

Notes on Fourier Analysis

The basics and beyond

by Dr. Colin Mercer, Technical Director, Prosig

Fourier analysis takes a signal and represents it either as a series of cosines (real part) and sines (imaginary part) or as a cosine with phase (modulus and phase form). As an illustration we will look at Fourier analysing the sum of the two sine waves shown below. The resultant summed signal is shown in the third graph.

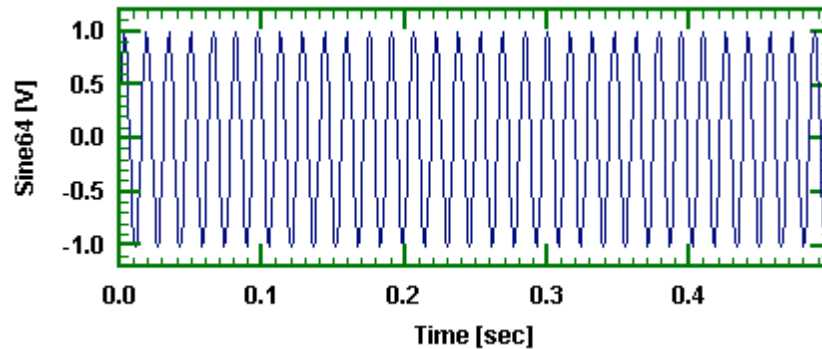


Figure 1 Sine wave, 64Hz, unit amplitude, zero phase

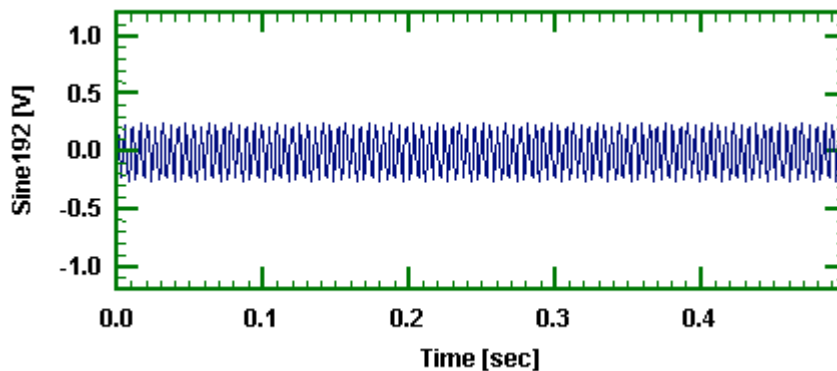


Figure 2 Sine wave, 192Hz, 0.25 amplitude, 30° phase

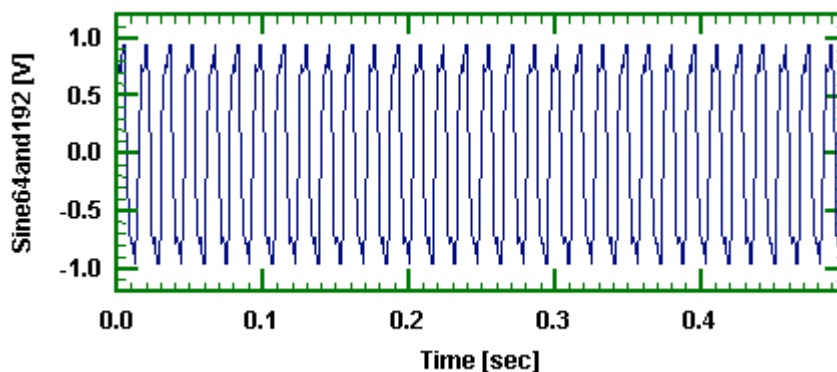


Figure 3 Combined Sine waves

If we now carry out a Fourier Analysis, in this case with an FFT, of the combined signal then we obtain the following result.

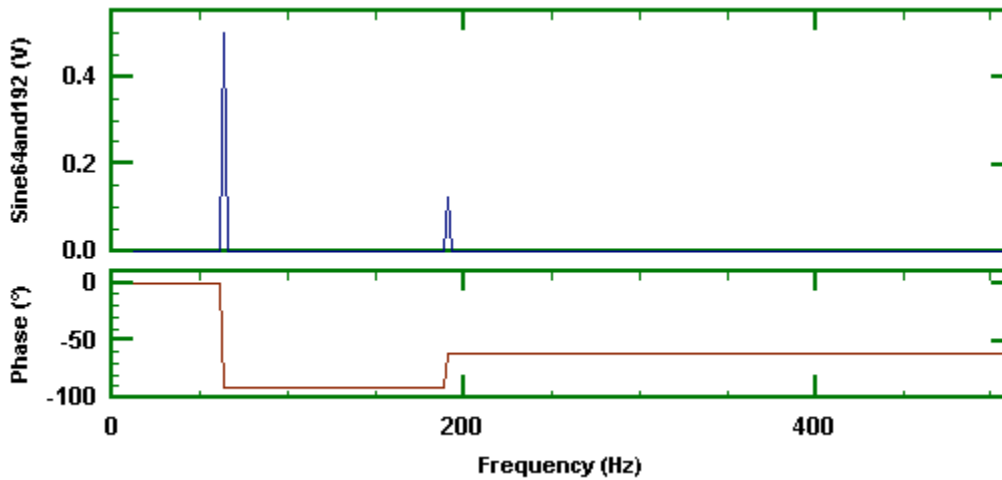


Figure 4. FFT of 64Hz & 192Hz signals

We see immediately that there are two distinct peaks in the modulus curve and two distinct changes in the phase curve at 64Hz and at 192Hz as expected.

The amplitude shown is exactly half of the original constituent sine waves. That is the sine wave of unity amplitude at 64Hz is shown as 0.5 and the sine wave of amplitude 0.25 is shown as 0.125. Why is this? The reason is that when we do a frequency analysis of a signal some of the 'energy' is represented for positive frequencies and half for the negative frequencies. For a real time signal, as opposed to a complex time signal, then this energy is split equally and we get exactly half. Some software packages do a doubling to overcome this but this is not done in DATS. This is to make so called half range analysis compatible with full range analyses.

Sine Wave Amplitude	Peak to Peak Value	FFT or DFT Value
A	2A	A/2

Table 1. Amplitude Relationship

Now consider the phase part. The original 64Hz sine had a zero degree phase and the 192Hz had a 30° phase. From the phase plot at 64Hz the phase jumps from 0° to -90°. Why? This is because Fourier analysis uses cosines and sines. It is cosines, not the sines, which are the basic reference. Because a sine wave is a -90° phase shifted cosine then that is what we get. The phase shift at 192Hz was not 30° but -60°. This is totally correct as we have (-90+30) = -60°. Further explanation of this is given in the slightly more mathematical part at the end of these notes.

In the above examples the signals were represented by 512 points at 1024 samples/second. That is we had 0.5 seconds of data. Hence when we using an FFT to carry out the Fourier analysis then the separation between frequency points is 2Hz. This is a fundamental relationship. If the length of the data frequency analysed is T seconds then the frequency spacing given by an FFT is (1/T)Hz.

FFT Frequency Spacing in Hz = 1/Analysed Signal Duration in Seconds

Selecting the FFT size, N, will dictate the effective duration of the signal being analysed. If we were to choose an FFT size of say 256 points with a 1024 points/second sample rate then we would use 1/4 seconds of data and the frequency spacing would be 4Hz.

As we are dealing with the engineering analysis of signals measuring physical events it is clearly more sensible to ensure we can set our frequency spacing rather than the arbitrary

choice of some FFT size which is not physically related to the problem in hand. That is DATS uses the natural default of physically meaningful quantities. However it is necessary to note that some people have become accustomed to specifying "block size". To accommodate this DATS includes an FFT module shown as FFT (Select) on the frequency analysis pull down menu. This module does allow a choice of block size.

"Non Exact" Frequencies

In the above examples the frequency of the sine waves were exact multiples of the frequency spacing. They were specifically chosen that way. As noted earlier 0.5 seconds of data gives a frequency spacing of exactly 2Hz. Now, suppose we have a sine wave like the original 64Hz sine wave but at a frequency of 63Hz. This frequency is not an exact multiple of the frequency spacing. What happens? Visually it is very difficult to see any difference in the time domain but there is a distinct difference in the Fourier results. The graph below shows an expanded version of the result of an FFT of unit amplitude, zero phase, 63Hz sine wave.

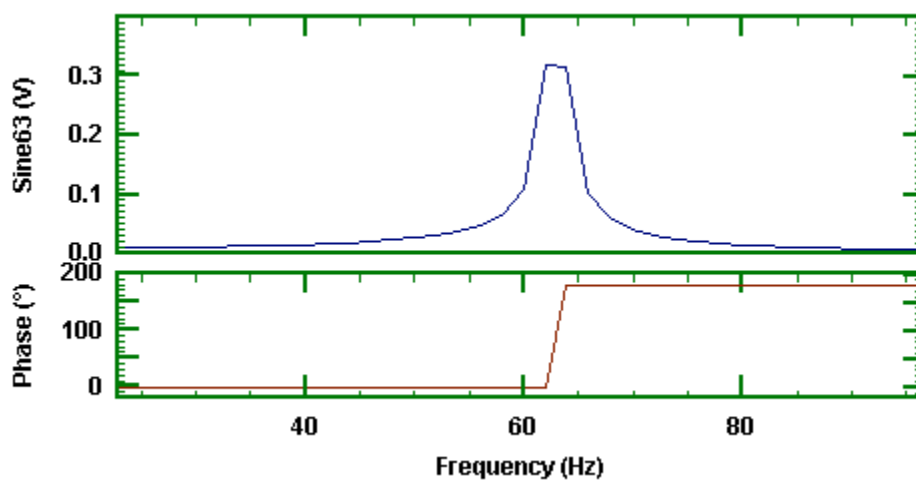


Figure 5 FFT of 63Hz, a "non-exact" frequency

Note that there is not a single spike but rather a 'spike' with the top cut off. The values at 62Hz and 64Hz are almost identical, but they are not 0.5, rather they are approximately 0.32. Furthermore the phase at 62Hz is 0° and at 64Hz it is 180°. That is the Fourier analysis is telling us we have a signal composed of multiple sine waves, the two principle ones being at 62 and 64Hz with half amplitudes of 0.32 and a phase of 0° and 180° respectively. In reality we know we had a sine wave at 63Hz.

If we overlay the modulus results at 63Hz and 64Hz then we note that the 63Hz curve is quite different in characteristic to the 64Hz curve.

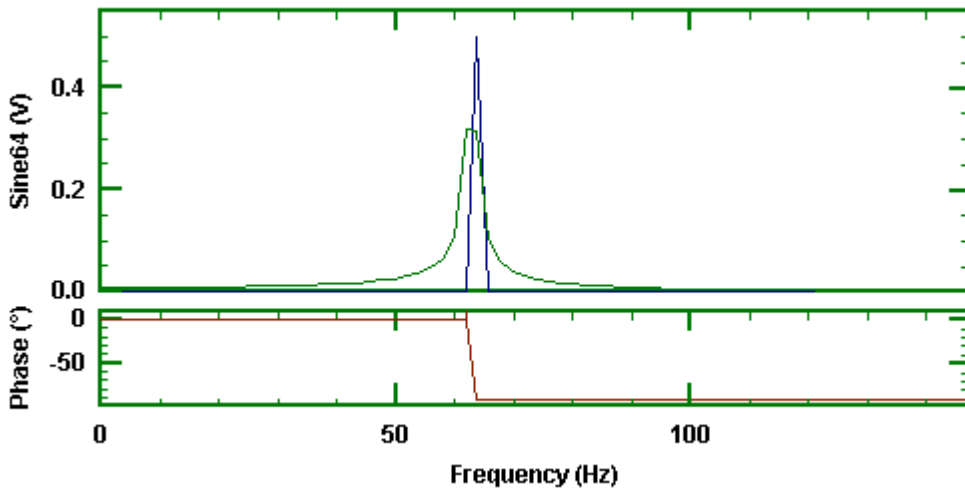


Figure 6. Overlay FFT of 63Hz and 64Hz signals

This shows that care needs to be taken when interpreting FFT results of analysing sine waves as the value shown will depend upon the relationship between the actual frequency of the signal and the “measurement” frequencies. Although the amplitudes vary significantly between these two cases if one compares the RMS value by using Spectrum RMS over Frequency Range then the 64Hz signal gives 0.707107 and the 63Hz signal gives 0.704936.

The above results were obtained using an FFT algorithm. With the FFT the frequency spacing is a function of the signal length. Now given the speed of the modern PC then we may also use an original Direct Fourier Transform method. In particular the DFT (Basic Mod Phase) version in Frequency Analysis (Advanced) allows a choice of start frequency, end frequency and frequency spacing. The DFT is much slower than the FFT. Choosing to analyse from 40Hz to 80Hz in 0.1Hz steps gives the results shown below with the continuous curve. The * marks are those points from the corresponding FFT analysis.

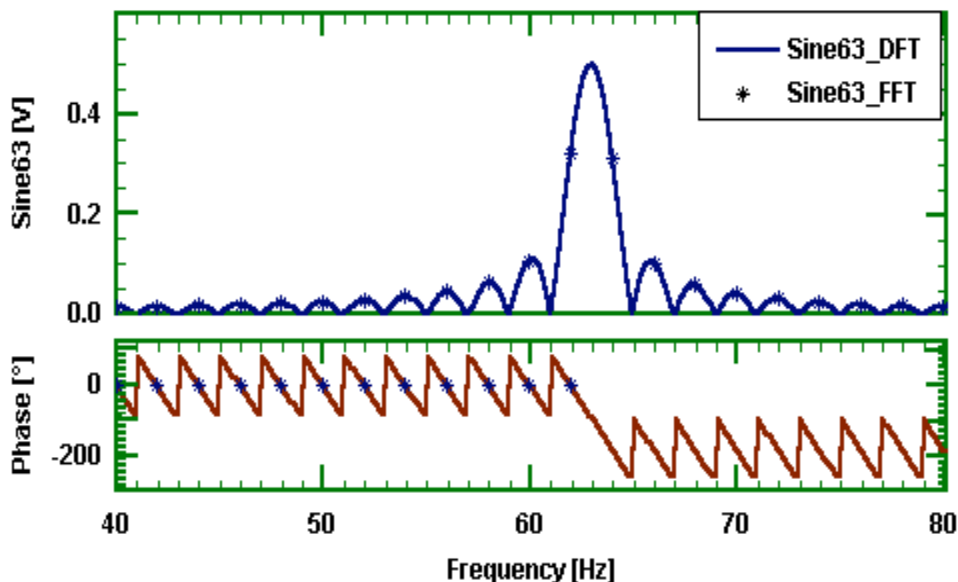


Figure 7. DFT analysis of 63Hz Sine wave

This now shows the main lobe of the response. The peak value is 0.5 at 63Hz and the phase is -90°. Also from 62Hz to 64Hz the phase goes from 0° to -180°. Note that this amount of phase change from one “Exact” frequency to the adjacent one is typical.

The above plot shows all the “side lobes” and illustrates another aspect of digital signal processing, namely the phenomenon known as spectral leakage. That is in principle the energy at one frequency “leaks” to every other frequency. This leakage may be reduced by a suitable choice of data window. The shape of the curve in Figure 7 is actually that of the so-called “spectral window” through which we are looking at the data. It is often better to think of this as the shape of the effective analysis filter. In this example the data window used is a Bartlet (rectangular) type. Details of different data windows and their corresponding spectral window are discussed in a separate article.

In this note we have been careful to use “frequency spacing” rather than “frequency resolution”. It is clear that with DFT and other techniques we can change the frequency spacing. For an FFT method the spacing is related to the “block size”. But what is the frequency resolution? This is a large subject but we will give the essence. The clue is the shape of the spectral window as illustrated in Figure 7. A working definition of frequency resolution is the ability to separate two close frequency responses. Another common definition is the half power (-3dB) points of the spectral window. In practice the most useful definition is a frequency bandwidth known as the Equivalent Noise Band Width (ENBW). This is very similar to the half power points definition. ENBW is determined entirely by the shape of the data window used and the duration of the data used in the FFT processing.

Signal Duration Effects

If we have data taken over a longer period then the frequency spacing will be narrower. In many cases this will assist the problem but if there is no exact match the same phenomenon will arise.

Fourier analysis tells us the amplitude and phase of that set of cosines which have the same duration as the original signal. Suppose now we take a signal which again is composed of unit amplitude 64Hz sine wave and a 0.25 amplitude 192Hz sine wave signals but this time the 64Hz signal occupies the first half and the 192Hz signal occupies the second half. That is we now have a one second signal as shown below.

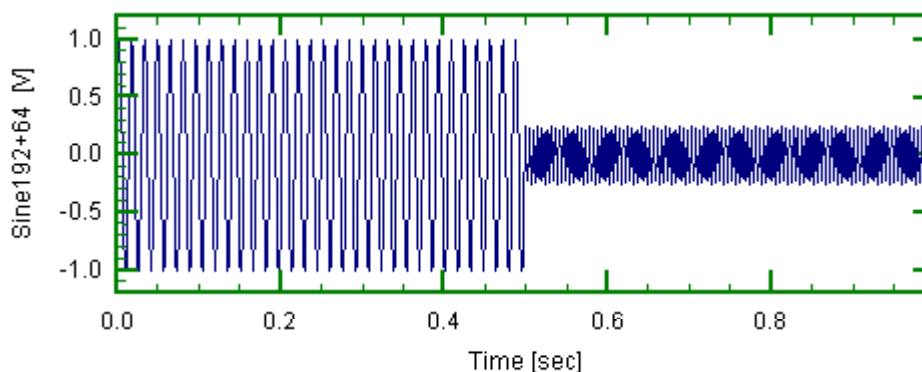


Figure 8. Two sines joined

The result of an FFT of these two joined signals is shown below.

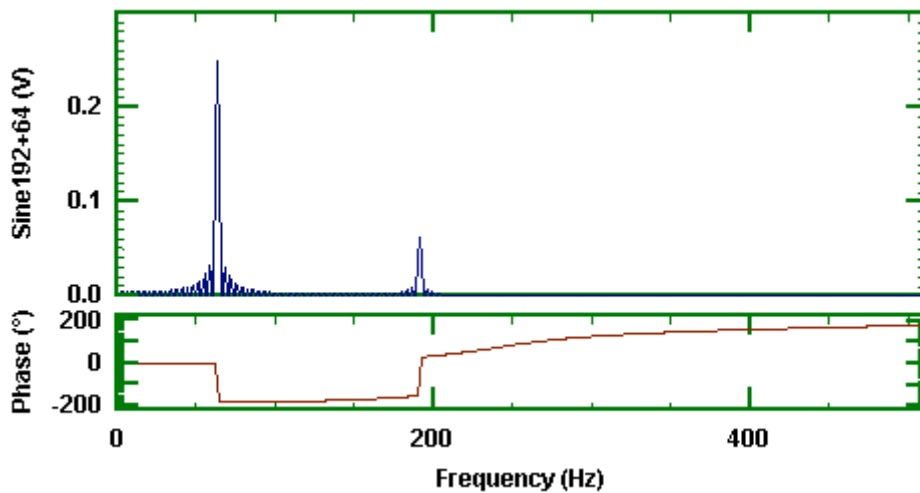


Figure 9 FFT of two joined sinewaves

There are as expected significant frequencies at 64Hz and 192Hz. However the half amplitudes are now 0.25 (instead of 0.5) and 0.0625 (instead of 0.125). One interpretation of what the FFT is telling us is that there is a cosine wave at 64Hz of half amplitude 0.25 for the whole one second duration and another one of half amplitude 0.0625 for the whole duration. But we know that we had a 64Hz signal with a half amplitude of 0.5 for the first part of the time and a 192Hz signal with a half amplitude of 0.125 for the second part. What is happening?

A closer look at the spectrum around 64Hz as shown below reveals that we have a large number of frequencies around 64Hz. This time they are 1Hz apart as we had one second of data. Their relative amplitudes and phases combine to double the amplitude at 64Hz over the first part and to cancel during the second part. The same of course happens in reverse around those frequencies close to 192Hz.

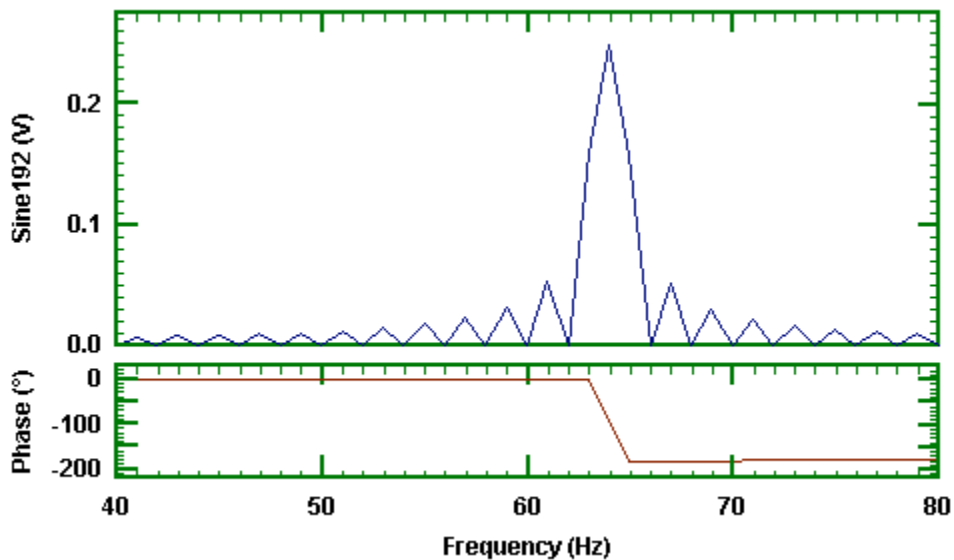


Figure 10. FFT (part) of joined signals

Another example is where a signal is extended by zeroes. Again the amplitude is reduced. In this case the reduction is proportional to the percentage extension by zeroes.

The important point to note is that the Fourier analysis assumes that the sines and cosines last for the entire duration.

Swept Sine Signal

With a swept sine signal theoretically each frequency only lasts for an instant in time. A swept sine signal sweeping from 10Hz to 100Hz is shown below.

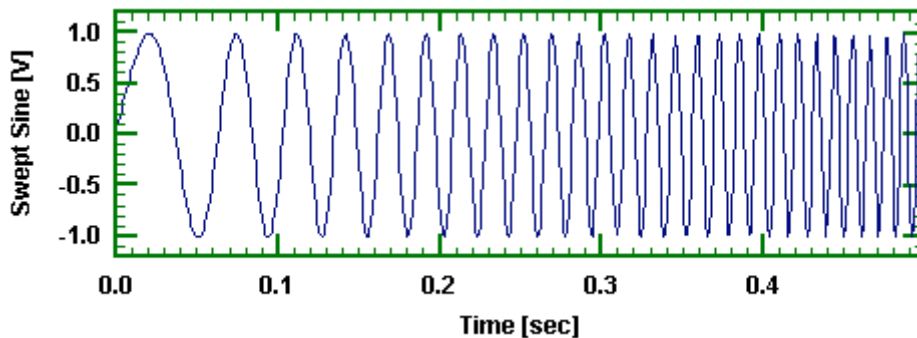


Figure 11 Swept sine, unit amplitude, 10 to 100Hz.

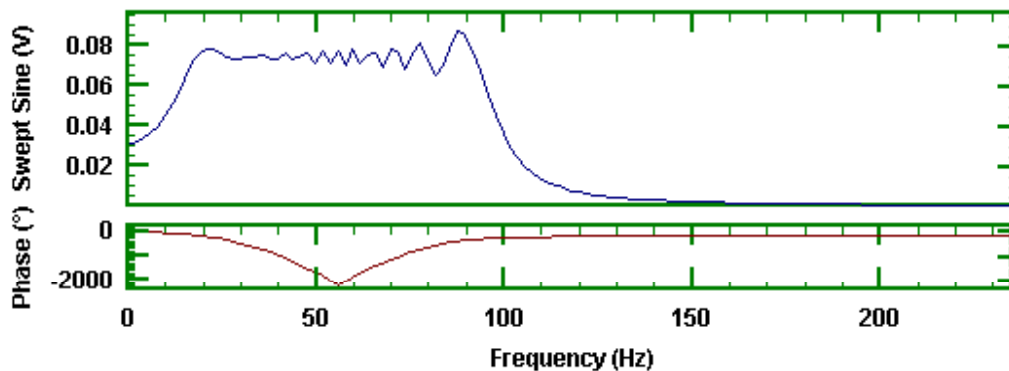


Figure 12 FFT of Swept Sine

This has 512 points at 1024 samples/second. Thus the sweep rate was 180Hz/second. The FFT of that signal shows an amplitude of about 0.075. Over the duration of the sweep the phase goes from around zero to -2000° and then settles to -180° above 100Hz. If the sweep rate is lowered to around 10Hz/second then the amplitude becomes about 0.019. The relationship between the spectrum level the amplitude and sweep rate of the original swept sine is not straight forward.

It is clear that one has to interpret a simple Fourier analysis, whether it is done by an FFT or by a DFT, with some care. A Fourier analysis shows the (half) amplitudes and phases of the constituent cosine waves that exist for the whole duration of that part of the signal that has been analysed. Although we have not discussed it, a Fourier analysed signal is invertible. That is if we have the Fourier analysis over the entire frequency range from zero to half sample rate then we may do an inverse Fourier transform to get back to the time signal. One point that arises from this is that if the signal being analysed has some random noise in it, then so does the Fourier transformed signal. Fourier analysis by itself does nothing to remove or minimise the effects of noise. Thus simple Fourier analysis is not suitable for random data, but it is for signals such as transients and complicated or simple periodic signals such as those generated by an engine running at a constant speed.

We have not considered Auto Spectral Density (also sometimes called Power Spectral Density) or RMS Spectrum Level Analyses here. They are discussed in another article. However for completeness it is worth noting that the essential difference between ASD analysis and FFT analysis is that ASDs are describing the distribution in frequency of the 'power' in the signal whilst Fourier analysis is determining (half) amplitudes and phases. While ASDs and RMS Spectrum Level analyses do reduce the effects of any randomness, Fourier analysis does not. Where confusion occurs is that both analysis methods may use FFT algorithms. This is not to do with the objective of the analysis or its properties but rather

with efficiency of implementation. After all every analysis will use addition. That is just a mathematical operation and so, in that sense, is the use of an FFT.

A Little Mathematics

We will not go into all the mathematical niceties except to see that a Fourier series could be written in the forms below. In real and imaginary terms we have

$$x(t_k) = \frac{a_o}{2} + \frac{1}{N} \sum_{n=1}^{N-1} (a_n \cos 2\pi f_n t_k + b_n \sin 2\pi f_n t_k)$$

and in modulus and phase form as

$$x(t_k) = \frac{1}{N} \sum_{n=0}^{N-1} \frac{C_n}{2} \cos 2\pi (f_n t_k + \theta_n / 360)$$

The above forms are a slightly unusual way of expressing the Fourier expansion. For instance θ is in degrees. More significantly the product $f_n t_k$ is shown explicitly. Usually in an FFT then f_n is expressed as $n/N\Delta t$ and t_k as $k\Delta t$ where Δt is the time between samples. This gives the relationship of the form

$$x_k = \frac{1}{N} \sum_{n=0}^{N-1} (a_n \cos 2\pi nk / N + b_n \sin 2\pi nk / N)$$

However the point of using $f_n t_k$ explicitly above is to indicate that nothing in the Fourier expansion inhibits the choice of actual frequency at which we evaluate the Fourier coefficients. The FFT gains speed by being selective about where it evaluates the coefficients and also restrictive in the values of N that are permitted. There are ways around these but in most implementations, for practical purposes N is restricted to being a power of 2.

This means that with a DFT we can actually evaluate the Fourier coefficients at any frequency provided we obey the anti aliasing (Nyquist) criterion. The DFT is slower than an FFT. Another way of getting at the finer detail and still getting some speed advantage is to use the so-called Zoom FFT based on the Chirp-z transform. Again the relative advantages are discussed elsewhere.

As a historical note it is perhaps interesting to recall that Fourier did not generate his series in order to carry out frequency analysis but rather to determine a least squares error approximation to a function.

C.A.Mercer
29 July 2003



Dr Colin Mercer

Colin Mercer is the Technical Director of Prosig and has prime responsibility for signal processing and its applications. He was formerly at the Institute of Sound and Vibration research (ISVR), Southampton University where he founded the Data Analysis Centre. He is a Chartered Engineer and a Fellow of the British Computer Society.



Prosig

Prosig provide data acquisition, noise & vibration, acoustic, health & condition monitoring and refinement solutions for the scientific and engineering communities and offer a number of standard hardware and software products. Take a look at www.prosig.com to discover why major automotive, military, aerospace, power and industrial companies rely on Prosig for their complete data acquisition, signal analysis and reporting software & hardware.