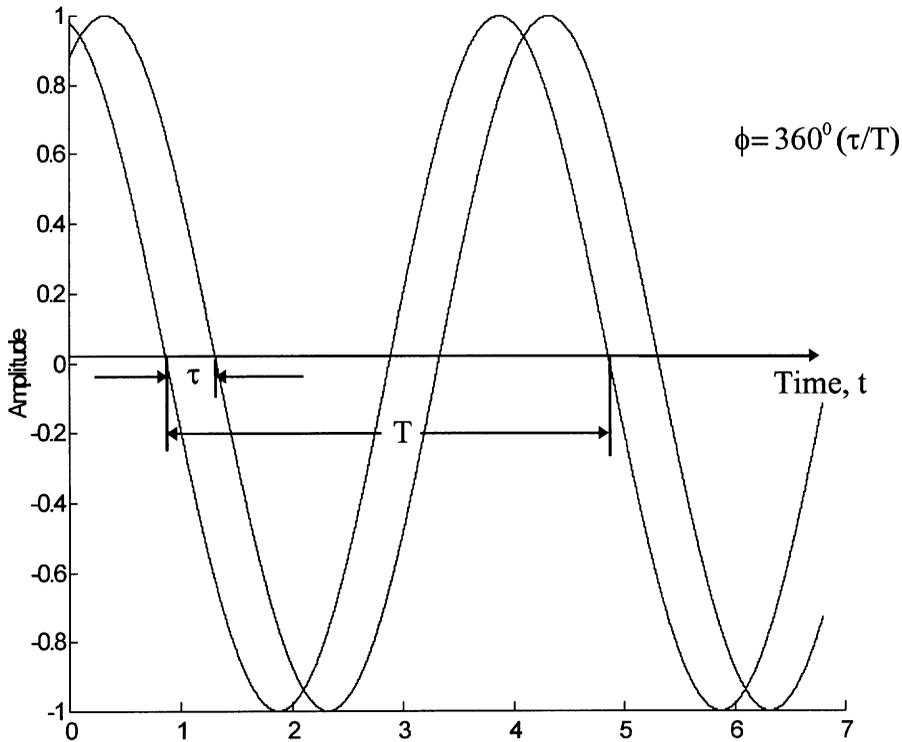


FIGURE 41.2 Two signals with a relative phase difference of ϕ between them. The time scale is arbitrary.

41.1 Amplitude, Frequency, and Phase of a Sinusoidal Signal

An arbitrary sinusoidal signal can be written in the form:

$$s(t) = A \cos(2\pi ft + \phi_0) = A \cos(\omega t + \phi_0) \quad (41.1)$$

where A = Peak amplitude

f = Frequency

ω = Angular frequency

ϕ_0 = Phase at time $t = 0$

This signal can be thought of as being the real part of a complex phasor that has amplitude, A , and which rotates at a constant angular velocity $\omega = 2\pi f$ in the complex plane (see Figure 41.3).

Mathematically, then, $s(t)$ can be written as:

$$s(t) = \Re\{Ae^{j2\pi ft + \phi_0}\} = \Re\{z(t)\} \quad (41.2)$$

where $z(t)$ is the complex phasor associated with $s(t)$, and $\Re\{\cdot\}$ denotes the real part. The “phase” of a signal at any point in time corresponds to the angle that the rotating phasor makes with the real axis. The initial phase (i.e., the phase at time $t = 0$) is ϕ_0 . The “frequency” f of the signal is $1/2\pi$ times the phasor’s angular velocity.

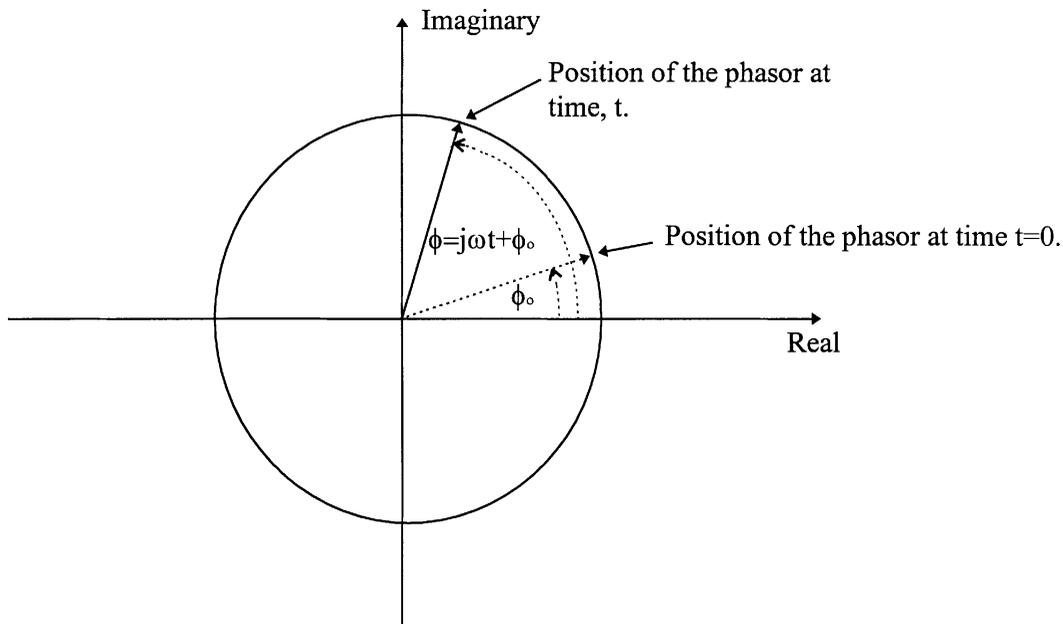


FIGURE 41.3 A complex rotating phasor, $A \exp(j\omega t + \phi_0)$. The length of the phasor is A and its angular velocity is ω . The real part of the phasor is $A \cos(\omega t + \phi_0)$.

There are a number of ways to define the phase of a real sinusoid with unknown amplitude, frequency, and initial phase. One way, as already discussed, is to define it as the fractional part of the period that has been completed. This is a valid and intuitively pleasing definition, and one that can be readily generalized to periodic signals that contain not only a sinusoid, but also a number of harmonics. It cannot, however, be elegantly generalized to allow for slow variations in the frequency of the signal, a scenario that occurs in communications with phase and frequency modulation. Gabor put forward a definition in 1946 that can be used for signals with slowly varying frequency. He proposed a mathematical definition for generating the complex phasor, $z(t)$, associated with the real signal, $s(t)$. The so-called *analytic signal* $z(t)$ is defined according to the following definition [1]:

$$z(t) = s(t) + j \mathcal{H} \{ s(t) \} \quad (41.3)$$

where $\mathcal{H} \{ \}$ denotes the *Hilbert transform* and is given by:

$$\mathcal{H} \{ s(t) \} = p.v. \left[\int_{-\infty}^{+\infty} \frac{s(t-\tau)}{\pi\tau} d\tau \right] \quad (41.4)$$

with *p.v.* signifying the Cauchy principal value of the integral [2].

The imaginary part of the analytic signal can be generated practically by passing the original signal through a “Hilbert transform” filter. From Equations 41.3 and 41.4, it follows that this filter has impulse response given by $1/\pi t$. The filter can be implemented, for example, with one of the HSP43xxx series of ICs from Harris Semiconductors. Details of how to determine the filter coefficients can be found in [2].

Having formally defined the analytic signal, it is possible to provide definitions for phase, frequency, and amplitude as functions of time. They are given below.

$$\text{Phase: } \phi(t) = \arg\{z(t)\} \quad (41.5)$$

$$\text{Frequency: } f(t) = \frac{1}{2\pi} \frac{d[\arg\{z(t)\}]}{dt} \quad (41.6)$$

$$\text{Amplitude: } A(t) = \text{abs}[z(t)] \quad (41.7)$$

The definitions for phase, frequency, and amplitude can be used for signals whose frequency and/or amplitude vary slowly with time. If the frequency and amplitude do vary with time, it is common to talk about the “instantaneous frequency” or “instantaneous amplitude” rather than simply the frequency or amplitude.

Note that in the analytic signal, the imaginary part lags the real part by 90° . This property actually holds not only for sinusoids, but for the real and imaginary parts of all frequency components in “multicomponent” analytic signals as well. The real and imaginary parts of the analytic signal then correspond to the “in-phase (I)” and “quadrature (Q)” components used in communications systems.

In a balanced three-phase electrical power distribution system, the analytic signal can be generated by appropriately combining the different outputs of the electrical power signal; that is, it can be formed according to:

$$z(t) = v_a(t) + \frac{j}{\sqrt{3}} [v_c(t) - v_b(t)] \quad (41.8)$$

where $v_a(t)$ = Reference phase

$v_b(t)$ = Phase that leads the reference by 120°

$v_c(t)$ = Phase that lags the reference by 120° .

41.2 The Phase of a Periodic Nonsinusoidal Signal

It is possible to define “phase” for signals other than sinusoidal signals. If the signal has harmonic distortion components present in addition to the fundamental, the signal will still be periodic, but it will no longer be sinusoidal. The phase can still be considered to be the fraction of the period completed. The “start” of the period is commonly taken to be the point at which the initial phase of the fundamental component is 0, or at a zero-crossing. This approach is equivalent to just considering the phase of the fundamental, and ignoring the other components. The Fourier method provides a very convenient method for determining this phase — the energy of the harmonics in the Fourier transform can be ignored.

41.3 Phase Measurement Techniques

Direct Oscilloscope Methods

Cathode-ray oscilloscopes (CROs) provide a simple means for measuring the phase difference between two sinusoidal signals. The most straightforward approach to use is direct measurement; that is, the signal of interest is applied to the vertical input of the CRO and an automatic time sweep is applied to the horizontal trace. The phase difference is the time delay between the two waveforms measured as a fraction of the period. The result is expressed as a fraction of 360° or of 2π radians; that is, if the time

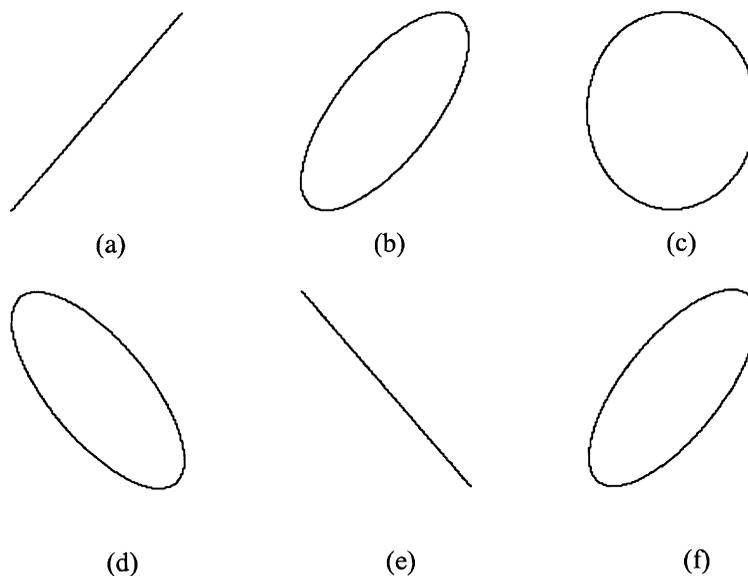


FIGURE 41.4 Lissajous figures for two equal-amplitude, frequency-synchronized signals with a relative phase difference of (a) 0, (b) $\pi/4$, (c) $\pi/2$, (d) $3\pi/4$, (e) π , (f) $-\pi/4$.

delay is 1/4 of the period, then the phase difference is 1/4 of $360^\circ = 90^\circ$ (see Figure 41.2). If the waveforms are not sinusoidal but are periodic, the same procedure can still be applied. The phase difference is just expressed as a fraction of the period or as a fractional part of 360° .

Care must be taken with direct oscilloscope methods if noise is present. In particular, the noise can cause triggering difficulties that would make it difficult to accurately determine the period and/or the time delay between two different waveforms. The “HF reject” option, if available, will alleviate the triggering problems.

Lissajous Figures

Lissajous figures are sometimes used for the measurement of phase. They are produced in an oscilloscope by connecting one signal to the vertical trace and the other to the horizontal trace. If the ratio of the first frequency to the second is a rational number (i.e., it is equal to one small integer divided by another), then a closed curve will be observed on the CRO (see Figures 41.4 and 41.5). If the two frequencies are unrelated, then there will be only a patch of light observed because of the persistence of the oscilloscope screen.

If the two signals have the same frequency, then the Lissajous figure will assume the shape of an ellipse. The ellipse’s shape will vary according to the the phase difference between the two signals, and according to the ratio of the amplitudes of the two signals. Figure 41.6 shows some figures for two signals with synchronized frequency and equal amplitudes, but different phase relationships. The formula used for determining the phase is:

$$\sin(\phi) = \pm \frac{Y}{H} \quad (41.9)$$

where H is half the maximum vertical height of the ellipse and Y is the intercept on the y -axis. Figure 41.7 shows some figures for two signals that are identical in frequency and have a phase difference of 45° , but with different amplitude ratios. Note that it is necessary to know the direction that the Lissajous trace is moving in order to determine the sign of the phase difference. In practice, if this is not known *a priori*,

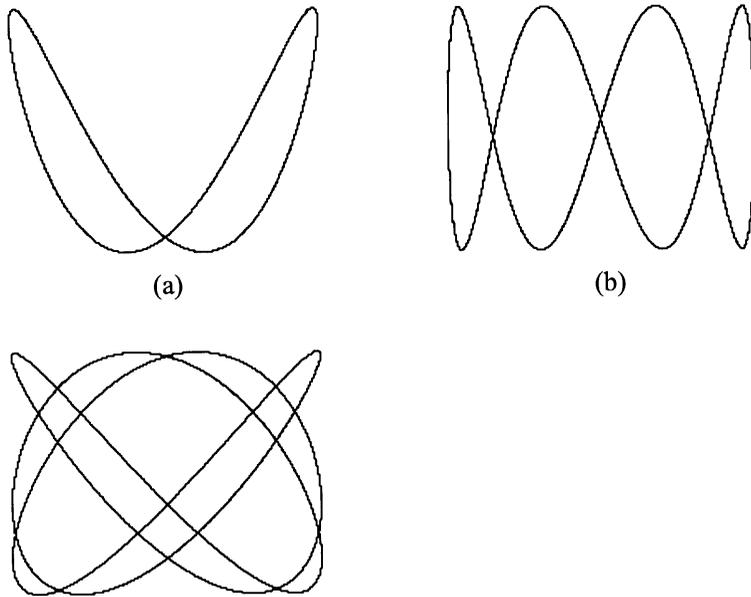


FIGURE 41.5 Lissajous figures for two signals with vertical frequency: horizontal frequency ratios of (a) 2:1, (b) 4:1, (c) 4:3.

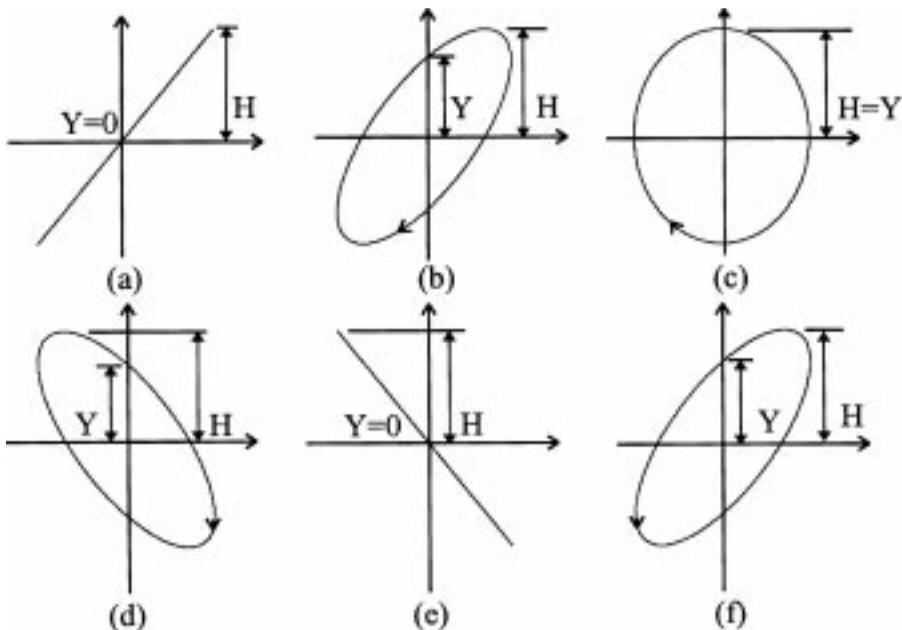


FIGURE 41.6 Lissajous figures for two signals with synchronized frequency and various phase differences: (a) phase difference = 0° , (b) phase difference = 45° , (c) phase difference = 90° , (d) phase difference = 135° , (e) phase difference = 180° , (f) phase difference = 315° .

then it can be determined by testing with a variable frequency signal generator. In this case, one of the signals under consideration is replaced with the variable frequency signal. The signal generator is adjusted until its frequency and phase equal that of the other signal input to the CRO. When this happens, a straight line will exist. The signal generator frequency is then increased a little, with the relative phase

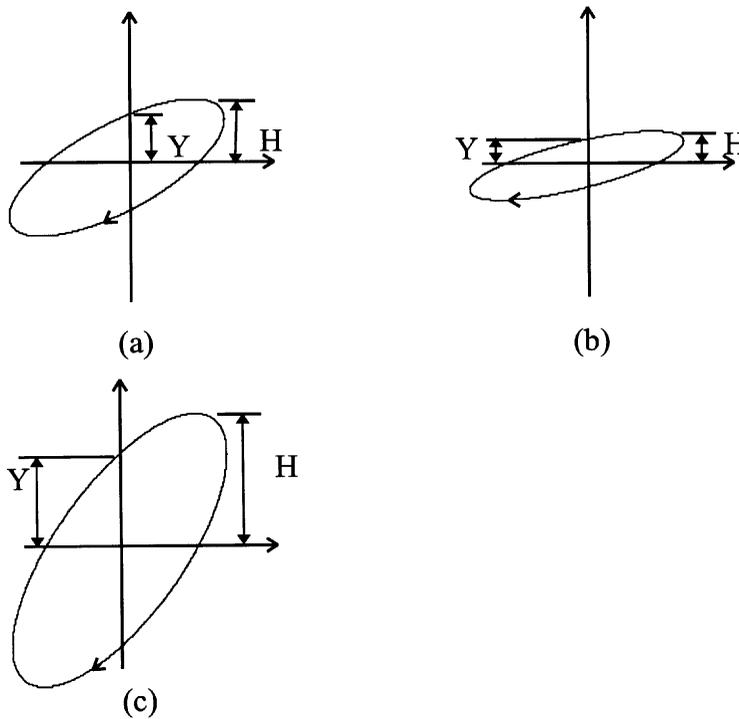


FIGURE 41.7 Lissajous figures for two signals with synchronized frequency, a phase difference of 45° , and various amplitude ratios: (a) amplitude ratio of 1, (b) amplitude ratio of 0.5, (c) amplitude ratio of 2.

thus being effectively changed in a known direction. This can be used to determine the correct sign in Equation 41.9.

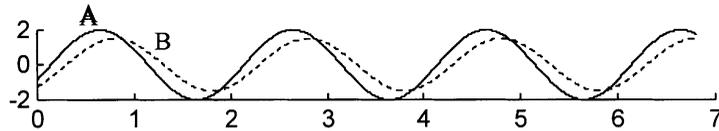
Lissajous figure methods are a little more robust to noise than direct oscilloscope methods. This is because there are no triggering problems due to random noise fluctuations. Direct methods are, however, much easier to interpret when harmonics are present. The accuracy of oscilloscope methods is comparatively poor. The uncertainty of the measurement is typically in excess of 1° .

Zero-Crossing Methods

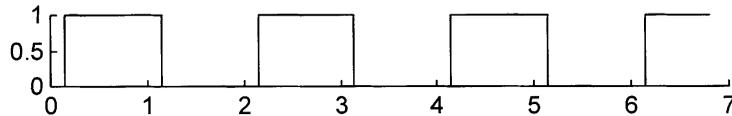
This method is currently one of the most popular methods for determining phase difference, largely because of the high accuracy achievable (typically 0.02°). The process is illustrated in Figure 41.8 for two signals, denoted A and B , which have the same frequency but different amplitudes. Each negative to positive zero-crossing of signal A triggers the start of a rectangular pulse, while each negative to positive zero-crossing of signal B triggers the end of the rectangular pulse. The result is a pulse train with a pulse width proportional to the phase angle between the two signals. The pulse train is passed through an averaging filter to yield a measure of the phase difference. It is also worth noting that if the positive to negative zero-crossings are also used in the same fashion, and the two results are averaged, the effects of dc and harmonics can be significantly reduced.

To implement the method practically, the analog input signals must first be converted to digital signals that are “high” if the analog signal is positive, and “low” if the analog signal is negative. This can be done, for example, with a Schmitt trigger, along with an RC stabilizing network at the output. Chapter 81 provides a circuit to do the conversion. In practice, high-accuracy phase estimates necessitate that the switching of the output between high and low be very sharp. One way to obtain these sharp transitions is to have several stages of “amplify and clip” preceding the Schmitt trigger.

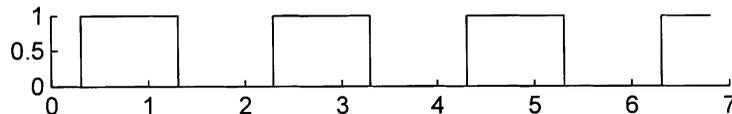
(a) The 2 input signals, A & B.



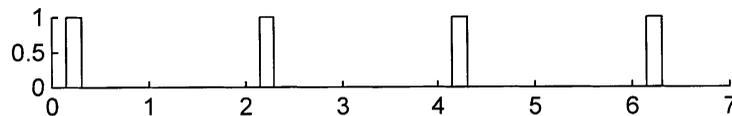
(b) Zero-crossing detector output for A.



(c) Zero-crossing detector output for B.



(d) Bistable flip-flop output. (The input signals are shown in (b) & (c)).



(e) Phase difference (degrees) obtained by averaging bistable output.

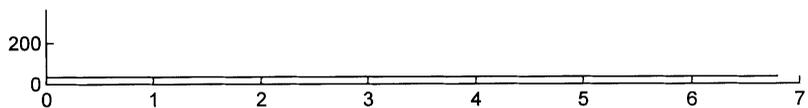


FIGURE 41.8 Input, output and intermediate signals obtained with the zero-crossing method for phase measurement. Note that the technique is not sensitive to signal amplitude.

The digital portion of the zero-crossing device can be implemented with an edge-triggered RS flip-flop and some ancillary circuitry, while the low-pass filter on the output stage can be implemented with an RC network. A simple circuit to implement the digital part of the circuit is shown in Chapter 81.

A simpler method for measuring phase based on zero-crossings involves the use of an exclusive or (XOR) gate. Again, the analog input signals must first be converted to digital pulse trains. The two inputs are then fed into an XOR gate and finally into a low-pass averaging filter. The circuit is illustrated in Chapter 81. A disadvantage with this method is that it is only effective if the duty cycle is 50% and if the phase shift between the two signals is between 0 and π radians. It is therefore not widely used.

The Three-Voltmeter Method

The measurement of a phase difference between two voltage signals, v_{ac} , and v_{bc} , can be expedited if there is a common voltage point, c . The voltage between points b and a (v_{ba}), the voltage between points b and c (v_{bc}), and the voltage between points c and a (v_{ca}) are measured with three different voltmeters. A vector diagram is constructed with the three measured voltages as shown in Figure 41.9. The phase difference between the two vectors, v_{ac} and v_{bc} , is determined using a vector diagram (Figure 41.9) and the cos rule. The formula for the phase difference, ϕ , in radians is given by:

$$\pi - \phi = \cos^{-1} \left(\frac{v_{ca}^2 + v_{bc}^2 - v_{ba}^2}{2v_{ca}v_{bc}} \right) \quad (41.10)$$

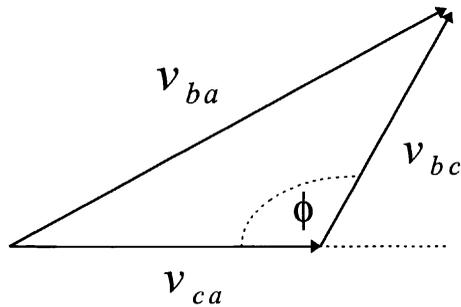


FIGURE 41.9 A vector diagram for determining the phase angle, ϕ , between two ac voltages, v_{ac} and v_{bc} , with the three-voltmeter method.

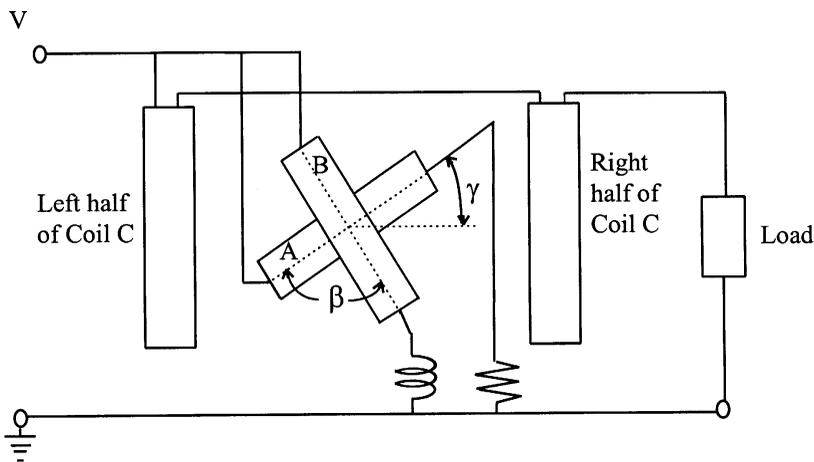


FIGURE 41.10 Diagram of a crossed-coil device for measuring phase. Coils A and B are the rotating coils. Coil C (left and right parts) is the stationary coil.

The Crossed-Coil Method

The crossed-coil phase meter is at the heart of many analog power factor meters. It has two crossed coils, denoted A and B, positioned on a common shaft but aligned at different angles (see Figure 41.10). The two coils move as one, and the angle between them, β , never changes. There is another independent nonrotating coil, C, consisting of two separate parts, “enclosing” the rotating part (see Figure 41.10). The separation of coil C into two separate parts (forming a Helmholtz pair) allows the magnetic field of coil C to be almost constant in the region where the rotating A and B coils are positioned.

Typically the system current, I , is fed into coil C, while the system voltage, V , is applied to coil A via a resistive circuit. The current in coil A is therefore in phase with the system voltage, while the current in coil C is in phase with the system current. Coil B is driven by V via an inductive circuit, giving rise to a current that lags V (and therefore the current in coil A) by 90° . In practice, the angle between the currents in coils A and B is not quite 90° because of the problems associated with achieving purely resistive and purely inductive circuits. Assume, then, that this angle is β . If the angle between the currents in coil B and in coil C is ϕ , then the angle between the currents in coils A and C is $\beta + \phi$. The average torque induced in coil A is proportional to the product of the average currents in coils A and C, and to the

cosine of the angle between coil A and the perpendicular to coil C. The average torque induced in coil A is therefore governed by the equation:

$$\bar{T}_A \propto I_A I_C \cos(\phi + \beta) \cos(\gamma) = k_A \cos(\phi + \beta) \cos(\gamma) \quad (41.11)$$

where I_A and I_C = Constants

ω = Angular frequency

$\phi + \beta$ = Relative phase between the currents in coils A and C

γ = Angle between coil A and the perpendicular to coil C

\propto = Signifies "is proportional to"

Assuming that the current in coil B lags the current in coil A by β , then the average torque in coil B will be described by:

$$\bar{T}_B \propto I_B I_C \cos(\phi) \cos(\gamma + \beta) = k_B \cos(\phi) \cos(\gamma + \beta) \quad (41.12)$$

where I_B is a constant, ϕ is the relative phase between the currents in coils B and C, and the other quantities are as in Equation 41.11.

Now, the torques due to the currents in coils A and B are designed to be in opposite directions. The shaft will therefore rotate until the two torques are equal; that is, until:

$$k_A \cos(\phi + \beta) \cos(\gamma) = k_B \cos(\phi) \cos(\gamma + \beta) \quad (41.13)$$

If $k_A = k_B$, then Equation 41.13 will be satisfied when $\phi = \gamma$. Thus, the A coil will be aligned in the direction of the phase shift between the load current and load voltage (apart from errors due to the circuits of the crossed coils not being perfectly resistive/inductive). Thus, a meter pointer attached to the A plane will indicate the angle between load current and voltage. In practice, the meter is usually calibrated to read the cosine of the phase angle rather than the phase angle, and also to allow for the errors that arise from circuit component imperfections.

The accuracy of this method is limited, due to the heavy use of moving parts and analog circuits. Typically, the measurement can only be made accurate to within about 1°.

Synchrosopes and Phasing Lamps

The crossed-coil meter described above is used as the basis for *synchrosopes*. These devices are often used in power generation systems to determine whether two generators are phase and frequency synchronized before connecting them together. In synchrosopes, the current from one generator is fed into the fixed coil and the current from the other generator is fed into the movable crossed coils. If the two generators are synchronized in frequency, then the meter needle will move to the position corresponding to the phase angle between the two generators. If the generators are not frequency synchronized, the meter needle will rotate at a rate equal to the difference between the two generator frequencies. The direction of rotation will indicate which generator is rotating faster.

When frequency synchronization occurs (i.e., the meter needle rotation ceases) and the phase difference is zero, the generators can be connected together. Often in practice, the generators are connected before synchronization occurs; the generator coming on-line is deliberately made a little higher in frequency so that it can provide extra power rather than be an extra drain on the system. The connection is still made, however, when the instantaneous phase difference is zero.

Phasing lamps are sometimes used as a simpler alternative to synchrosopes. A lamp is connected between the two generators, and any lack of frequency synchronization manifests as a flickering of the lamp. A zero phase difference between the two generators corresponds to maximum brightness in the lamp.

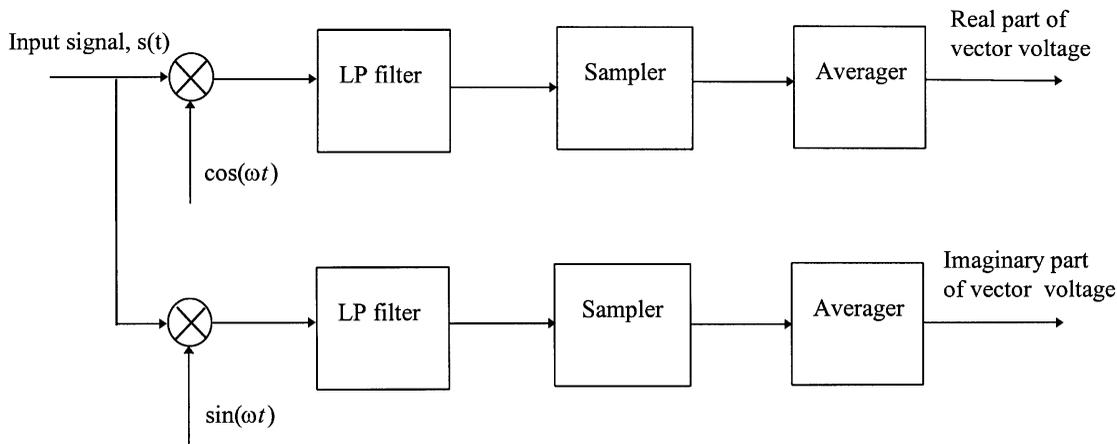


FIGURE 41.11 Vector voltmeter block diagram. The vector voltmeter determines the voltage (amplitude and phase or real and imaginary parts) of the component of the input signal at frequency f .

Vector Voltmeters and Vector Impedance Methods

Alternating voltages (and currents) are often characterized as vectors consisting of a magnitude and a phase, with the phase being measured relative to some desired reference. Many instruments exist that can display the voltage amplitude and phase of a signal across a wide range of frequencies. These instruments are known as *vector voltmeters* or *network analyzers*. The phase and amplitude as a function of frequency can be obtained very simply in principle by taking the Fourier transform of the signal and simply reading the amplitude and phase across the continuum of frequencies. To achieve good accuracy, this is typically done with down-conversion and digital processing in the baseband region. The down-conversion can be analog, or it can be digital. The procedure is described more fully in the succeeding paragraphs.

To determine the real part of the voltage vector at a given frequency f , the signal is first down-converted by mixing with a local oscillator signal, $\cos(2\pi ft)$. This mixing of the signal recenters the frequency component of interest at 0 Hz. The resultant signal is low-pass filtered, digitally sampled (if not in the digital domain already), and averaged. The digital sampling and averaging enables the amplitude of the newly formed 0-Hz component to be evaluated. The imaginary part is obtained in similar fashion by mixing the signal with $\sin(2\pi ft)$, low-pass filtering, digitally sampling, and again averaging the samples. The amplitude and phase of the voltage vector, V , are obtained from the real and imaginary parts using the standard trigonometric relationships:

$$\text{Magnitude} = \text{Abs}(V) = \sqrt{[\Re\{V\}]^2 + [\Im\{V\}]^2} \quad (41.14)$$

$$\text{Phase} = \arg(V) = \arctan(\Im\{V\} / \Re\{V\}) \quad (41.15)$$

where $\Re\{\cdot\}$ and $\Im\{\cdot\}$ denote the real and imaginary parts, respectively.

The procedure for forming the vector voltage is summarized in the block diagram in [Figure 41.11](#). In practice, the down-conversion can be carried out in more than one step. For high-frequency signals, for example, the first stage might shift a large band of frequencies to the audio region, where further down-conversion is carried out. Alternatively, the first stage might shift a band of frequencies to the intermediate frequency (IF) band, and the second stage to the audio band. More details on the physics of the down-conversion process are available in the article on “Modulation” in Chapter 81.

Just as it is possible to analyze a voltage signal and produce a magnitude and phase across any given frequency band, so it is possible to obtain a frequency profile of the magnitude and phase of a *current* signal. If current vectors and voltage vectors can be obtained for a given impedance, then it is possible to obtain a “vector impedance.” This impedance is defined simply as the result of the complex division of voltage by current:

$$Z = \frac{V}{I} \quad (41.16)$$

The calculation of vector impedances are useful for such applications as impedance matching, power factor correction, and equalization.

Typically, much of the current processing for vector voltmeters and vector impedance meters is done digitally. One of the great advantages of this type of processing is the high accuracy achievable. Accuracies of 0.02° are common, but this figure is improving with developing technology. The high sampling rates that can be employed (typically beyond 1 GHz) cause the errors in the A/D conversion to be spread out over very large bandwidths. Since the ultimate measurement of a vector voltage or impedance is usually made over a very narrow bandwidth, the errors are substantially eliminated. The developments in technology that enable greater accuracy are (1) sampling rate increases, (2) word-length increases, and (3) increased A/D converter fidelity.

Phase Standard Instruments

For high-precision phase measurements and calibration, “phase standard” instruments can be used. These instruments provide two sinusoidal signal outputs, whose phase difference can be controlled with great accuracy. They typically use crystal-controlled timing to digitally synthesize two independent sinusoids with a variable user-defined phase difference. The Clarke-Hess 5500 Digital Phase Standard is one such instrument. For this standard, the sinusoids can have frequencies ranging from 1 Hz to 100 kHz, while amplitudes can vary between 50 mV and 120 V rms. The phase can be set with a resolution of 0.001° , with a typical accuracy of about 0.003° .

The Fast Fourier Transform Method

This method is one in which virtually all the processing is done in the digital domain. It operates on the pulse code modulated (PCM) digital samples of a signal. This and other similar methods are very promising means for measuring phase. This is because of the digital revolution that has resulted in cheap, fast, accurate, and highly versatile digital signal processors (DSPs). The latter are small computer chips capable of performing fast additions and multiplications, and which can be programmed to emulate conventional electronic functions such as filtering, coding, modulation, etc. They can also be programmed to perform a wide variety of functions not possible with analog circuitry. Up until the end of the 1980s, digital measurement was limited by the relatively inaccurate analog-to-digital (A/D) conversion process required before digital processing could be performed. Developments in the early 1990s, however, saw the introduction of oversampling analog-to-digital converters (ADCs), which can achieve accuracies of about one part in 100,000 [3]. ADC speeds as well as DSP chips are now running reliably at very high speeds.

In the fast Fourier transform (FFT) method, the digital signal samples are Fourier transformed with an FFT [2]. If the signal is sinusoidal, the initial phase is estimated as that value of the phase where the Fourier transform is maximized [4]. The frequency of the signal is estimated as that value of frequency where the Fourier transform is maximized. Once measurements of the frequency f and initial phase ϕ_0 have been obtained, the phase at any point in time can be calculated according to:

$$\phi = 2\pi ft + \phi_0 \quad (41.17)$$

One important practical issue in the measurement of the frequency and initial phase with an FFT arises because the FFT yields only *samples* of the Fourier transform; that is, it does not yield a continuous Fourier transform. It is quite possible that the true maximum of the Fourier transform will fall *between* samples of the FFT. For accurate measurement of frequency and initial phase, then, it is necessary to *interpolate* between the FFT samples. An efficient algorithm to do this is described in [5].

The FFT method is particularly appealing where there is significant noise present, as it is effective down to quite low signal-to-noise ratios (SNRs). Furthermore, it provides an optimal estimate of the frequency and initial phase, providing the background noise is white and Gaussian, and that no harmonic components are present [4]. If harmonics are present, the estimate of the phase is commonly taken as the phase at the FFT peak corresponding to the fundamental; this is not an optimal estimate, but serves well in many applications. An optimal estimate in the case when harmonics are present can be obtained, if necessary, with the algorithms in [6], [7], and [8]. DSP chips such as the Texas Instruments TMS320C3x family, the Analog Devices ADSP21020 chip, or the Motorola MC5630x series can be used to implement the real-time FFTs.

If long word-lengths are used (say 32 bits) to perform the arithmetic for the FFTs, then determination of the phase from the samples of the FFT is virtually error-free. The only significant inaccuracy incurred in determining the phase is due to the ADC errors. Moreover, the error due to the digitization will typically be spread out over a large bandwidth, only a small amount of which will be “seen” in the phase measurement. With a high-quality ADC, accuracies of less than 0.001° are possible.

Phase-Locked Loops

If the frequency of a signal changes significantly over the period of the measurement, the FFT method described above will provide inaccurate results. If the signal’s frequency does change substantially during measurement, one means to estimate the phase of the signal is to use a phase-locked loop (PLL). In this case, the signal, $s(t) = A \sin(\omega t + \phi(t))$, can be thought of as a constant frequency component, $A \sin(\omega t)$, which is phase modulated by a time-varying phase component, $\phi(t)$. The problem then reduces largely to one of demodulating a phase-modulated (PM) signal. A PLL can be used to form an estimate of the “phase-modulating” component, $\hat{\phi}(t)$, and the overall phase of the signal, ϕ_{oa} , can be estimated according to:

$$\phi_{\text{oa}}(t) = \omega t + \hat{\phi}(t) \quad (41.18)$$

Either analog or digital PLLs can be used, although higher accuracy is attainable with digital PLLs. Analog PLLs for demodulating a frequency-modulated (FM) signal are discussed in Chapter 81 and in [12]. The digital PLL (DPLL) was developed as an extension of the conventional analog PLL and is therefore similar in structure to its analog counterpart. The DPLL is discussed in [9] and [10]. The equation to demodulate the digital modulated signal with a first order DPLL is a simple recursive equation [9].

A block diagram of the DPLL for demodulating a PM signal is shown in [Figure 41.12](#). In this diagram, n represents the discrete-time equivalent of continuous time t . It can be seen that there are strong similarities between [Figure 41.12](#) and the analog PLL-based FM demodulator in Chapter 81. Both have phase comparators (implemented by the multiplier in [Figure 41.12](#)), both have loop filters, and both have modulators (either PM or FM) in the feedback loop.

The DPLL can easily be extended to measure the phase of a signal consisting not just of a fundamental component, but also of harmonically related components. Details are provided in [11]. The DPLL is near optimal for phase estimation in white Gaussian background noise down to a signal power-to-noise power ratio of about 8 dB [10].

The DPLL will function effectively whether the phase is constant or time-varying. Unlike the FFT, the DPLL is a recursive algorithm, with the feedback involved in the recursion creating a vulnerability to quantization errors. However, with proper precautions and long word-lengths, the DPLL will introduce

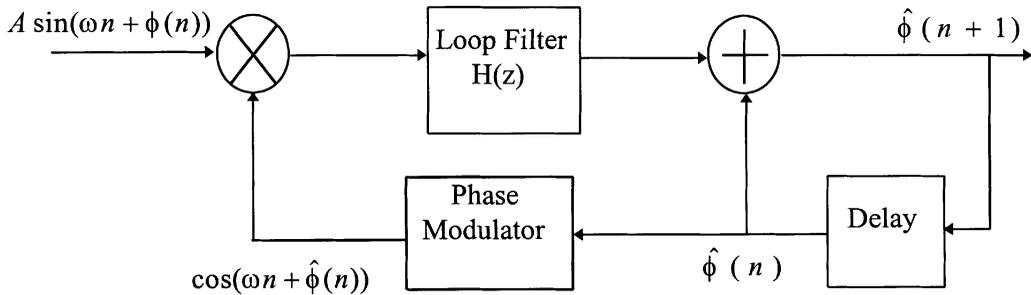


FIGURE 41.12 Block diagram of a digital phase-locked loop to implement phase demodulation.

minimal processing error. The main error would then arise from the inaccuracy of the ADC. With appropriate conditioning, one could expect the DPLL to provide accuracies approaching 0.001° .

41.4 Phase-Sensitive Demodulation

It is frequently necessary to track the phase of a carrier that “jitters” in some uncertain manner. This tracking of the carrier phase is necessary for synchronous demodulation schemes, where the phase of the demodulating signal must be made equal to the phase of the carrier. This need is explained in Chapter 81, and is briefly re-explained here. Consider, for example, double sideband (DSB) amplitude modulation. In DSB, the modulated signal is given by $f_s(t) = A[k + \mu m(t)]\cos(\omega_c t)$, where $m(t)$ is the message signal, A is the amplitude of the unmodulated carrier, μ is the modulation index, k is the proportion of modulating signal present in the modulated signal, and $\cos(\omega_c t)$ is the carrier. Demodulation is typically carried out by multiplying $f_s(t)$ by the carrier, and then low-pass filtering so that the demodulated signal is given by:

$$f_d(t) = \frac{A_c [k + \mu m(t)]}{2} \quad (41.19)$$

However, if because of carrier uncertainty, one multiplies the modulated signal by $\cos(\omega_c t + \phi)$, then the demodulated signal is given by:

$$f_d(t) = \frac{A_c [k + \mu m(t)] \cos(\phi)}{2} \quad (41.20)$$

It can be seen from Equation 41.20 that the error in the carrier phase can affect both the amplitude and the sign of the demodulated signal. The phase errors can thus yield substantial errors in system output. The following sections outline important techniques used for tracking the phase of carriers, and thus reducing phase errors.

The Phase-Locked Loop for Carrier Phase Tracking

The phase-locked loop (PLL) is well known as a means for demodulating frequency-modulated signals. It is also frequently used for tracking the phase of a carrier in noise, so that a copy of the carrier with correct phase is available for demodulation. This tracking is simple enough if a (noisy) copy of the carrier is directly available; either a digital or analog PLL can be used. In either case, the input can be assumed to have the form, $A \sin(\omega t + \phi(t))$, where $\phi(t)$ is the carrier phase. The PLL consists of a multiplier (effectively a phase comparator), a phase modulator, and a loop filter, arranged as shown in Chapter 81. The design of the loop filter is critical if noise is to be optimally removed. In proper

operation, the PLL output will track the phase of the incoming signal (i.e., of the carrier). If a copy of the carrier is not available but needs to be inferred from the modulated signal, the demodulation task is more difficult. Digital PLLs using a DSP chip can be particularly helpful in this case; the carrier can be adaptively estimated using intelligent algorithms, with convergence to the “correct” signal being registered when certain desired features of the demodulated signal are observed.

The PLL is quite versatile. It can function in relatively low noise environments (typically down to about 8 dB SNR). It can be implemented digitally. It can also cope with substantial carrier frequency variations by increasing the order of the loop filter [12]; (this is often necessary, for example, in satellite communications because of the Doppler effect). At very low SNR, however, the PLL fails badly. Recent developments in digital signal processing have seen the development of an alternative based on hidden Markov models (HMMs), which will function down to about -5 dB SNR [13]. The HMM method is discussed in the next section.

Hidden Markov Model-Based Carrier Phase Tracker

In the HMM method, the problem of estimating the phase and frequency of a noisy waveform is couched as a “state estimation” problem. The phase of the signal at any point in time can go from 0 to 360°. The 0 to 360° value range is divided into a finite number of intervals or “states,” so that the phase at any time occupies a particular (though unknown) state. Similarly, the angular frequency normalized by the sampling frequency at any time in a digital system must be between $-\pi$ to $+\pi$. This value range is also divided into a number of states, so that the frequency at any time has a (hidden or unknown) state associated with it. The frequency is assumed to be a first-order Markov process and probabilities are assigned to the possibility of the frequency changing from one state to another for successive values of time, i.e., frequency transition probabilities are assigned. Large frequency changes are assigned low probabilities, while small changes are assigned high probabilities. The problem of estimating the true phase and frequency states underlying the noisy signal then reduces to one of estimating which states the phase and frequency occupy as time evolves, given the observed noisy signal and the transition probabilities. Computationally efficient optimal algorithms have been developed to estimate these “optimal state sequences” for both the phase and frequency. Details are provided in [13].

41.5 Power Factor

Of particular interest in many applications is the phase angle between the current and voltage of a system. This angle is important because it is a measure of the power which is dissipated in the system. The following paragraphs discuss this angle, its cosine (the system power factor), and its measurement.

In a linear electric circuit that is fed by a current of peak amplitude, I_M , with an angular frequency of ω , the current will have the form, $I_M \cos(\omega t)$. The system voltage will be given by $V_M \cos(\omega t + \phi)$, where V_M is the peak voltage and ϕ is the phase difference between the current and voltage. Then the average power dissipated in the circuit will be given by:

$$P_{av} = \frac{1}{2} V_M I_M \cos(\phi) = V_{rms} I_{rms} \cos(\phi) \quad (41.21)$$

where V_{rms} and I_{rms} are the root mean square (rms) values of the voltage and current respectively. The term $\cos(\phi)$ is known as the *power factor*. It may alternatively be expressed as the ratio of real average power to the product of the rms values of voltage and current, respectively:

$$PF = \frac{P_{av}}{V_{rms} I_{rms}} \quad (41.22)$$

TABLE 41.1 Integrated Circuits Used in Phase Measurement

Function	Designation	Manufacturer	Approximate price
Phase-locked loop	LM566	National, Motorola, Phillips	\$2.70
Phase-locked loop	74HC4046	Harris, Motorola	\$2
Phase/frequency detector	MC4044P	Motorola	\$18.29
Pair of retriggerable monostables (one-shot)	74HC4538	Motorola, Harris	\$2
DSP Chip	TMS320C32	Texas Instruments	\$50
DSP Chip	TMS320C31	Texas Instruments	\$80
DSP Chip	MC56303	Motorola	\$60
DSP Chip	ADSP21020	Analog Devices	\$110

The above expression is, in fact, a general definition of power factor for any current and voltage waveforms. For the special case of sinusoidal voltage and currents, PF reduces to $\cos(\phi)$.

There are a number of ways to measure the power factor. One way is to use a wattmeter to measure the real average power and a voltmeter and an ammeter to measure the rms voltage and current, respectively. The power factor is then determined according to Equation 41.22. This is probably the most effective way when the currents and/or voltages are nonsinusoidal. This procedure can easily be implemented with “digital power meters.” The power is measured by time-averaging the product of the instantaneous voltage and current, while the rms values are calculated by taking the square root of the time averaged value of the square of the parameter of interest — current or voltage. Some digital power meters also provide an analysis of the individual harmonics via FFT processing. These meters are accurate and versatile, and consequently very popular.

A more direct method is based on the crossed-coil meter, the operation of which was described earlier in this chapter. Note that this meter is a “single-phase meter,” which is accurately designed for one frequency only. Errors will occur at other frequencies because of the dependence of the crossed-coil meter method on a constant and known phase angle between the currents in the crossed coils.

With balanced polyphase circuits, it is possible to use a single-phase meter applied to one of the phases. Alternatively, one can use specially designed polyphase meters. In a three-phase meter, for example, one phase is connected to the fixed coil, while the other two phases are connected to the crossed coils on the rotating shaft. The crossed coils are constructed with a 60° angle between them. With four-phase systems, consecutive lines are 90° out of phase. Two of these consecutive lines are connected to the two crossed-coils and the angle between the coils is made equal to 90° .

With unbalanced polyphase circuits amid the presence of harmonics, each of the harmonic components has its own power factor, and so it is likely to be misleading to use a meter that measures a single angle. These methods based on the crossed-coil meter are thus much more limited than their digital counterparts.

41.6 Instrumentation and Components

Table 41.1 lists some integrated circuits and DSP chips that can be used in the various techniques for measuring phase. The list is really only illustrative of what is available and prices are approximate costs in U.S. dollars for small quantities at the end of 1996. Table 41.2 lists some companies that manufacture these products. An extensive (and indeed rapidly increasing) product range exists for DSP chip-based products, with details being available from the companies listed in Table 41.2. Table 41.3 lists instruments used for phase measurement. These instruments include CROs, vector voltage meters, vector impedance meters, crossed-coil meters and digital power meters, zero-crossing meters, and phase standards. Again, the table is only representative, as the full range of available instruments is enormous. Addresses of some of the relevant companies are provided in Table 41.4.

TABLE 41.2 Companies Making Integrated Circuits and DSP Chips Which Can Be Used for Phase Measurement

Analog Devices, Inc. One Technology Way Box 9106, Norwood, MA 02062 Tel: (617) 329-4700.	National Semiconductor Corp. 2900 Semiconductor Dr. P.O. Box 58090 Santa Clara, CA 95052-8090
Harris Semiconductor Products Division P.O. Box 883 Melbourne, FL 37902 Tel: (407) 724-3730	Texas Instruments Incorporated P.O. Box 1443 Houston, Texas 77251-1443
Motorola, Semiconductor Products Sector 3102 N. 56th St. Phoenix, AZ 85018 Tel: (602) 952-3248	

TABLE 41.3 Instruments for Measuring Phase

Description	Model number	Manufacturer	Approximate price
CRO	HP54600B	Hewlett Packard	\$2,495
CRO	HP54602B	Hewlett Packard	\$2,995
CRO	HP54616	Hewlett Packard	\$5,595
CRO	TDS220	Tektronix	\$1,695
CRO	TDS510A	Tektronix	\$9,956
Vector signal analyzer	HP89410A	Hewlett Packard	\$29,050
Vector signal analyzer	HP89440A	Hewlett Packard	\$52,500
Vector signal analyzer	HP89441A	Hewlett Packard	\$58,150
Gain/phase Impedance meter	HP4193A	Hewlett Packard	\$13,700
Zero-crossing phase meter	KH6500	Krohn-Hite	\$1,300
Digital power analyzer (with power factor & phase)	NANOVIP	Elcontrol	\$660
Digital analyzing vector voltmeter	NA2250	North Atlantic Instruments	
Digital power analyzer (with power factor & phase, FFT analysis)	3195	Hioki	\$25,000
Crossed-coil meter	246-425G	Crompton Industries	\$290
Digital phase standard	5500	Clarke-Hess	\$11,000

TABLE 41.4 Companies Making Instruments for Measuring Phase

Hewlett-Packard Co. Test and Measurement Sector P.O. Box 58199 Santa Clara, CA 95052-9943 Tel: (800) 452-4844	Krohn-Hite Corporation Bodwell St., Avon Industrial Park Avon, MA	Hioki 81 Koizumi Veda, Nagano 386-11 Japan
Tektronix Inc. Corporate Offices 26600 SW Parkway P.O. Box 1000 Wilsonville, OR 97070-1000 Tel: (503) 682-3411, (800) 426-2200	Crompton Instruments Freebournes Road, Witham Essex, CM83AH England	Clarke-Hess Comm. Research Corporation 220 W. 19 Street New York, NY
	Elcontrol Via San Lorenzo 1/4 - 40037 Sasso Marconi Bologna, Italy	North Atlantic Instruments http://www.naii.com

References

1. D. Gabor, The theory of communication, *J. Inst. Elec. Eng.*, 93(III), 429-457, 1946.
2. A. V. Oppenheim and R. W. Schaffer, *Discrete-Time Signal Processing*, Englewood-Cliffs, NJ: Prentice-Hall, 1989.
3. K. C. Pohlmann (ed.), *Advanced Digital Audio*, Carmel, IN: Howard Sams and Co., 1991.
4. D. Rife and R. Boorstyn, Single tone parameter estimation from discrete-time observations, *IEEE Trans. Inf. Theory*, 20, 591-598, 1974.
5. T. Abotzoglou, A fast maximum likelihood algorithm for estimating the frequency of a sinusoid based on Newton's algorithm, *IEEE Trans. Acoust., Speech Signal Process.*, 33, 77-89, 1985.
6. D. McMahon and R. Barrett, ML estimation of the fundamental frequency of a harmonic series, *Proc. of ISSPA 87*, Brisbane, Australia, 1987, 333-336.
7. A. Nehorai and B. Porat, Adaptive comb filtering for harmonic signal enhancement, *IEEE Trans. Acoust., Speech Signal Process.*, 34, 1124-1138, 1986.
8. L. White, An iterative method for exact maximum likelihood estimation of the parameters of a harmonic series, *IEEE Trans. Automat. Control*, 38, 367-370 1993.
9. C. Kelly and S. Gupta, Discrete-time demodulation of continuous time signals, *IEEE Trans. Inf. Theory*, 18, 488-493, 1972.
10. B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Englewood Cliffs, NJ: Prentice-Hall, 1979.
11. P. Parker and B. Anderson, Frequency tracking of periodic noisy signals, *Signal Processing*, 20(2), 127-152, 1990.
12. R. E. Best, *Phase-Locked Loops; Theory, Design and Applications*, 2nd ed., New York: McGraw-Hill, 1993.
13. L. White, Estimation of the instantaneous frequency of a noisy signal, in B. Boashash (ed.), *Time-Frequency Signal Analysis, Methods and Applications*, Melbourne, Australia: Longman-Cheshire; New York: Halsted Press, 1992.

Further Information

- A. D. Helfrick and W. D. Cooper, *Modern Electronic Instrumentation and Measurement Techniques*, Englewood Cliffs, NJ: Prentice-Hall, 1990.
- McGraw-Hill Encyclopedia of Science and Technology*, 8th ed., New York: McGraw-Hill, 1997.
- H. Taub and D. L. Schilling, *Principles of Communication Systems*, 2nd ed., New York: McGraw-Hill, 1986.
- J. D. Lenk, *Handbook of Oscilloscopes: Theory and Application*, Englewood Cliffs, NJ: Prentice-Hall, 1982.