A A Short Introduction to Frequency, Phase and Amplitude

v1.3 – June 2021

Prerequisite knowledge required: None

A.1 Introduction

The term *frequency* comes up frequently in electronics, but the word has a very specific meaning in electronics which is slightly different from how the word is sometimes used in non-technical application, and it's very important to be clear about this. You will also hear people talking about the *frequency content* of a signal, and again this term has a specific meaning. The purpose of this note is to try and make the definitions of these terms clear, as well as introducing the concept of phase.

A.1.1 Some definitions

- A *periodic signal* is a signal which exactly repeats for ever (for example a sine wave).
- The *period* of a periodic signal is the smallest time after which the signal exactly repeats.
- The *frequency* of a signal is the inverse of the period.
- The *phase difference* between two signals is the offset between two identical signals, expressed as a proportion of the period, and in terms of an angle (one period is 360 degrees or 2π radians).
- The *amplitude* of a signal is a measure of the size of the waveform (usually expressed in terms of volts or amps).
- The *frequency content* of a periodic signal is the set of amplitudes and frequencies of the sine and cosine waves that added together make up the periodic signal[1].

There are quite a lot of important observations to make about these definitions, so I'll go through them one at a time:

A.2 Periodic signals and the period of a signal

First, note that the definition of a period (and by extension the definition of frequency) only applies to signals that exactly repeat after a certain time interval. Such signals are termed *periodic signals*.

Second, note that the period is defined in terms of the smallest time after which the signal repeats. For example, consider the following signal:

[1] This really isn't obvious (and it's one of the most amazing results in mathematical physics), but any real signal can be made up of a whole series of sine and cosine waves with different frequencies added together. This is the subject of Fourier Analysis, which you'll study in much more detail later.

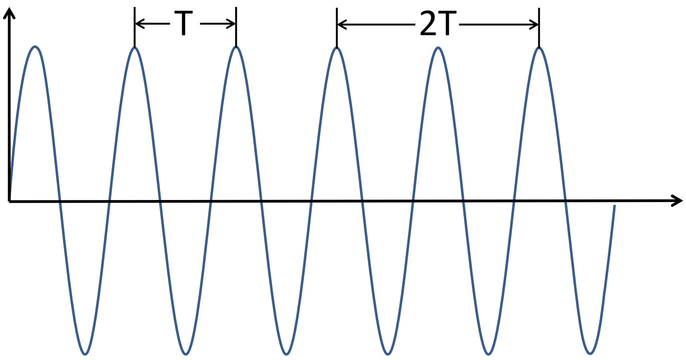

Figure A-1 A periodic waveform with period T

The waveform exactly repeats after the period (shown as 'T'), but also after 2T, 3T, and so on. However the period is defined as the smallest such period, which is unambiguously just T.

A.3 The frequency of a signal

The frequency of a signal is the inverse of the period, in this case 1/T. Or in other words, it's the number of periods in one second. It's usually measured in Hertz (usually written as 'Hz') which is just another way of saying s^{-1}.

However you should also know that frequency can also be measured in radians per second, and that a frequency in radians per second (sometimes called the *angular frequency*) is defined as 2π/T, so the frequency in radians is 2π times the frequency in Hz[2].

This unit derives from considering a point moving in a circle around the origin. If the point returns to where it had started after *T* seconds, then it could be said to have a period of *T* seconds and hence a frequency of 1/*T* Hz. However it would be moving through an angle of 2π radians in every period (there are 2π radians in a complete revolution), so the rate of change of the angle in radians per second is 2π / *T*. This is termed the *angular frequency*.

The fact that there are two common units used to express frequency means that it is always very important to specify which units you are using when you quote a frequency. It can make quite a difference: for example, a frequency of 10 Hz is the same thing as 62.8 radians/sec, and 10 radians/sec is 1.59 Hz.

To try and minimise the confusion, I always use f to represent a frequency in hertz, and ω (a lower-case Greek omega) to represent a frequency in radians per second.

A.4 The phase of a signal

As hinted above, there is really no such thing as the phase of a signal, in the sense that if all you see on an oscilloscope screen is one oscillating waveform, it is impossible to say what its phase is.

While it's impossible to talk about the phase of a signal, it is possible (and often very useful) to talk about the phase difference between two signals.

[2] There are very good reasons why frequencies are sometimes specified in radians/sec; it makes some mathematical operations on frequencies much easier. You'll come across examples of this later when we discuss phasors.

Despite this, you'll often see equations such as:

$$y_1(t) = \cos(\omega t + \theta) \tag{A.1}$$

with the quantity θ defined as the phase of the signal. This is just engineer's shorthand for saying that this signal has a phase difference of θ compared to a simple cosine wave which has its peak value of one at a time chosen to be $t = 0$:

$$y_2(t) = \cos(\omega t) \tag{A.2}$$

As also noted in the definition above, the phase difference between two signals is determined by the ratio of the offset between the two signals to the period. So for example, the phase difference between the two waveforms shown in Figure A-2 below is:

$$\begin{aligned}\theta &= 360\left(\frac{t}{T}\right) \text{ degrees} \\ &= 2\pi\left(\frac{t}{T}\right) \text{ radians}\end{aligned} \tag{A.3}$$

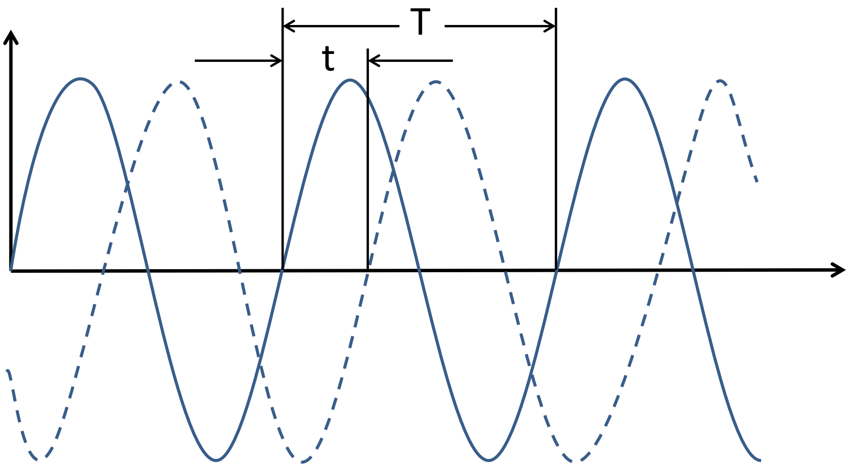

Figure A-2 Phase difference between two sinusoidal waveforms

The most important points about the concept of phase here are:

- The frequencies of the two waveforms being compared must be the same. This is essential: you cannot define the phase difference between two signals which have different frequencies.
- You can talk about positive and negative phase differences, sometimes called *lags* and *leads*.
    - In the diagram above, the dotted waveform lags the solid waveform (since it goes through zero and reaches its maximum value after the solid waveform)
    - Alternatively you could say that the solid waveform leads the dotted waveform (since it goes though zero and reaches its maximum value before the dotted waveform)
- There is no difference between a phase lead of 180 degrees and a phase lag of 180 degrees. In both cases one signal is an upside-down version of the other one.

- There is no difference between a phase lead of θ degrees and a phase lag of (360 – θ) degrees: they are the same thing.
  - This is equivalent to noting that the phase difference between the two waveforms above could equally well have been drawn as shown below, with the dotted waveform leading the solid waveform by *T* - *t* seconds.

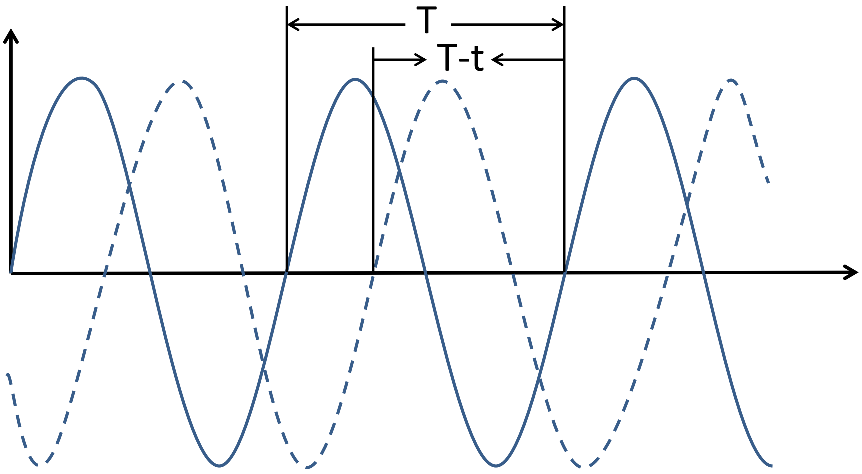

Figure A-3 Alternate way to describe the phase difference

A.5 Sinusoidal waveforms

You might have noticed that every waveform described in this note so far has had the same shape. This shape is known as a sinusoid, and is a general case of a sine or cosine wave. (A sine wave is just a cosine wave shifted through phase by 90 degrees, since:

$$\begin{aligned}\cos(\theta - 90°) &= \cos\theta\cos 90° + \sin\theta\sin 90° \\ &= \sin\theta\end{aligned} \tag{A.4}$$

using the fact that $\sin 90° = 1$ and $\cos 90° = 0$.)

This is not an arbitrary choice. As mentioned above, the frequency content of any periodic signal is often expressed in terms of the set of amplitudes and frequencies of the sine and cosine waves that when added together give the signal of interest. Sine and cosine waves are given this special status because they are particularly easy to use when doing calculus: you can differentiate and integrate sine and cosine waves as much as you like, and you always get other sine and cosine waves:

$$\begin{aligned} &y(t) = \sin(\omega t) && \frac{dy}{dt} = \omega\cos(\omega t) \\ &\frac{d^2y}{dt^2} = -\omega^2\sin(\omega t) && \frac{d^3y}{dt^3} = -\omega^4\cos(\omega t) \\ &\int y\,dt = \frac{-1}{\omega}\cos(\omega t) && \int\left(\int y\,dt\right)dt = \frac{-1}{\omega^2}\sin(\omega t)\end{aligned} \tag{A.5}$$

and so on.

(Note that the above only works if the angles are expressed in radians, and the frequencies in radians per second. If the angles were expressed in degrees, and the frequencies in Hz, then we'd have to write:

$$y(t)=\sin(360ft) \qquad \frac{dy}{dt}=\frac{f}{2\pi}\cos(360ft)$$
$$\frac{d^2y}{dt^2}=-\left(\frac{f}{2\pi}\right)^2\sin(360ft) \qquad \frac{d^3y}{dt^3}=-\left(\frac{f}{2\pi}\right)^3\cos(360ft) \tag{A.6}$$

which rapidly gets confusing. If you ever have to differentiate or integrate a sine-wave, it's much easier to work in radians and angular frequencies.)

A.6 Frequency content

Back to that rather extraordinary claim made earlier: it is possible to make any real[3] periodic waveform by adding up a series of sines and cosines with different frequencies and amplitudes. I'll have to leave the proof of this statement to a later module, but for now I can give an example.

Consider the waveform shown below:

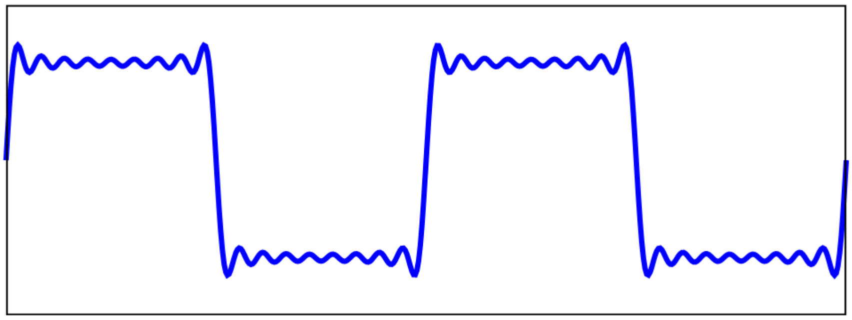

Figure A-4 An approximation to a square-wave made from sine waves

This has been constructed by adding up a series of sine waves with frequencies of 10, 30, 50, 70, ..., 150 Hz, with decreasing amplitudes. As you can see, it almost looks like a square wave. The individual sine waves that combine to make this waveform look like this:

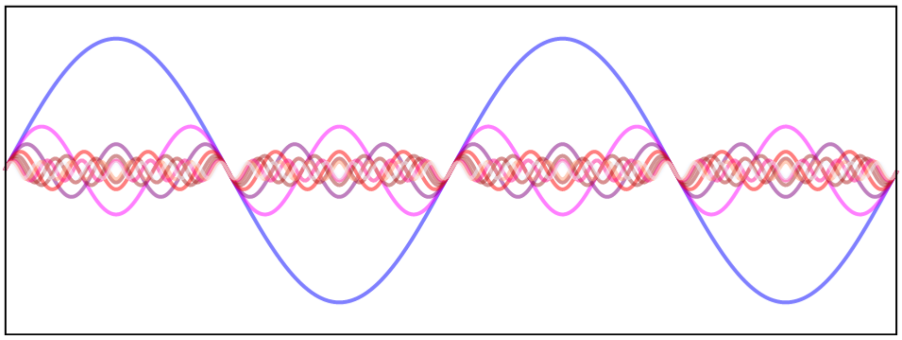

Figure A-5 Individual waveforms that together make up the approximate square wave

---

[3] "real" in the sense of a waveform that might exist in a real circuit in real life. There are some pathological mathematical waveforms which cannot be expressed as the sum of sines and cosines, but they are impossible to produce in reality.

A.7 The amplitudes of a signal

That's not a typo, I do mean "amplitudes" in the plural. As noted before, the amplitude of a signal can be expressed in several different terms, including the zero-to-peak amplitude, the peak-to-peak amplitude, and the rms amplitude. It's important to specify which one you are using, since just saying "the amplitude is 2 volts" is ambiguous: your audience won't know which amplitude you're talking about.

The most obvious way to measure the amplitude (size) of a sine-wave is using the zero-to-peak or peak-to-peak amplitudes. The zero-to-peak is the most intuitive, since in the general formula for a sine-wave:

$$x(t) = A\cos(\omega t + \theta) \qquad \text{(A.7)}$$

the term *A* is the zero-to-peak amplitude.

However peak-to-peak amplitudes are more common for measurements of sine waves, and this is because they do not suffer from any problems due to dc offsets in the measured waveform. For example, consider the waveform shown below:

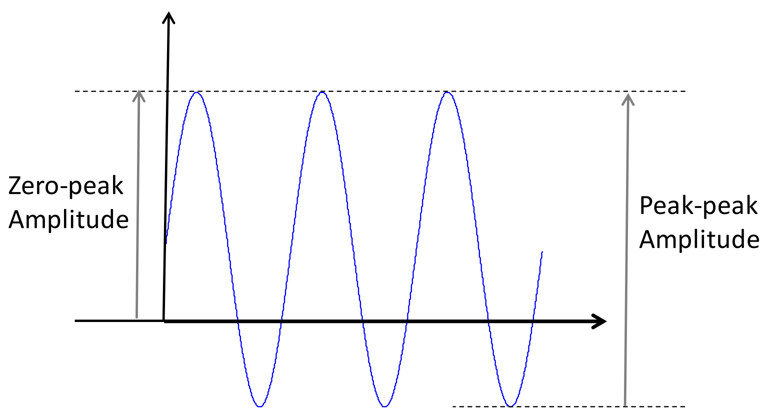

Figure A-6 A sine-wave with a non-zero offset

Here a measurement of the zero-peak amplitude would not give half of the peak-to-peak amplitude. (The effect gets even more dramatic if the dc offset is greater than half the amplitude, in which case the minimum value of the oscillation is above zero, and the zero-to-peak amplitude is actually greater than the peak-to-peak amplitude.) In these cases the peak-to-peak amplitude more accurately represents how much of the sine wave is there, which is what you usually want to know.

However, most signals are not pure sine waves. For other types of signals, the peak-to-peak amplitude is often not the most important parameter. For example, for many types of noise, the actual maximum value can be extremely large, but is so rarely achieved that the peak-to-peak amplitude is not a useful measure of the size of the signal.

For these signals, a much more useful parameter is to define the size of the signal in terms of the average power contained within it. The power in a signal varies with the square of the voltage, so a useful measure of the power in a signal is the mean (average) value of the square of the signal voltage[4].

Of course, the square of a signal is measured in volts-squared, so it isn't really an amplitude. To get a quantity that is measured in volts, we can simply take the square root of the mean value of the square of the signal. This is known as the root-mean-square (rms) value of the signal, and it is measured in volts.

For example: consider a sine-wave, a square-wave and a noise waveform, as shown below:

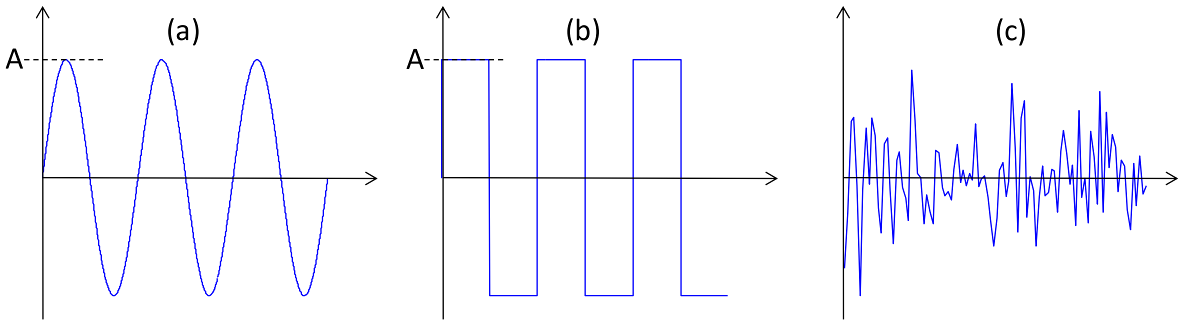

Figure A-7 Sine, square and noise waveforms

Taking these cases in turn:

A.7.1 A simple sine wave

In Figure A-7 (a), the sine wave has a zero-to-peak amplitude of *A*, and a peak-to-peak amplitude of 2*A*. The mean value of the square of the waveform averaged over one period can most easily[5] be determined by considering the well-known formula:

$$\sin^2(t) = \frac{1-\cos(2t)}{2} = \frac{1}{2} - \frac{\cos(2t)}{2} \tag{A.8}$$

The mean value of the right-hand-side is just 0.5, since the mean value of cos(2t) is zero (the cosine waveform spends as much time positive as it does negative).

Hence, the average value of the square of a sine wave of zero-to-peak amplitude *A* is:

---

[4] This is sometimes called the power in the signal, but of course in a real circuit where the signal represents a real voltage across a resistor, the power is given by $P = V^2 / R$, so the mean power dissipated in the resistor is the square of the rms voltage divided by the resistance.

[5] Although perhaps not most rigorously. For those who know calculus, the mean-square value can be derived as follows (taking the average over one cycle from zero to 2π:

$$\begin{aligned} mean\_square &= \frac{1}{2\pi}\int_0^{2\pi}\left(A\sin(t)\right)^2 dt = \frac{A^2}{2\pi}\int_0^{2\pi}\frac{1-\cos(2t)}{2}dt \\ &= \frac{A^2}{2\pi}\int_0^{2\pi}\frac{1}{2}dt + \frac{A^2}{2\pi}\int_0^{2\pi}\frac{\cos(2t)}{2}dt \\ &= \frac{A^2}{2\pi}\left[\frac{t}{2}\right]_0^{2\pi} = \frac{A^2\pi}{2\pi} = \frac{A^2}{2} \end{aligned}$$

$$\overline{\left(A\sin(t)\right)^2} = A^2\overline{\left(\frac{1-\cos(2t)}{2}\right)} = \overline{\frac{A^2}{2}} - \overline{\frac{A^2\cos(2t)}{2}} = \overline{\frac{A^2}{2}} \tag{A.9}$$

Therefore the root-mean-square value (rms) is $A/\sqrt{2}$.

This is a very useful fact to know: for a sine wave with zero mean, the rms value is the zero-to-peak amplitude divided by $\sqrt{2}$ (or the peak-to-peak amplitude divided by $2\sqrt{2}$).

For example, a sine wave with an rms amplitude of 10 mV has a peak-to-peak amplitude of $10\text{m}\times 2\sqrt{2} = 28.3$ mV, and a sine wave with a peak-to-peak value of 10 mV has an rms value of $\frac{10\text{m}}{2\sqrt{2}} = 3.54$ mV.

A.7.2 A simple square wave

In Figure A-7 (b), the square wave also has a peak-to-peak amplitude of 2*A*. It's very easy to work out the rms value in this case, since the square of the value is always A^2. (The amplitude is always either *A* or –*A*, and both, when squared, give A^2.)

That makes the mean square value A^2, and the rms value is just *A*.

A.7.3 A random noise waveform

In Figure A-7 (c), the waveform is a small sample of a very common type of noise known as Gaussian noise. In theory, Gaussian noise has an infinite peak-to-peak amplitude, although it almost never gets to very large values. Waveforms like this are always classified in terms of their rms amplitude, since the peak-to-peak amplitude isn't very useful.

Provided the noise has a zero mean value (which most noise does), the rms value of this noise is the standard deviation of the noise signal (if you've not come across the concept of *standard deviation* yet don't worry about this).

A.7.4 Back to the offset sine-wave

What if the sine-wave has a non-zero mean (as shown in Figure A-6)?

We can determine the rms average using the same technique: square the waveform, take the average, and then take the square root.

If this sine-wave has a peak-to-peak amplitude of 2A and mean value of B, then the equation of the sine-wave could be written as:

$$y(t) = B + A\cos(t) \tag{A.10}$$

Taking the square of this gives:

$$\begin{aligned} y^2(t) &= \left(B + A\cos(t)\right)^2 \\ &= B^2 + 2AB\cos(t) + A^2\cos^2(t) \\ &= B^2 + 2AB\cos(t) + \frac{A^2}{2} - \frac{A^2\cos(2t)}{2} \end{aligned} \tag{A.11}$$

so here the mean square value is given by:

$$\begin{aligned}\overline{y^2(t)} &= \overline{B^2} + \overline{2AB\cos(t)} + \frac{\overline{A^2}}{2} - \frac{\overline{A^2\cos(2t)}}{2} \\ &= \overline{B^2} + \frac{\overline{A^2}}{2}\end{aligned} \tag{A.12}$$

since the other two terms are oscillations and have a mean value of zero. That makes the rms value:

$$\sqrt{\overline{y^2(t)}} = \sqrt{B^2 + \frac{A^2}{2}} \tag{A.13}$$

The technique works for any waveform: to find the rms value, square the waveform, then take the average value of the square, then take the (positive) square root of the mean value of the square.