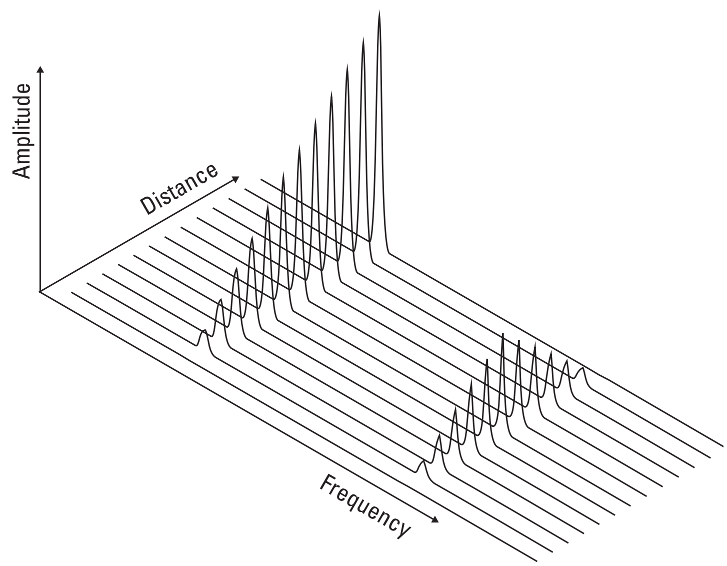
Keysight Technologies
The Fundamentals of Signal Analysis

Application Note



Chapter 1
Introduction

The analysis of electrical signals is a fundamental problem for many engineers and scientists. Even if the immediate problem is not electrical, the basic parameters of interest are often changed into electrical signals by means of transducers. Common transducers include accelerometers and load cells in mechanical work, EEG electrodes and blood pressure probes in biology and medicine, and pH and conductivity probes in chemistry. The rewards for transforming physical parameters to electrical signals are great, as many instruments are available for the analysis of electrical signals in the time, frequency and modal domains. The powerful measurement and analysis capabilities of these instruments can lead to rapid understanding of the system under study.

This note is a primer for those who are unfamiliar with the advantages of analysis in the frequency and modal domains and with the class of analyzers we call Dynamic Signal Analyzers. In Chapter 2 we develop the concepts of the time, frequency and modal domains and show why these different ways of looking at a problem often lend their own unique insights. We then introduce classes of instrumentation available for analysis in these domains.

In Chapter 3 we develop the properties of one of these classes of analyzers, Dynamic Signal Analyzers. These instruments are particularly appropriate for the analysis of signals in the range of a few millihertz to about a hundred kilohertz.

Chapter 4 shows the benefits of Dynamic Signal Analysis in a wide range of measurement situations. The powerful analysis tools of Dynamic Signal Analysis are introduced as needed in each measurement situation.

This note avoids the use of rigorous mathematics and instead depends on heuristic arguments. We have found in over a decade of teaching this material that such arguments lead to a better understanding of the basic processes involved in the various domains and in Dynamic Signal Analysis. Equally important, this heuristic instruction leads to better instrument operators who can intelligently use these analyzers to solve complicated measurement problems with accuracy and ease*.

Because of the tutorial nature of this note, we will not attempt to show detailed solutions for the multitude of measurement problems which can be solved by Dynamic Signal Analysis. Instead, we will concentrate on the features of Dynamic Signal Analysis, how these features are used in a wide range of applications and the benefits to be gained from using Dynamic Signal Analysis.

Those who desire more details on specific applications should look to Appendix B. It contains abstracts of Keysight Technologies, Inc. Application Notes on a wide range of related subjects. These can be obtained free of charge from your local Keysight field engineer or representative.

* A more rigorous mathematical justification for the arguments developed in the main text can be found in Appendix A.

Chapter 2
The Time, Frequency and Modal Domains:

A Matter of Perspective

In this chapter we introduce the concepts of the time, frequency and modal domains. These three ways of looking at a problem are interchangeable; that is, no information is lost in changing from one domain to another. The advantage in introducing these three domains is that of a change of perspective. By changing *perspective* from the time domain, the solution to difficult problems can often become quite clear in the frequency or modal domains.

After developing the concepts of each domain, we will introduce the types of instrumentation available. The merits of each generic instrument type are discussed to give the reader an appreciation of the advantages and disadvantages of each approach.

Section 1:
The Time Domain

The traditional way of observing signals is to view them in the time domain. The time domain is a record of what happened to a parameter of the system versus time. For instance, Figure 2.1 shows a simple spring-mass system where we have attached a pen to the mass and pulled a piece of paper past the pen at a constant rate. The resulting graph is a record of the displacement of the mass versus time, *a time domain view of displacement*.

Such direct recording schemes are sometimes used, but it usually is much more practical to convert the parameter of interest to an electrical signal using a transducer. Transducers are commonly available to change a wide variety of parameters to electrical signals. Microphones, accelerometers, load cells, conductivity and pressure probes are just a few examples.

This electrical signal, which represents a parameter of the system, can be recorded on a strip chart recorder as in Figure 2.2. We can adjust the gain of the system to calibrate our measurement. Then we can reproduce exactly the results of our simple direct recording system in Figure 2.1.

Why should we use this indirect approach? One reason is that we are not always measuring displacement. We then must convert the desired parameter to the displacement of the recorder pen. Usually, the easiest way to do this is through the intermediary of electronics. However, even when measuring displacement we would normally use an indirect approach. Why? Primarily because the system in Figure 2.1 is hopelessly ideal. The mass must be large enough and the spring stiff enough so that the pen's mass and drag on the paper will not

Figure 2.1 Direct recording of displacement - a time domain view.

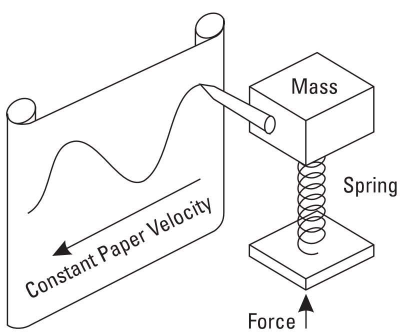
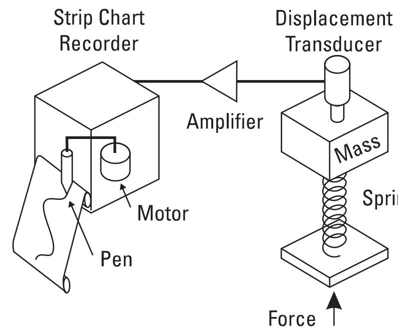


Figure 2.2 Indirect recording of displacement.



affect the results appreciably. Also, the deflection of the mass must be large enough to give a usable result, otherwise a mechanical lever system to amplify the motion would have to be added with its attendant mass and friction.

With the indirect system a transducer can usually be selected which will not significantly affect the measurement. This can go to the extreme of commercially available displacement transducers which do not even contact the mass. The pen deflection can be easily set to any desired value by controlling the gain of the electronic amplifiers.

This indirect system works well until our measured parameter begins to change rapidly. Because of the mass of the pen and recorder mechanism and the power limitations of its drive, the pen can only move at finite velocity. If the measured parameter changes faster, the output of the recorder will be in error. A common way to reduce this problem is to eliminate the pen and record on a photosensitive paper by deflecting a light beam. Such a device is called an *oscillograph*. Since it is only necessary to move a small, light-weight mirror through a very small angle, the oscillograph can respond much faster than a strip chart recorder.

Figure 2.3 Simplified oscillograph operation.

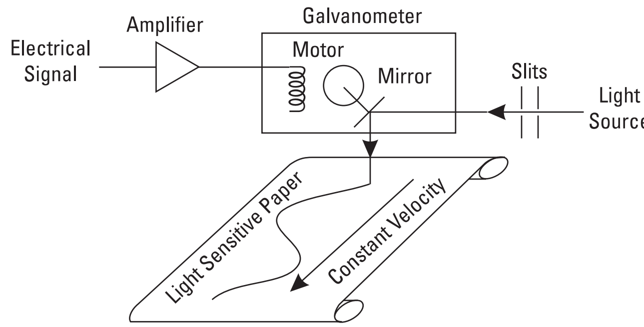
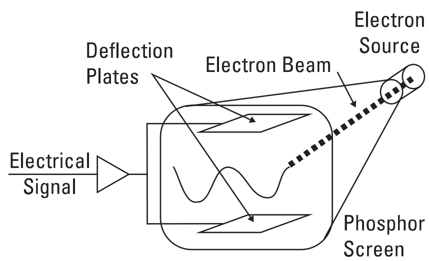


Figure 2.4 Simplified oscilloscope operation (Horizontal deflection circuits omitted for clarity).



Another common device for displaying signals in the time domain is the *oscilloscope*. Here an electron beam is moved using electric fields. The electron beam is made visible by a screen of phosphorescent material. It is capable of accurately displaying signals that vary even more rapidly than the oscillograph can handle. This is because it is only necessary to move an electron beam, not a mirror.

The strip chart, oscillograph and oscilloscope all show displacement versus time. We say that changes in this displacement represent the variation of some parameter versus time. We will now look at another way of representing the variation of a parameter.

Section 2: The Frequency Domain

It was shown over one hundred years ago by Baron Jean Baptiste Fourier that any waveform that exists in the real world can be generated by adding up sine waves. We have illustrated this in Figure 2.5 for a simple waveform composed of two sine waves. By picking the amplitudes, frequencies and phases of these sine waves correctly, we can generate a waveform identical to our desired signal.

Conversely, we can break down our real world signal into these same sine waves. It can be shown that this combination of sine waves is unique; any real world signal can be represented by only one combination of sine waves.

Figure 2.6a is a three dimensional graph of this addition of sine waves. Two of the axes are time and amplitude, familiar from the time domain. The third axis is frequency which allows us to visually separate the sine waves which add to give us our complex waveform. If we view this three-dimensional graph along the frequency axis we get the view in Figure 2.6b. This is the time domain view of the sine waves. Adding them together at each instant of time gives the original waveform.

Figure 2.5
Any real waveform can be produced by adding sine waves together.

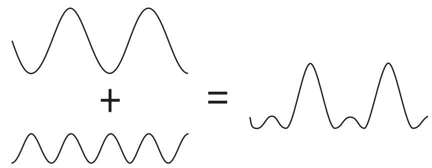
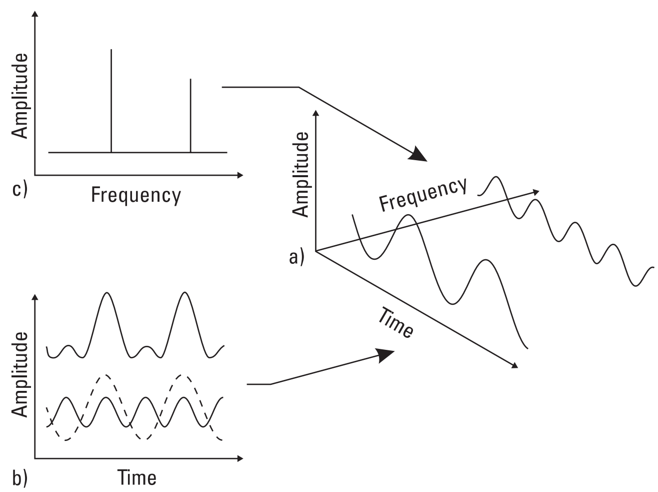


Figure 2.6
The relationship between the time and frequency domains.
a) Three-dimensional coordinates showing time, frequency and amplitude
b) Time domain view
c) Frequency domain view.



However, if we view our graph along the time axis as in Figure 2.6c, we get a totally different picture. Here we have axes of amplitude versus frequency, what is commonly called the frequency domain. Every sine wave we separated from the input appears as a vertical line. Its height represents its amplitude and its position represents its frequency. Since we know that each line represents a sine wave, we have uniquely characterized our input signal in the frequency domain*. This frequency domain representation of our signal is called the *spectrum* of the signal. Each sine wave line of the spectrum is called a *component* of the total signal.

* Actually, we have lost the phase information of the sine waves. How we get this will be discussed in Chapter 3.

Since one of the major uses of the frequency domain is to resolve small signals in the presence of large ones, let us now address the problem of how we can see both large and small signals on our display simultaneously.

Suppose we wish to measure a distortion component that is 0.1% of the signal. If we set the fundamental to full scale on a four inch (10 cm) screen, the harmonic would be only four thousandths of an inch (0.1 mm) tall. Obviously, we could barely see such a signal, much less measure it accurately. Yet many analyzers are available with the ability to measure signals even smaller than this.

Since we want to be able to see all the components easily at the same time, the only answer is to change our amplitude scale. A logarithmic scale would compress our large signal amplitude and expand the small ones, allowing all components to be displayed at the same time.

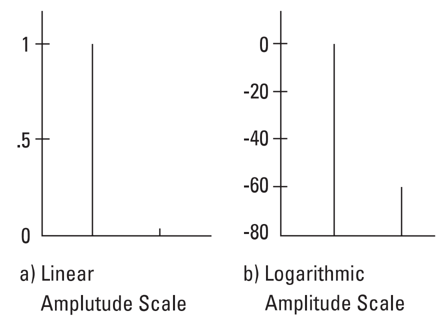
Alexander Graham Bell discovered that the human ear responded logarithmically to power difference and invented a unit, the Bel, to help him measure the ability of people to hear. One tenth of a Bel, the deciBel (dB) is the most common unit used in the frequency domain today. A table of the relationship between volts, power and dB is given in Figure 2.8. From the table we can see that our 0.1% distortion component example is 60 dB below the fundamental. If we had an 80 dB display as in Figure 2.9, the distortion component would occupy 1/4 of the screen, not 1/1000 as in a linear display.

Figure 2.8 The relationship between decibels, power and voltage.

db	Power Ratio	db	Voltage Ratio
+20	100	+40	100
+10	10	+20	10
+ 3	2	+ 6	2
0	1	0	1
- 3	1/2	- 6	1/2
- 10	1/10	- 20	1/10
- 20	1/100	- 40	1/100

db = 10 log (Power Ratio) = 20 log (Voltage Ratio)

Figure 2.9 Small signals can be measured with a logarithmic amplitude scale.



It is very important to understand that *we have neither gained nor lost information, we are just representing it differently.* We are looking at the same three-dimensional graph from different angles. This different perspective can be very useful.

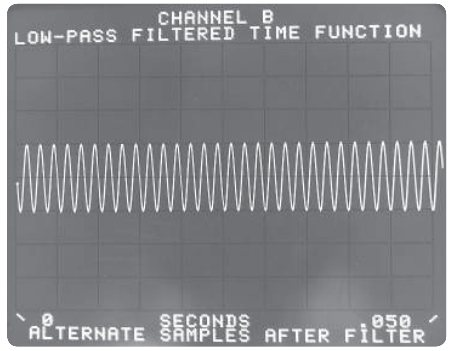
Why the Frequency Domain?

Suppose we wish to measure the level of distortion in an audio oscillator. Or we might be trying to detect the first sounds of a bearing failing on a noisy machine. In each case, we are trying to detect a small sine wave in the presence of large signals. Figure 2.7a shows a time domain waveform which seems to be a single sine wave. But Figure 2.7b shows in the frequency domain that the same signal is composed of a large sine wave and significant other sine wave components (distortion components). When these components are separated in the frequency domain, the small components are easy to see because they are not masked by larger ones.

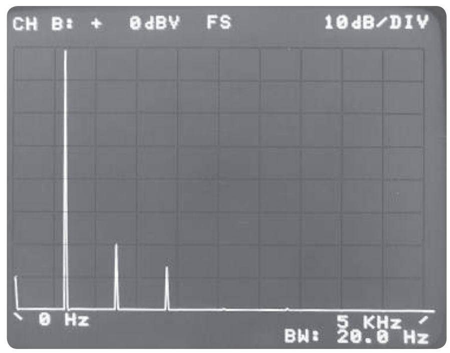
The frequency domain's usefulness is not restricted to electronics or mechanics. All fields of science and engineering have measurements like these where large signals mask others in the time domain. The frequency domain provides a useful tool in analyzing these small but important effects.

Figure 2.7 Small signals are not hidden in the frequency domain.

a) Time Domain - small signal not visible



b) Frequency Domain - small signal easily resolved



The Frequency Domain: A Natural Domain

At first the frequency domain may seem strange and unfamiliar, yet it is an important part of everyday life. Your ear-brain combination is an excellent frequency domain analyzer. The ear-brain splits the audio spectrum into many narrow bands and determines the power present in each band. It can easily pick small sounds out of loud background noise thanks in part to its frequency domain capability. A doctor listens to your heart and breathing for any unusual sounds. He is listening for frequencies which will tell him something is wrong. An experienced mechanic can do the same thing with a machine. Using a screwdriver as a stethoscope, he can hear when a bearing is failing because of the frequencies it produces.

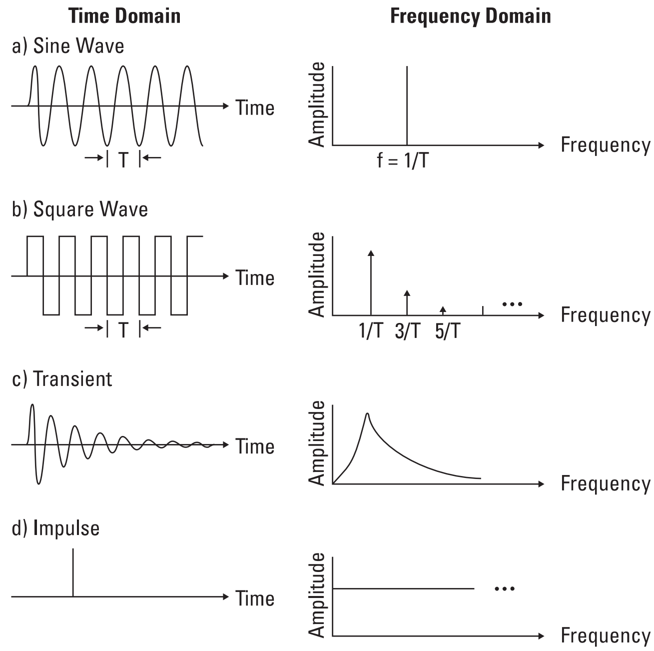
So we see that the frequency domain is not at all uncommon. We are just not used to seeing it in graphical form. But this graphical presentation is really not any stranger than saying that the temperature changed with time like the displacement of a line on a graph.

Spectrum Examples

Let us now look at a few common signals in both the time and frequency domains. In Figure 2.10a, we see that the spectrum of a sine wave is just a single line. We expect this from the way we constructed the frequency domain. The square wave in Figure 2.10b is made up of an infinite number of sine waves, all harmonically related. The lowest frequency present is the reciprocal of the square wave period. These two examples illustrate a property of the frequency transform: a signal which is periodic and exists for all time has a discrete frequency spectrum. This is in contrast to the transient signal in Figure 2.10c which has a continuous spectrum. This means that the sine waves that make up this signal are spaced infinitesimally close together.

Another signal of interest is the impulse shown in Figure 2.10d. The frequency spectrum of an impulse is flat, i.e., there is energy at all frequencies. It would, therefore, require infinite energy to generate a true impulse. Nevertheless, it is possible to generate an approximation to an impulse which has a fairly flat spectrum over the desired frequency range of interest. We will find signals with a flat spectrum useful in our next subject, network analysis.

Figure 2.10 Frequency spectrum examples.



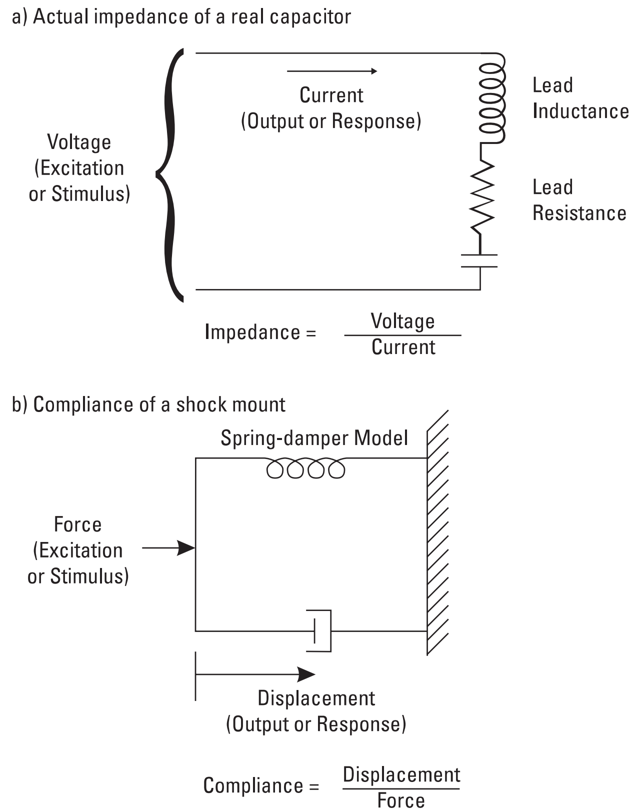
Network Analysis

If the frequency domain were restricted to the analysis of signal spectrums, it would certainly not be such a common engineering tool. However, the frequency domain is also widely used in analyzing the behavior of networks (network analysis) and in design work.

Network analysis is the general engineering problem of determining how a network will respond to an input*. For instance, we might wish to determine how a structure will behave in high winds. Or we might want to know how effective a sound absorbing wall we are planning on purchasing would be in reducing machinery noise. Or perhaps we are interested in the effects of a tube of saline solution on the transmission of blood pressure waveforms from an artery to a monitor.

All of these problems and many more are examples of network analysis. As you can see a "network" can be any system at all. *One-port network analysis* is the variation of one parameter with respect to another, both measured at the same point (port) of the network. The impedance or compliance of the electronic or mechanical networks shown in Figure 2.11 are typical examples of one-port network analysis.

Figure 2.11 One-port network analysis examples.



* Network Analysis is sometimes called Stimulus/Response Testing. The input is then known as the stimulus or excitation and the output is called the response.

Two-port analysis gives the response at a second port due to an input at the first port. We are generally interested in the transmission and rejection of signals and in insuring the integrity of signal transmission. The concept of two-port analysis can be extended to any number of inputs and outputs. This is called *N-port analysis*, a subject we will use in modal analysis later in this chapter.

We have deliberately defined network analysis in a very general way. It applies to all networks with no limitations. If we place one condition on our network, *linearity*, we find that network analysis becomes a very powerful tool.

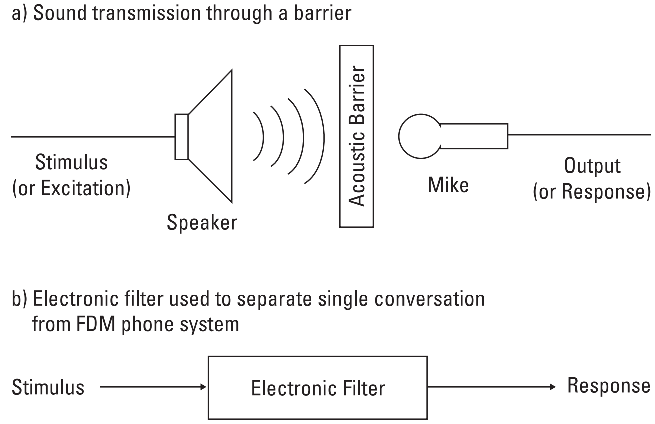


Figure 2.12
Two-port network analysis.

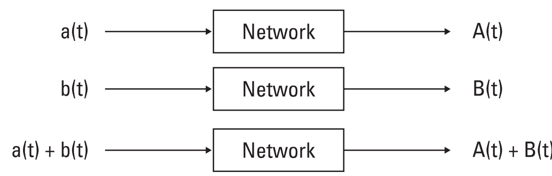


Figure 2.13
Linear network.

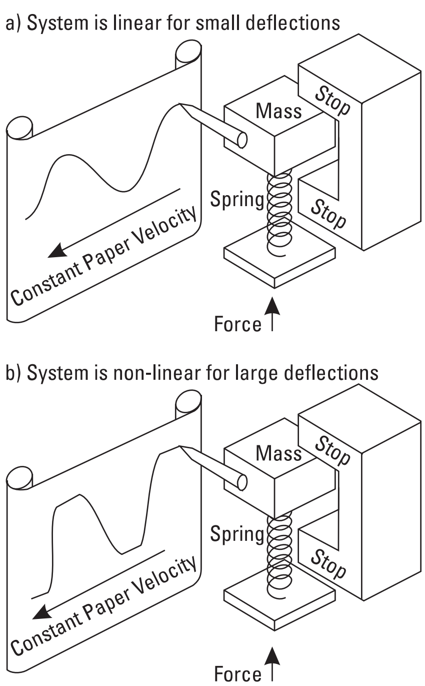


Figure 2.14
Non-linear system example.

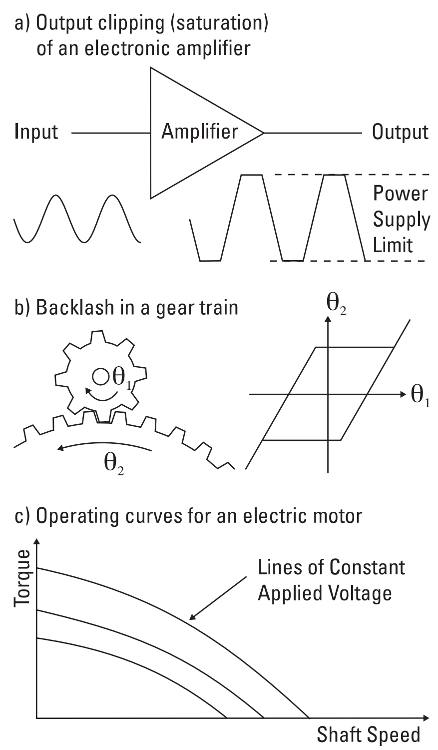


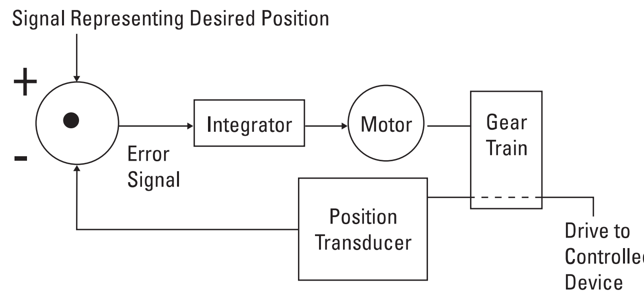
Figure 2.15
Examples of non-linearities.

When we say a network is *linear*, we mean it behaves like the network in Figure 2.13. Suppose one input causes an output A and a second input applied at the same port causes an output B. If we apply both inputs at the same time to a linear network, the output will be the sum of the individual outputs, A + B.

At first glance it might seem that all networks would behave in this fashion. A counter example, a *non-linear* network, is shown in Figure 2.14. Suppose that the first input is a force that varies in a sinusoidal manner. We pick its amplitude to ensure that the displacement is small enough so that the oscillating mass does not quite hit the stops. If we add a second identical input, the mass would now hit the stops. Instead of a sine wave with twice the amplitude, the output is clipped as shown in Figure 2.14b.

This spring-mass system with stops illustrates an important principal: *no real system is completely linear*. A system may be approximately linear over a wide range of signals, but eventually the assumption of linearity breaks down. Our spring-mass system is linear before it hits the stops. Likewise, a linear electronic amplifier clips when the output voltage approaches the internal supply voltage. A spring may compress linearly until the coils start pressing against each other.

Figure 2.16
A positioning system.



Other forms of non-linearities are also often present. Hysteresis (or backlash) is usually present in gear trains, loosely riveted joints and in magnetic devices. Sometimes the non-linearities are less abrupt and are smooth, but nonlinear, curves. The torque versus rpm of an engine or the operating curves of a transistor are two examples that can be considered linear over only small portions of their operating regions.

The important point is not that all systems are nonlinear; it is that *most systems can be approximated as linear systems*. Often a large engineering effort is spent in making the system as linear as practical. This is done for two reasons. First, it is often a design goal for the output of a network to be a scaled, linear version of the input. A strip chart recorder is a good example. The electronic amplifier and pen motor must both be designed to ensure that the deflection across the paper is linear with the applied voltage.

The second reason why systems are linearized is to reduce the problem of nonlinear instability. One example would be the positioning system shown in Figure 2.16. The actual position is compared to the desired position and the error is integrated and applied to the motor. If the gear train has no backlash, it is a straightforward problem to design this system to the desired specifications of positioning accuracy and response time.

However, if the gear train has excessive backlash, the motor will "hunt," causing the positioning system to oscillate around the desired position. The solution is either to reduce the loop gain and therefore reduce the overall performance of the system, or to reduce the backlash in the gear train. Often, reducing the backlash is the only way to meet the performance specifications.

Analysis of Linear Networks

As we have seen, many systems are designed to be reasonably linear to meet design specifications. This has a fortuitous side benefit when attempting to analyze networks*.

Recall that a real signal can be considered to be a sum of sine waves. Also, recall that the response of a linear network is the sum of the responses to each component of the input. Therefore, if we knew the response of the network to each of the sine wave components of the input spectrum, we could predict the output.

It is easy to show that the steady-state response of a linear network to a sine wave input is a sine wave of the same frequency. As shown in Figure 2.17, the amplitude of the output sine wave is proportional to the input amplitude. Its phase is shifted by an amount which depends only on the frequency of the sine wave. As we vary the frequency of the sine wave input, the amplitude proportionality factor (gain) changes as does the phase of the output. *If we divide the output of the network by the input, we get a*

Figure 2.17
Linear network response to a sine wave input.

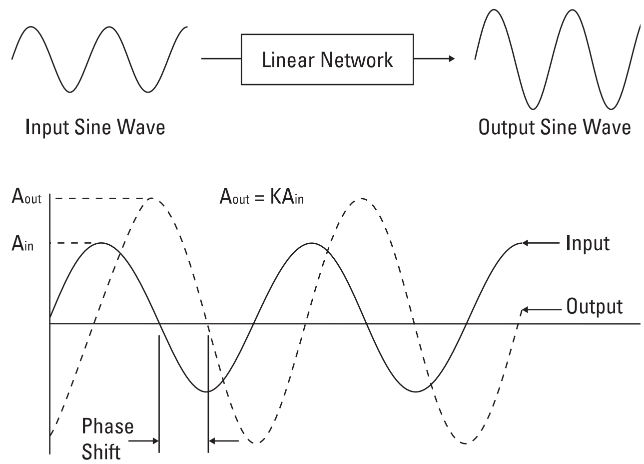
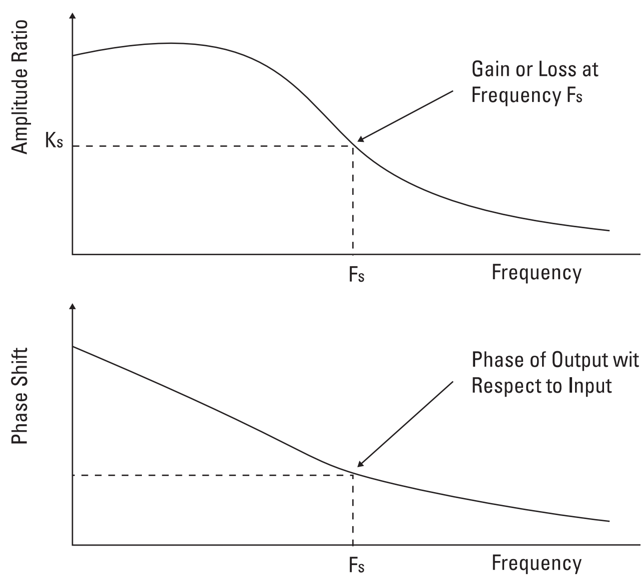


Figure 2.18
The frequency response of a network.

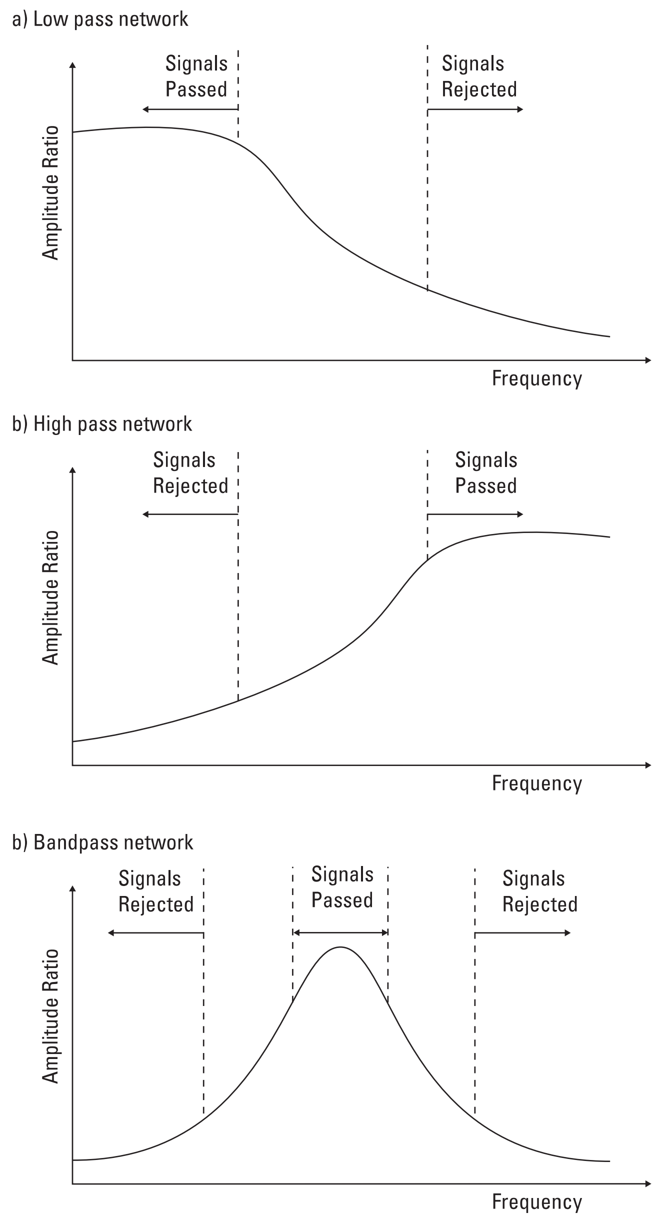


* We will discuss the analysis of networks which have not been linearized in Chapter 3, Section 6.

normalized result called the frequency response of the network. As shown in Figure 2.18, the frequency response is the gain (or loss) and phase shift of the network as a function of frequency. Because the network is linear, the frequency response is independent of the input amplitude; *the frequency response is a property of a linear network*, not dependent on the stimulus.

The frequency response of a network will generally fall into one of three categories; low pass, high pass, bandpass or a combination of these. As the names suggest, their frequency responses have relatively high gain in a band of frequencies, allowing these frequencies to pass through the network. Other frequencies suffer a relatively high loss and are rejected by the network. To see what this means in terms of the response of a filter to an input, let us look at the bandpass filter case.

Figure 2.19
Three classes of frequency response.



In Figure 2.20, we put a square wave into a bandpass filter. We recall from Figure 2.10 that a square wave is composed of harmonically related sine waves. The frequency response of our example network is shown in Figure 2.20b. Because the filter is narrow, it will pass only one component of the square wave. Therefore, the steady-state response of this bandpass filter is a sine wave.

Notice how easy it is to predict the output of any network from its frequency response. The spectrum of the input signal is multiplied by the frequency response of the network to determine the components that appear in the output spectrum. This frequency domain output can then be transformed back to the time domain.

In contrast, it is very difficult to compute in the time domain the output of any but the simplest networks. A complicated integral must be evaluated which often can only be done numerically on a digital computer*. If we computed the network response by both evaluating the time domain integral and by transforming to the frequency domain and back, we would get the same results. However, it is usually easier to compute the output by transforming to the frequency domain.

Transient Response

Up to this point we have only discussed the steady-state response to a signal. By steady-state we mean the output after any transient responses caused by applying the input have died out. However, the frequency response of a network also contains all the information necessary to predict the transient response of the network to any signal.

Figure 2.20 Bandpass filter response to a square wave input.

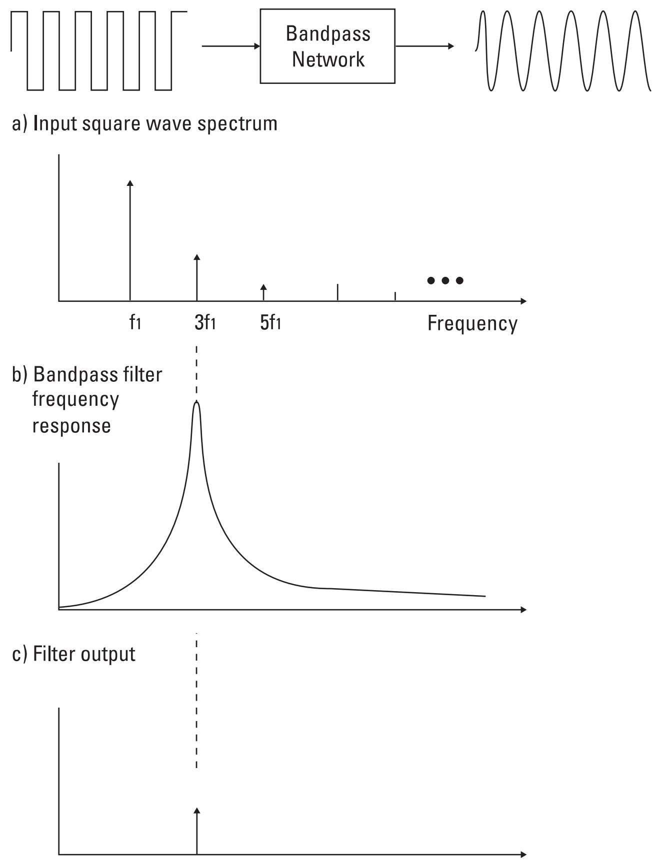
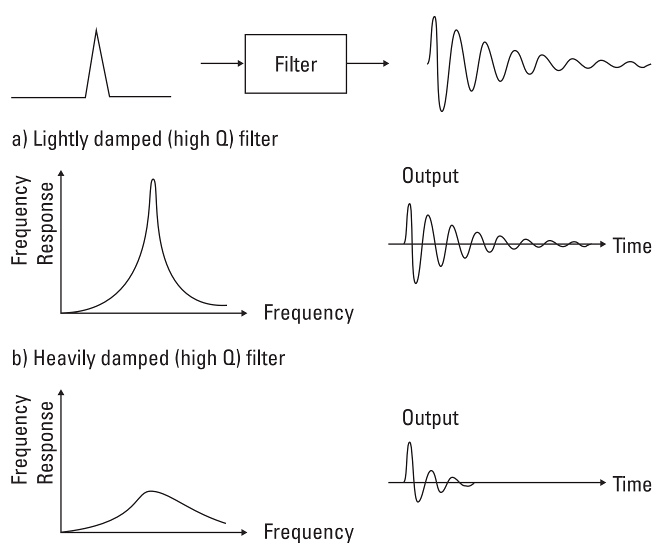


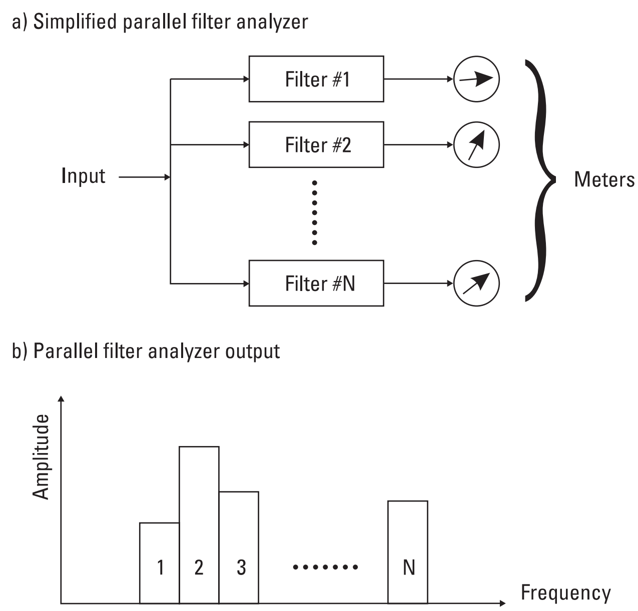
Figure 2.21 Time response of bandpass filters.



* This operation is called convolution.

Let us look qualitatively at the transient response of a bandpass filter. If a resonance is narrow compared to its frequency, then it is said to be a high "Q" resonance*. Figure 2.21a shows a high Q filter frequency response. It has a transient response which dies out very slowly. A time response which decays slowly is said to be lightly damped. Figure 2.21b shows a low Q resonance. It has a transient response which dies out quickly. This illustrates a general principle: *signals which are broad in one domain are narrow in the other.* Narrow, selective filters have very long response times, a fact we will find important in the next section.

Figure 2.22 Parallel filter analyzer.



Section 3: Instrumentation for the Frequency Domain

Just as the time domain can be measured with strip chart recorders, oscillographs or oscilloscopes, the frequency domain is usually measured with spectrum and network analyzers.

Spectrum analyzers are instruments which are optimized to characterize signals. They introduce very little distortion and few spurious signals. This insures that the signals on the display are truly part of the input signal spectrum, not signals introduced by the analyzer.

Network analyzers are optimized to give accurate amplitude and phase measurements over a wide range of network gains and losses. This design difference means that these two traditional instrument families are not interchangeable.** A spectrum analyzer can not be used as a network analyzer because it does not measure amplitude accurately and cannot measure phase. A network analyzer would make a very poor spectrum analyzer because spurious responses limit its dynamic range.

In this section we will develop the properties of several types of analyzers in these two categories.

The Parallel-Filter Spectrum Analyzer

As we developed in Section 2 of this chapter, electronic filters can be built which pass a narrow band of frequencies. If we were to add a meter to the output of such a bandpass filter, we could measure the power in the portion of the spectrum passed by the filter. In Figure 2.22a we have done this for a bank of filters, each tuned to a different frequency. If the center frequencies of these filters are chosen so that the filters overlap properly, the spectrum covered by the filters can be completely characterized as in Figure 2.22b.

* Q is usually defined as:

$$Q = \frac{\text{Center Frequency of Resonance}}{\text{Frequency Width of -3 dB Points}}$$

** Dynamic Signal Analyzers are an exception to this rule, they can act as both network and spectrum analyzers.

How many filters should we use to cover the desired spectrum? Here we have a trade-off. We would like to be able to see closely spaced spectral lines, so we should have a large number of filters. However, each filter is expensive and becomes more expensive as it becomes narrower, so the cost of the analyzer goes up as we improve its resolution. Typical audio parallel-filter analyzers balance these demands with 32 filters, each covering 1/3 of an octave.

Swept Spectrum Analyzer

One way to avoid the need for such a large number of expensive filters is to use only one filter and sweep it slowly through the frequency range of interest. If, as in Figure 2.23, we display the output of the filter versus the frequency to which it is tuned, we have the spectrum of the input signal. This swept analysis technique is commonly used in rf and microwave spectrum analysis.

We have, however, assumed the input signal hasn't changed in the time it takes to complete a sweep of our analyzer. If energy appears at some frequency at a moment when our filter is not tuned to that frequency, then we will not measure it.

One way to reduce this problem would be to speed up the sweep time of our analyzer. We could still miss an event, but the time in which this could happen would be shorter. Unfortunately though, we cannot make the sweep arbitrarily fast because of the response time of our filter.

To understand this problem, recall from Section 2 that a filter takes a finite time to respond to changes in its input. The narrower the filter, the longer it takes to respond.

Figure 2.23 Simplified swept spectrum analyzer.

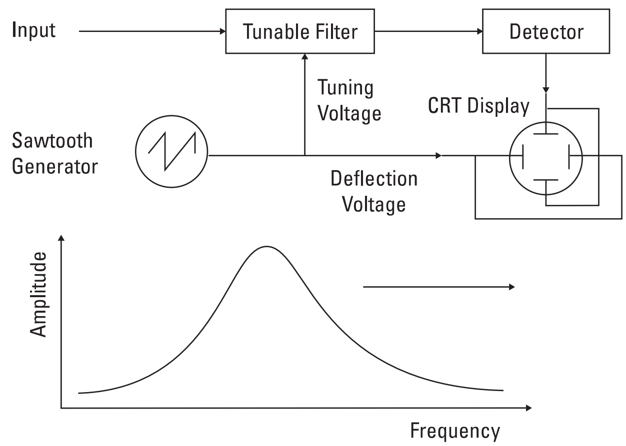
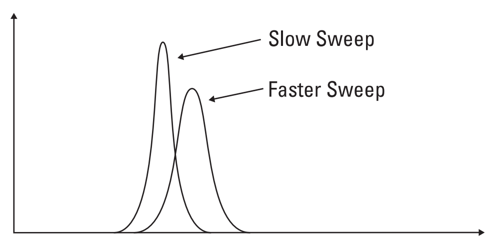


Figure 2.24 Amplitude error form sweeping too fast.



If we sweep the filter past a signal too quickly, the filter output will not have a chance to respond fully to the signal. As we show in Figure 2.24, the spectrum display will then be in error; our estimate of the signal level will be too low.

In a parallel-filter spectrum analyzer we do not have this problem. All the filters are connected to the input signal all the time. Once we have waited the initial settling time of a single filter, all the filters will be settled and the spectrum will be valid and not miss any transient events.

So there is a basic trade-off between parallel-filter and swept spectrum analyzers. The parallel-filter analyzer is fast, but has limited resolution and is expensive. The swept analyzer can be cheaper and have higher resolution but the measurement takes longer (especially at high resolution) and it can not analyze transient events*.

Dynamic Signal Analyzer

In recent years another kind of analyzer has been developed which offers the best features of the parallel-filter and swept spectrum analyzers. Dynamic Signal Analyzers are based on a high speed calculation routine which acts like a parallel filter analyzer with *hundreds* of filters and yet are cost-competitive with swept spectrum analyzers. In

* More information on the performance of swept spectrum analyzers can be found in Keysight Application Note Series 150.

addition, two channel Dynamic Signal Analyzers are in many ways better network analyzers than the ones we will introduce next.

Network Analyzers

Since in network analysis it is required to measure both the input and output, network analyzers are generally two channel devices with the capability of measuring the amplitude ratio (gain or loss) and phase difference between the channels. All of the analyzers discussed here measure frequency response by using a sinusoidal input to the network and slowly changing its frequency. Dynamic Signal Analyzers use a different, much faster technique for network analysis which we discuss in the next chapter.

Gain-phase meters are broadband devices which measure the amplitude and phase of the input and output sine waves of the network. A sinusoidal source must be supplied to stimulate the network when using a gain-phase meter as in Figure 2.25. The source can be tuned manually and the gain-phase plots done by hand or a sweeping source, and an x-y plotter can be used for automatic frequency response plots.

The primary attraction of gain-phase meters is their low price. If a sinusoidal source and a plotter are already available, frequency response measurements can be made for a very low investment. However, because gain-phase meters are broadband, they measure all the noise of the network as well as the desired sine wave. As the network attenuates the input, this noise eventually becomes a floor below which the meter cannot measure. This typically becomes a problem with attenuations of about 60 dB (1,000:1).

Figure 2.25 Gain-phase meter operation.

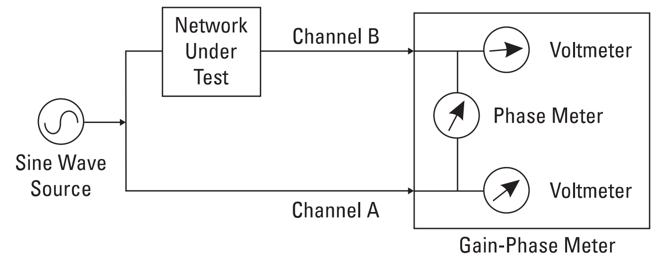
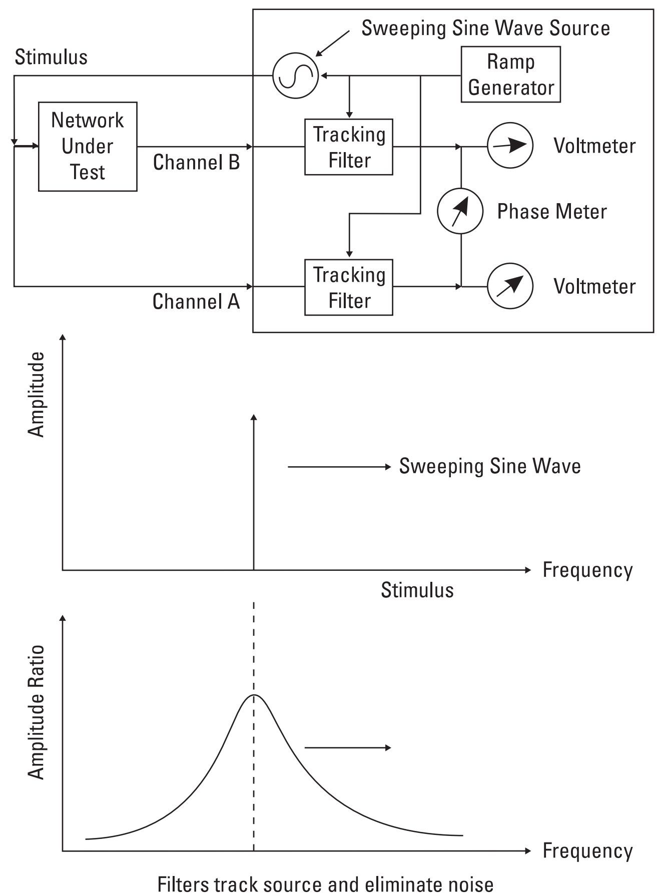


Figure 2.26 Tuned network analyzer operation.



Tuned network analyzers minimize the noise floor problems of gain-phase meters by including a bandpass filter which tracks the source frequency. Figure 2.26 shows how this tracking filter virtually eliminates the noise and any harmonics to allow measurements of attenuation to 100 dB (100,000:1).

By minimizing the noise, it is also possible for tuned network analyzers to make more accurate measurements of amplitude and phase. These improvements do not come without their price, however, as tracking filters and a dedicated source must be added to the simpler and less costly gain-phase meter.

Tuned analyzers are available in the frequency range of a few Hertz to many Gigahertz (109 Hertz). If lower frequency analysis is desired, a frequency response analyzer is often used. To the operator, it behaves exactly like a tuned network analyzer. However, it is quite different inside. It integrates the signals in the time domain to effectively filter the signals at very low frequencies where it is not practical to make filters by more conventional techniques. Frequency response analyzers are generally limited to from 1 mHz to about 10 kHz.

Section 4: The Modal Domain

In the preceding sections we have developed the properties of the time and frequency domains and the instrumentation used in these domains. In this section we will develop the properties of another domain, the modal domain. This change in perspective to a new domain is particularly useful if we are interested in analyzing the behavior of mechanical structures.

To understand the modal domain let us begin by analyzing a simple mechanical structure, a tuning fork. If we strike a tuning fork, we easily conclude from its tone that it is primarily vibrating at a single frequency. We see that we have excited a network (tuning fork) with a force impulse (hitting the fork). The time domain view of the sound caused by the deformation of the fork is a lightly damped sine wave shown in Figure 2.27b.

In Figure 2.27c, we see in the frequency domain that the frequency response of the tuning fork has a major peak that is very lightly damped, which is the tone we hear. There are also several smaller peaks.

Figure 2.27 The vibration of a tuning fork.

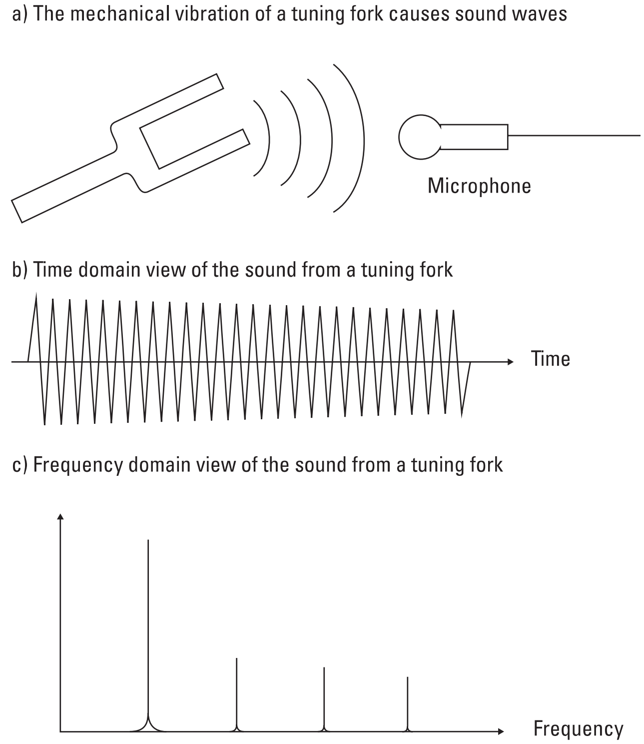
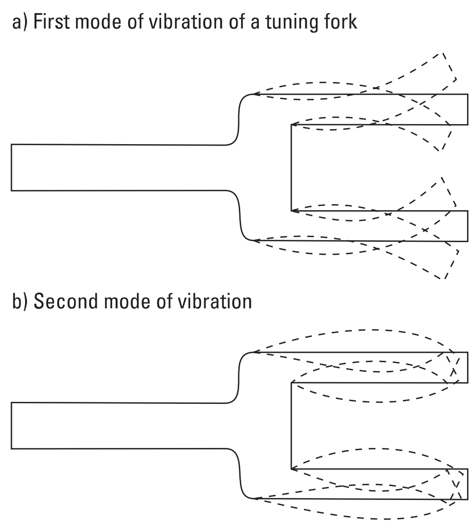


Figure 2.28 Example vibration modes of a tuning fork.



Each of these peaks, large and small, corresponds to a "vibration mode" of the tuning fork. For instance, we might expect for this simple example that the major tone is caused by the vibration mode shown in Figure 2.28a. The second harmonic might be caused by a vibration like Figure 2.28b

We can express the vibration of any structure as a sum of its vibration modes. Just as we can represent a real waveform as a sum of much simpler sine waves, we can represent any vibration as a sum of much simpler vibration modes. The task of "modal" analysis is to determine the shape and the magnitude of the structural deformation in each vibration mode. Once these are known, it usually becomes apparent how to change the overall vibration.

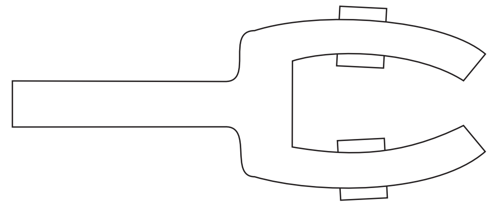
For instance, let us look again at our tuning fork example. Suppose that we decided that the second harmonic tone was too loud. How should we change our tuning fork to reduce the harmonic? If we had measured the vibration of the fork and determined that the modes of vibration were those shown in Figure 2.28, the answer becomes clear. We might apply damping material at the center of the tines of the fork. This would greatly affect the second mode which has maximum deflection at the center while only slightly affecting the desired vibration of the first mode. Other solutions are possible, but all depend on knowing the geometry of each mode.

The Relationship Between the Time, Frequency and Modal Domain

To determine the total vibration of our tuning fork or any other structure, we have to measure the vibration at several points on the structure. Figure 2.30a shows some points we might pick. If we transformed this time domain data to the frequency domain, we would get results like Figure 2.30b. We measure frequency response because we want to measure the properties of the structure independent of the stimulus*.

Figure 2.29 Reducing the second harmonic by damping the second vibration mode.

a) Damping material applied at maximum displacement in second mode



b) Damping material has little effect on first mode

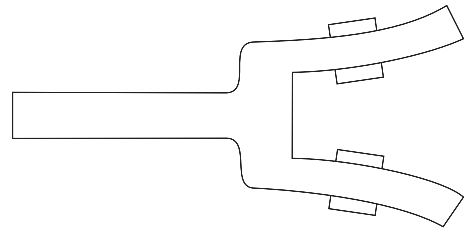
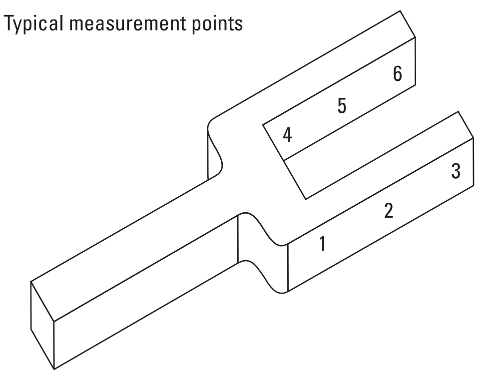
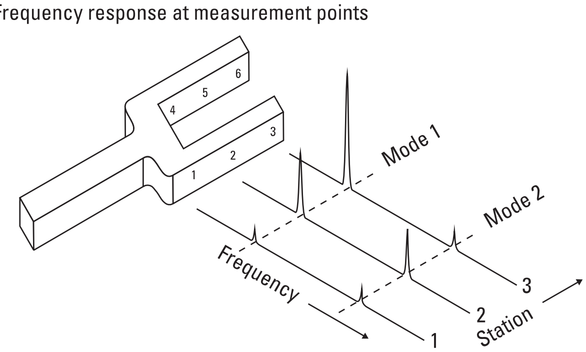


Figure 2.30 Modal analysis of a tuning fork.

a) Typical measurement points



b) Frequency response at measurement points



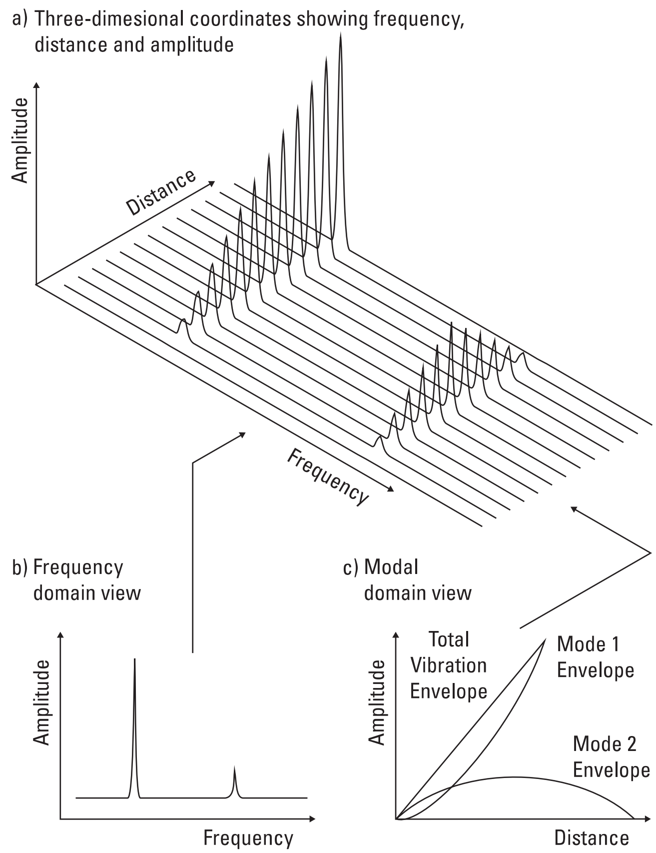
* Those who are more familiar with electronics might note that we have measured the frequency response of a network (structure) at N points and thus have performed an N-port Analysis.

We see that the sharp peaks (resonances) all occur at the same frequencies independent of where they are measured on the structure. Likewise we would find by measuring the width of each resonance that the damping (or Q) of each resonance is independent of position. The only parameter that varies as we move from point to point along the structure is the relative height of resonances.* By connecting the peaks of the resonances of a given mode, we trace out the mode shape of that mode.

Experimentally we have to measure only a few points on the structure to determine the mode shape. However, to clearly show the mode shape in our figure, we have drawn in the frequency response at many more points in Figure 2.31a. If we view this three-dimensional graph along the distance axis, as in Figure 2.31b, we get a combined frequency response. Each resonance has a peak value corresponding to the peak displacement in that mode. If we view the graph along the frequency axis, as in Figure 2.31c, we can see the mode shapes of the structure.

We have not lost any information by this change of perspective. Each vibration mode is characterized by its mode shape, frequency and damping from which we can reconstruct the frequency domain view.

Figure 2.31
The relationship between the frequency and the modal domains.



However, the equivalence between the modal, time and frequency domains is not quite as strong as that between the time and frequency domains. Because the modal domain portrays the properties of the network independent of the stimulus, transforming back to the time domain gives the impulse response of the structure, no matter what the stimulus. A more important limitation of this equivalence is that curve fitting is used in transforming from our frequency response measurements to the modal domain to minimize the effects of noise and small experimental errors. No information is lost in this curve fitting, so all three domains contain the same information, but not the same noise. Therefore, transforming from the frequency domain to the modal domain and back again will give results like those in Figure 2.32. The results are not exactly the same, yet in all the important features, the frequency responses are the same. This is also true of time domain data derived from the modal domain.

* The phase of each resonance is not shown for clarity of the figures but it too is important in the mode shape. The magnitude of the frequency response gives the magnitude of the mode shape while the phase gives the direction of the deflection.

Section 5: Instrumentation for the Modal Domain

There are many ways that the modes of vibration can be determined. In our simple tuning fork example we could guess what the modes were. In simple structures like drums and plates it is possible to write an equation for the modes of vibration. However, in almost any real problem, the solution can neither be guessed nor solved analytically because the structure is too complicated. In these cases it is necessary to measure the response of the structure and determine the modes.

There are two basic techniques for determining the modes of vibration in complicated structures: 1) exciting only one mode at a time, and 2) computing the modes of vibration from the total vibration.

Single Mode Excitation Modal Analysis

To illustrate single mode excitation, let us look once again at our simple tuning fork example. To excite just the first mode we need two shakers, driven by a sine wave and attached to the ends of the tines as in Figure 2.33a. Varying the frequency of the generator near the first mode resonance frequency would then give us its frequency, damping and mode shape.

In the second mode, the ends of the tines do not move, so to excite the second mode we must move the shakers to the center of the tines. If we anchor the ends of the tines, we will constrain the vibration to the second mode alone.

Figure 2.32 Curve fitting removes measurement noise.

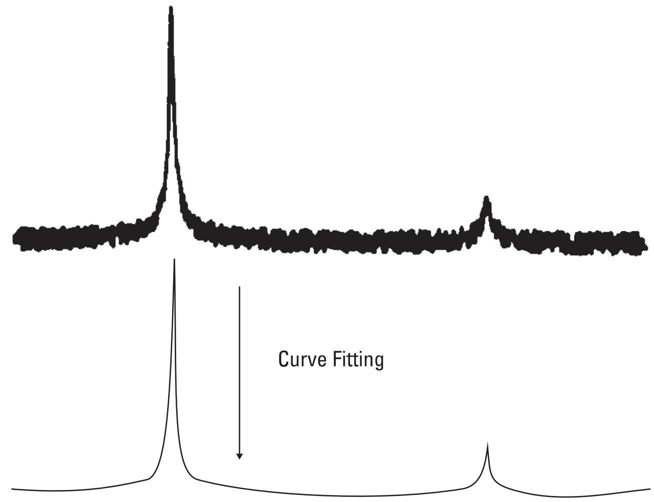
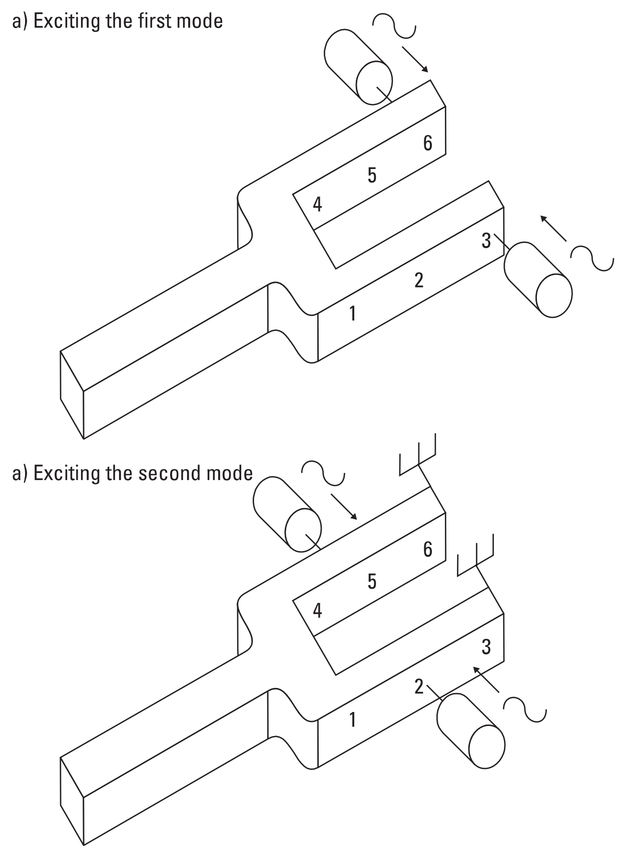


Figure 2.33 Single mode excitation modal analysis.



In more realistic, three dimensional problems, it is necessary to add many more shakers to ensure that only one mode is excited. The difficulties and expense of testing with many shakers has limited the application of this traditional modal analysis technique.

Modal Analysis From Total Vibration

To determine the modes of vibration from the total vibration of the structure, we use the techniques developed in the previous section. Basically, we determine the frequency response of the structure at several points and compute at each resonance the frequency, damping and what is called the residue (which represents the height of the resonance). This is done by a curve-fitting routine to smooth out any noise or small experimental errors. From these measurements and the geometry of the structure, the mode shapes are computed and drawn on a CRT display or a plotter. If drawn on a CRT, these displays may be animated to help the user understand the vibration mode.

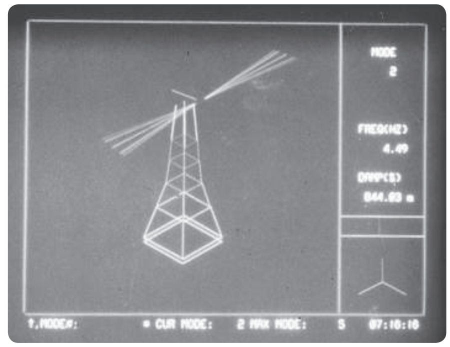


Figure 2.34 Measured mode shape.

From the above description, it is apparent that a modal analyzer requires some type of network analyzer to measure the frequency response of the structure and a computer to convert the frequency response to mode shapes. This can be accomplished by connecting a Dynamic Signal Analyzer through a digital interface* to a computer furnished with the appropriate software. This capability is also available in a single instrument called a Structural Dynamics Analyzer. In general, computer systems offer more versatile performance since they can be programmed to solve other problems. However, Structural Dynamics Analyzers generally are much easier to use than computer systems.

Section 6: Summary

In this chapter we have developed the concept of looking at problems from different perspectives. These perspectives are the time, frequency and modal domains. Phenomena that are confusing in the time domain are often clarified by changing perspective to another domain. Small signals are easily resolved in the presence of large ones in the frequency domain. The frequency domain is also valuable for predicting the output of any kind of linear network. A change to the modal domain breaks down complicated structural vibration problems into simple vibration modes.

No one domain is always the best answer, so the ability to easily change domains is quite valuable. Of all the instrumentation available today, only Dynamic Signal Analyzers can work in all three domains. In the next chapter we develop the properties of this important class of analyzers.

* GPIB, Keysight's implementation of IEEE-488-1975 is ideal for this application.

Chapter 3
Understanding Dynamic Signal Analysis

We saw in the previous chapter that the Dynamic Signal Analyzer has the speed advantages of parallel-filter analyzers without their low resolution limitations. In addition, it is the only type of analyzer that works in all three domains. In this chapter we will develop a fuller understanding of this important analyzer family, Dynamic Signal Analyzers. We begin by presenting the properties of the Fast Fourier Transform (FFT) upon which Dynamic Signal Analyzers are based. No proof of these properties is given, but heuristic arguments as to their validity are used where appropriate. We then show how these FFT properties cause some undesirable characteristics in spectrum analysis like aliasing and leakage. Having demonstrated a potential difficulty with the FFT, we then show what solutions are used to make practical Dynamic Signal Analyzers. Developing this basic knowledge of FFT characteristics makes it simple to get good results with a Dynamic Signal Analyzer in a wide range of measurement problems.

Section 1: FFT Properties

The Fast Fourier Transform (FFT) is an algorithm* for *transforming* data from the time domain to the frequency domain. Since this is exactly what we want a spectrum analyzer to do, it would seem easy to implement a Dynamic Signal Analyzer based on the FFT. However, we will see that there are many factors which complicate this seemingly straightforward task.

First, because of the many calculations involved in transforming domains, the transform must be implemented on a digital computer if the results are to be sufficiently accurate. Fortunately, with the advent of microprocessors, it is easy and inexpensive to incorporate all the needed computing power in a small instrument package. Note, however, that we cannot now transform to the frequency domain in a continuous manner, but instead must sample and digitize the time domain input. This means that our algorithm transforms digitized samples from the time domain to samples in the frequency domain as shown in Figure 3.1.**

Because we have sampled, we no longer have an exact representation in either domain. However, a sampled representation can be as close to ideal as we desire by placing our samples closer together. Later in this chapter, we will consider what sample spacing is necessary to guarantee accurate results.

Figure 3.1
The FFT samples in both the time and frequency domains.

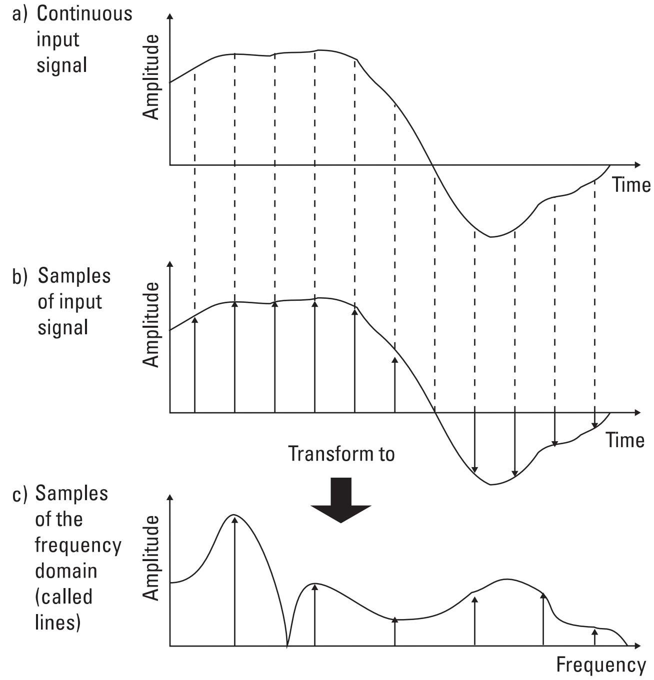
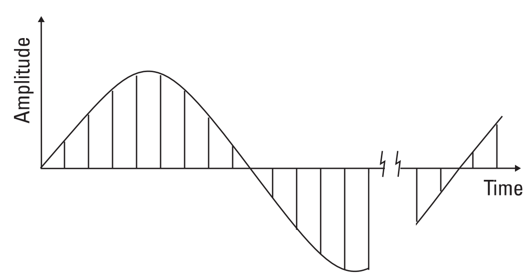


Figure 3.2
A time record is N equally spaced samples of the input.



* An algorithm is any special mathematical method of solving a certain kind of problem; e.g., the technique you use to balance your checkbook.

** To reduce confusion about which domain we are in, samples in the frequency domain are called lines.

Time Records

A *time record* is defined to be N consecutive, equally spaced samples of the input. Because it makes our transform algorithm simpler and much faster, N is restricted to be a multiple of 2, for instance 1024.

As shown in Figure 3.3, this time record is transformed as a complete *block* into a complete *block* of frequency lines. All the samples of the time record are needed to compute each and every line in the frequency domain. This is in contrast to what one might expect, namely that a single time domain sample transforms to exactly one frequency domain line. Understanding this *block processing* property of the FFT is crucial to understanding many of the properties of the Dynamic Signal Analyzer.

For instance, because the FFT transforms the entire time record block as a total, there cannot be valid frequency domain results until a complete time record has been gathered. However, once completed, the oldest sample could be discarded, all the samples shifted in the time record, and a new sample added to the end of the time record as in Figure 3.4. Thus, once the time record is initially filled, we have a new time record at every time domain sample and therefore could have new valid results in the frequency domain at every time domain sample.

This is very similar to the behavior of the parallel-filter analyzers described in the previous chapter. When a signal is first applied to a parallel-filter analyzer, we must wait for the filters to respond, then we can see very rapid changes in the frequency domain. With a Dynamic Signal Analyzer we do not get a valid result until a full time record has been gathered. Then rapid changes in the spectra can be seen.

Figure 3.3
The FFT works on blocks of data.

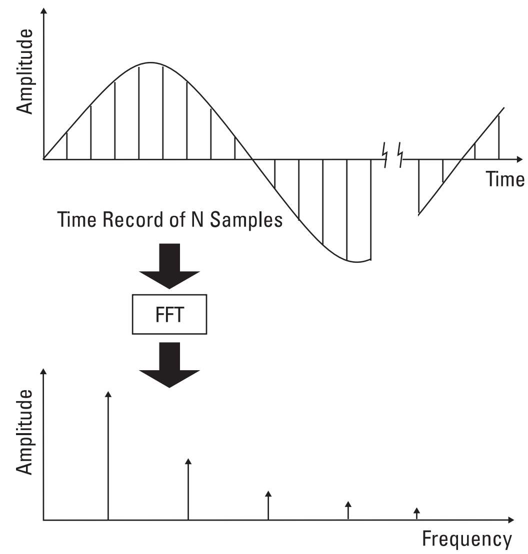
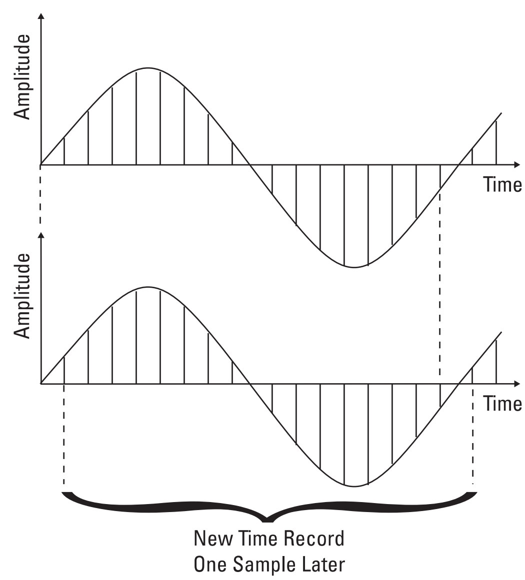


Figure 3.4
A new time record every sample after the time record is filled.



It should be noted here that a new spectrum every sample is usually too much information, too fast. This would often give you *thousands* of transforms per second. Just how fast a Dynamic Signal Analyzer should transform is a subject better left to the sections in this chapter on real time bandwidth and overlap processing.

How Many Lines are There?

We stated earlier that the time record has N equally spaced samples. Another property of the FFT is that it transforms these time domain samples to N/2 equally spaced lines in the frequency domain. We only get half as many lines because each frequency line actually contains two pieces of information, amplitude and phase. The meaning of this is most easily seen if we look again at the relationship between the time and frequency domain.

Figure 3.5 reproduces from Chapter 2 our three-dimensional graph of this relationship. Up to now we have implied that the amplitude and frequency of the sine waves contains all the information necessary to reconstruct the input. But it should be obvious that the phase of each of these sine waves is important too. For instance, in Figure 3.6, we have shifted the *phase* of the higher frequency sine wave components of this signal. The result is a severe distortion of the original wave form.

We have not discussed the phase information contained in the spectrum of signals until now because none of the traditional spectrum analyzers are capable of measuring phase. When we discuss measurements in Chapter 4, we shall find that phase contains valuable information in determining the cause of performance problems.

Figure 3.5 The relationship between the time and frequency domains.

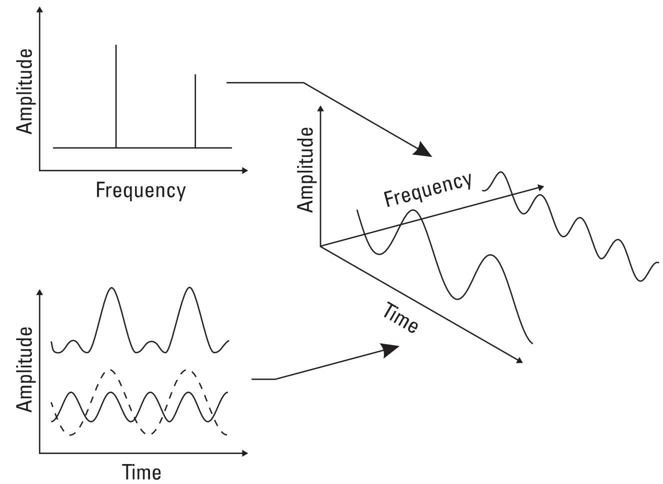
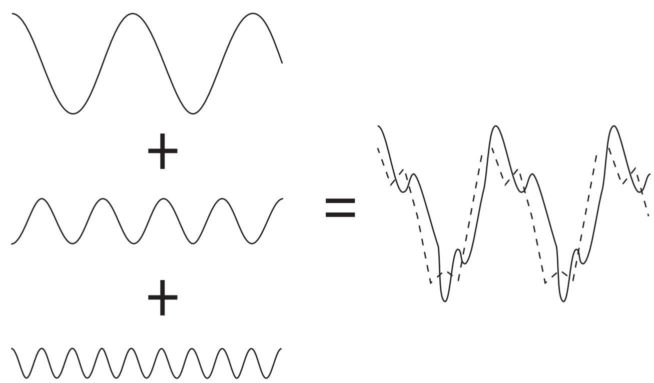


Figure 3.6 Phase of frequency domain components is important.



What is the Spacing of the Lines?

Now that we know that we have N/2 equally spaced lines in the frequency domain, what is their spacing? The lowest frequency that we can resolve with our FFT spectrum analyzer must be based on the length of the time record. We can see in Figure 3.7 that if the period of the input signal is longer than the time record, we have no way of determining the period (or frequency, its reciprocal). Therefore, the lowest frequency line of the FFT must occur at frequency equal to the reciprocal of the time record length.

In addition, there is a frequency line at zero Hertz, DC. This is merely the average of the input over the time record. It is rarely used in spectrum or network analysis. But, we have now established the spacing between these two lines and hence every line; it is the reciprocal of the time record.

What is the Frequency Range of the FFT?

We can now quickly determine that the highest frequency we can measure is:

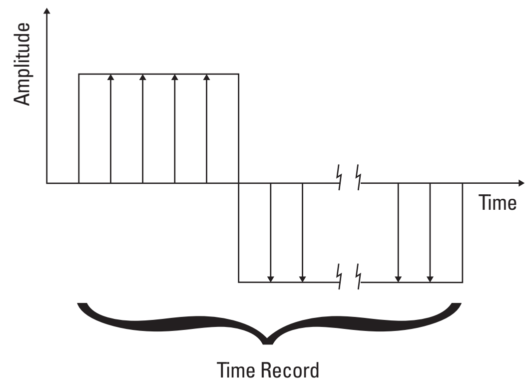
$$f_{max} = \frac{N}{2} \bullet \frac{1}{\text{Period of Time Record}}$$

because we have N/2 lines spaced by the reciprocal of the time record starting at zero Hertz *.

Since we would like to adjust the frequency range of our measurement, we must vary f_{max}. The number of time samples N is fixed by the implementation of the FFT algorithm. Therefore, we must vary the period of the time record to vary f_{max}. To do this, we must vary the sample rate so that we always have N samples in our variable time record period. This is illustrated in Figure 3.9. Notice that to cover higher frequencies, we must sample faster.

Figure 3.7 Lowest frequency resolvable by the FFT.

a) Period of input signal equals time record. Lowest resolvable frequency.



b) Period of input signal longer than the time record. Frequency of the input signal is unknown..

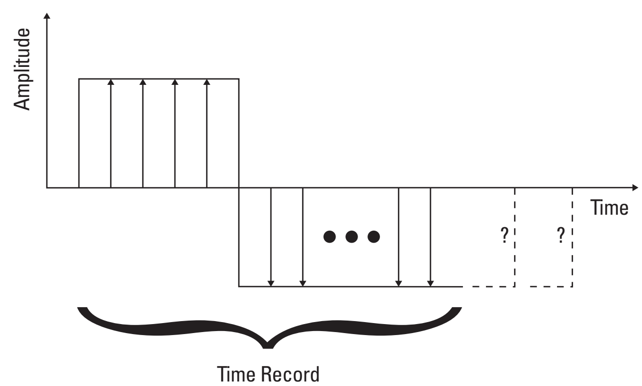


Figure 3.8 Frequencies of all the spectral lines of the FFT.

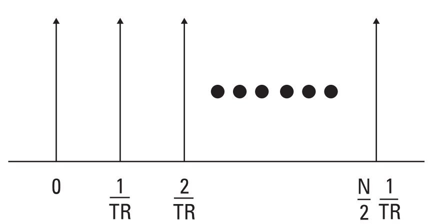
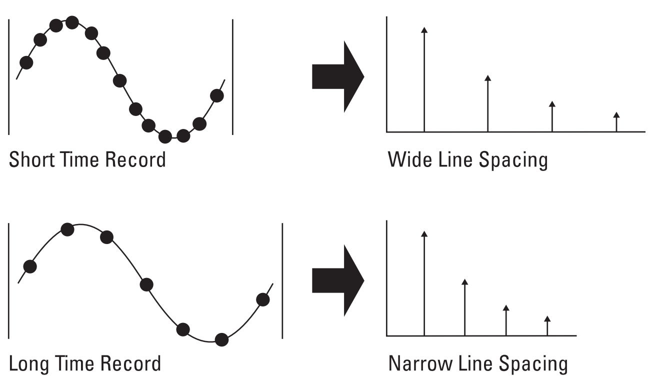


Figure 3.9 Frequency range of Dynamic Signal Analyzers is determined by sample rate.



* The usefulness of this frequency range can be limited by the problem of aliasing. Aliasing is discussed in Section 3.

Section 2*: Sampling and Digitizing

Recall that the input to our Dynamic Signal Analyzer is a continuous analog voltage. This voltage might be from an electronic circuit or could be the output of a transducer and be proportional to current, power, pressure, acceleration or any number of other inputs. Recall also that the FFT requires digitized samples of the input for its digital calculations. Therefore, we need to add a sampler and analog to digital converter (ADC) to our FFT processor to make a spectrum analyzer. We show this basic block diagram in Figure 3.10.

Figure 3.10
Block diagram of dynamic Signal Analyzer.

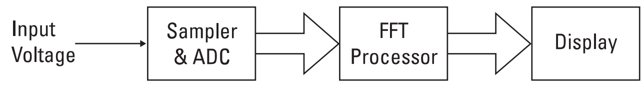
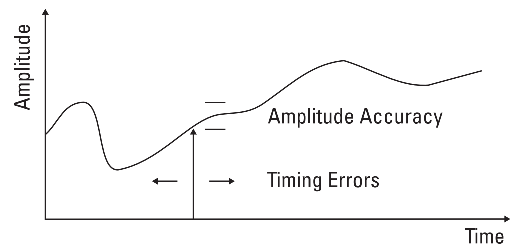


Figure 3.11
The Sampler and ADC must not introduce errors.



For the analyzer to have the high accuracy needed for many measurements, the sampler and ADC must be quite good. The sampler must sample the input at exactly the correct time and must accurately hold the input voltage measured at this time until the ADC has finished its conversion. The ADC must have high resolution and linearity. For 70 dB of dynamic range the ADC must have at least 12 bits of resolution and one half least significant bit linearity.

A good Digital Voltmeter (DVM) will typically exceed these specifications, but the ADC for a Dynamic Signal Analyzer must be much faster than typical fast DVM's. A fast DVM might take a thousand readings per second, but in a typical Dynamic Signal Analyzer the ADC must take at least a hundred thousand readings per second.

Section 3: Aliasing

The reason an FFT spectrum analyzer needs so many samples per second is to avoid a problem called aliasing. Aliasing is a potential problem in any sampled data system. It is often overlooked, sometimes with disastrous results.

A Simple Data Logging Example of Aliasing

Let us look at a simple data logging example to see what aliasing is and how it can be avoided. Consider the example for recording temperature shown in Figure 3.12. A thermocouple is connected to a digital voltmeter which is in turn connected to a printer. The system is set up to print the temperature every second. What would we expect for an output?

If we were measuring the temperature of a room which only changes slowly, we would expect every reading to be almost the same as the previous one. In fact, we are sampling much more often than necessary to determine the temperature of the room with time. If we plotted the results of this "thought experiment", we would expect to see results like Figure 3.13.

Figure 3.12
A simple sampled data system.

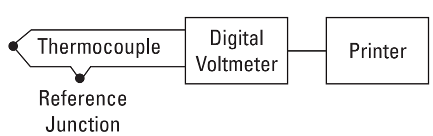
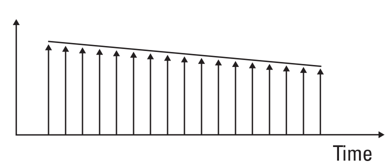


Figure 3.13
Plot of temperature variation of a room.



* This section and the next can be skipped by those not interested in the internal operation of a Dynamic Signal Analyzer. However, those who specify the purchase of Dynamic Signal Analyzers are especially encouraged to read these sections. The basic knowledge to be gained from these sections can insure specifying the best analyzer for your requirements.

The Case of the Missing Temperature

If, on the other hand, we were measuring the temperature of a small part which could heat and cool rapidly, what would the output be? Suppose that the temperature of our part cycled exactly once every second. As shown in Figure 3.14, our printout says that the temperature never changes.

What has happened is that we have sampled at exactly the same point on our periodic temperature cycle with every sample. We have not sampled fast enough to see the temperature fluctuations.

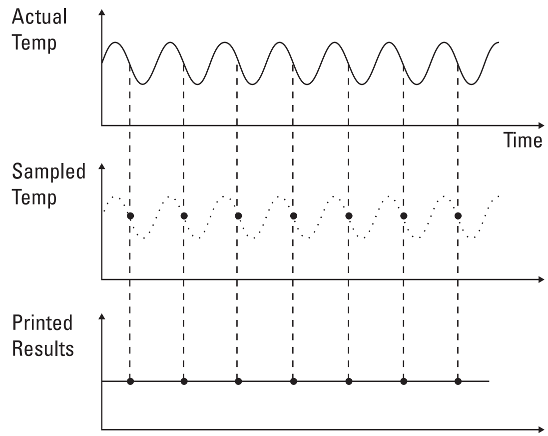


Figure 3.14
Plot of temperature variation of a small part.

Aliasing in the Frequency Domain

This completely erroneous result is due to a phenomena called aliasing.* Aliasing is shown in the frequency domain in Figure 3.15. Two signals are said to alias if the difference of their frequencies falls in the frequency range of interest. This difference frequency is always generated in the process of sampling. In Figure 3.15, the input frequency is slightly higher than the sampling frequency so a low frequency alias term is generated. If the input frequency equals the sampling frequency as in our small part example, then the alias term falls at DC (zero Hertz) and we get the constant output that we saw above.

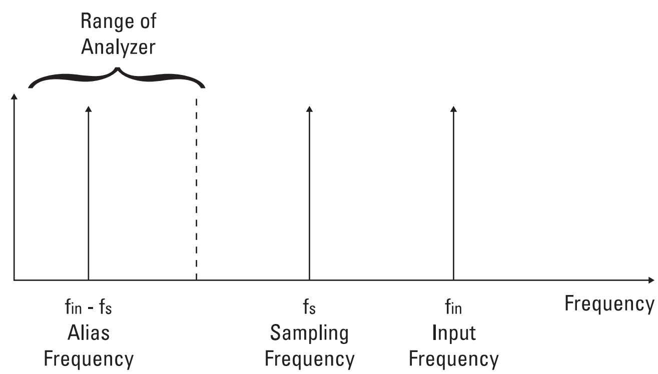


Figure 3.15
The problem of aliasing viewed in the frequency domain.

Aliasing is not always bad. It is called mixing or heterodyning in analog electronics, and is commonly used for tuning household radios and televisions as well as many other communication products. However, in the case of the missing temperature variation of our small part, we definitely have a problem. How can we guarantee that we will avoid this problem in a measurement situation?

Figure 3.16 shows that if we sample at greater than twice the highest frequency of our input, the alias products will not fall within the frequency range of our input. Therefore, a filter (or our FFT processor which acts like a filter) after the sampler will remove the alias products while passing the desired input signals *if the sample rate is greater than twice the highest frequency of the input*. If the sample rate is lower, the alias products will fall in the frequency range of the input and no amount of filtering will be able to remove them from the signal.

* Aliasing is also known as fold-over or mixing.

This minimum sample rate requirement is known as the Nyquist Criterion. It is easy to see in the time domain that a sampling frequency exactly twice the input frequency would not always be enough. It is less obvious that slightly more than two samples in each period is sufficient information. It certainly would not be enough to give a high quality time display. Yet we saw in Figure 3.16 that meeting the Nyquist Criterion of a sample rate greater than twice the maximum input frequency is sufficient to avoid aliasing and preserve all the information in the input signal.

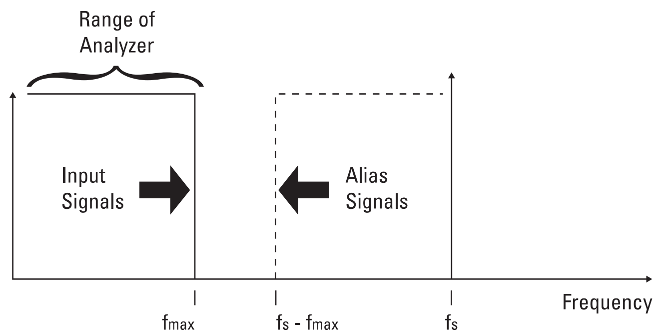


Figure 3.16 A frequency domain view of how to avoid aliasing - sample at greater than twice the highest input frequency.

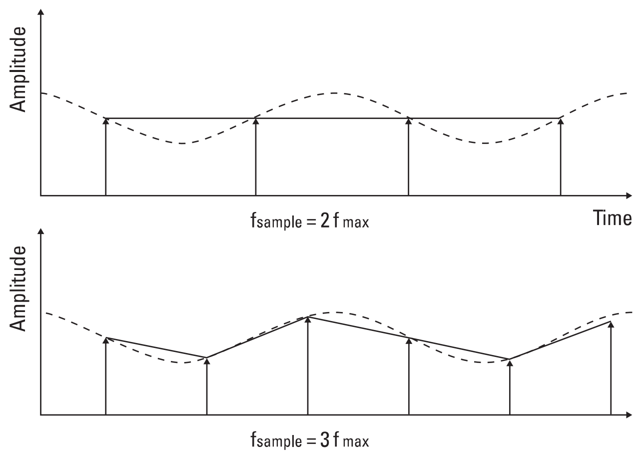


Figure 3.17 Nyquist Criterion in the time domain.

The Need for an Anti-Alias Filter

Unfortunately, the real world rarely restricts the frequency range of its signals. In the case of the room temperature, we can be reasonably sure of the maximum rate at which the temperature could change, but we still can not rule out stray signals. Signals induced at the powerline frequency or even local radio stations could alias into the desired frequency range. The only way to be really certain that the input frequency range is limited is to add a low pass filter before the sampler and ADC. Such a filter is called an anti-alias filter.

An ideal anti-alias filter would look like Figure 3.18a. It would pass all the desired input frequencies with no loss and completely reject any higher frequencies which otherwise could alias into the input frequency range. However, it is not even theoretically possible to build such a filter, much less practical. Instead, all real filters look something like Figure 3.18b with a gradual roll off and finite rejection of undesired signals. Large input signals which are not well attenuated in the transition band could still alias into the desired input frequency

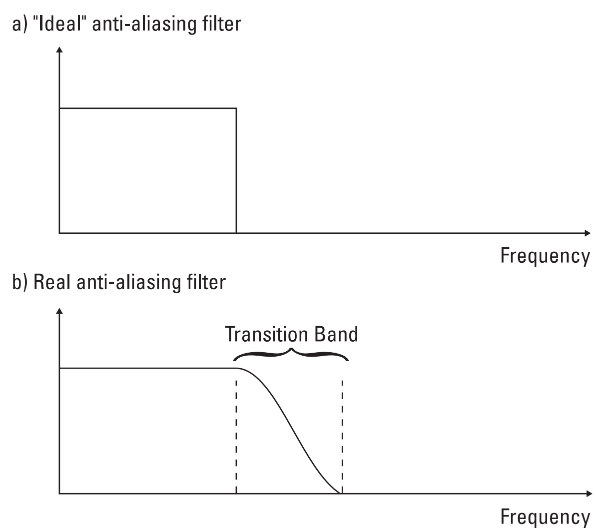


Figure 3.18 Actual anti-alias filters require higher sampling frequencies.

range. To avoid this, the sampling frequency is raised to twice the highest frequency of the transition band. This guarantees that any signals which could alias are well attentuated by the stop band of the filter. Typically, this means that the sample rate is now two and a half to four times the maximum desired input frequency. Therefore, a 25 kHz FFT Spectrum Analyzer can require an ADC that runs at 100 kHz as we stated without proof in Section 2 of this Chapter*.

The Need for More Than One Anti-Alias Filter

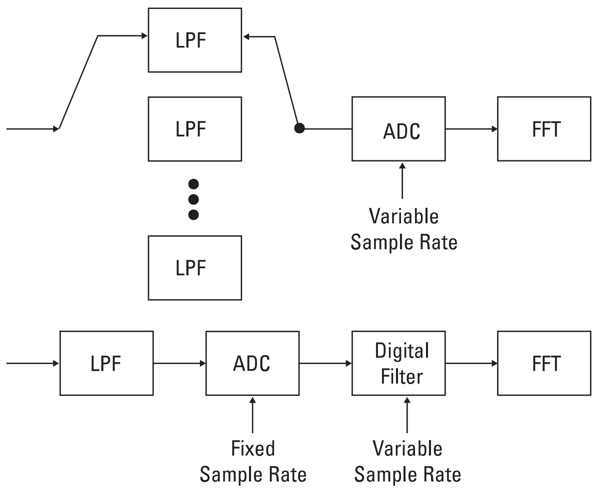
Recall from Section 1 of this Chapter, that due to the properties of the FFT we must vary the sample rate to vary the frequency span of our analyzer. To reduce the frequency span, we must reduce the sample rate. From our considerations of aliasing, we now realize that we must also reduce the anti-alias filter frequency by the same amount.

Since a Dynamic Signal Analyzer is a very versatile instrument used in a wide range of applications, it is desirable to have a wide range of frequency spans available. Typical instruments have a minimum span of 1 Hertz and a maximum of tens to hundreds of kilohertz. This four decade range typically needs to be covered with at least three spans per decade. This would mean at least twelve anti-alias filters would be required for each channel.

Each of these filters must have very good performance. It is desirable that their transition bands be as narrow as possible so that as many lines as possible are free from alias products. Additionally, in a two channel analyzer, each filter pair must be well matched for accurate network analysis measurements. These two points unfortunately mean that each of the filters is expensive. Taken together they can add significantly to the price of the analyzer. Some manufacturers don't have a low enough frequency anti-alias filter on the lowest frequency spans to save some of this expense. (The lowest frequency filters cost the most of all.) But as we have seen, this can lead to problems like our "case of the missing temperature".

* Unfortunately, because the spacing of the FFT lines depends on the sample rate, increasing the sample rate decreases the number of lines that are in the desired frequency range. Therefore, to avoid aliasing problems Dynamic Signal Analyzers have only .25N to .4N lines instead of N/2 lines.

Figure 3.19
Block diagrams of analog and digital filtering.



Digital Filtering

Fortunately, there is an alternative which is cheaper and when used in conjunction with a single analog anti-alias filter, always provides aliasing protection. It is called digital filtering because it filters the input signal after we have sampled and digitized it. To see how this works, let us look at Figure 3.19.

In the analog case we already discussed, we had to use a new filter every time we changed the sample rate of the Analog to Digital Converter (ADC). When using digital filtering, the ADC sample rate is left constant at the rate needed for the highest frequency span of the analyzer. This means we need not change our anti-alias filter. To get the reduced sample rate and filtering we need for the narrower frequency spans, we follow the ADC with a digital filter.

This digital filter is known as a decimating filter. It not only filters the digital representation of the signal to the desired frequency span, it also reduces the sample rate at its output to the rate needed for that frequency span. Because this filter is digital, there are no manufacturing variations, aging or drift in the filter. Therefore, in a two channel analyzer the filters in each channel are identical. It is easy to design a single digital filter to work on many frequency spans so the need for multiple filters per channel is avoided. All these factors taken together mean that digital filtering is much less expensive than analog anti-aliasing filtering.

Section 4:
Band Selectable Analysis

Suppose we need to measure a small signal that is very close in frequency to a large one. We might be measuring the powerline sidebands (50 or 60 Hz) on a 20 kHz oscillator. Or we might want to distinguish between the stator vibration and the shaft imbalance in the spectrum of a motor.*

Recall from our discussion of the properties of the Fast Fourier Transform that it is equivalent to a set of filters, starting at zero Hertz, equally spaced up to some maximum frequency. Therefore, our frequency resolution is limited to the maximum frequency divided by the number of filters.

To just resolve the 60 Hz sidebands on a 20 kHz oscillator signal would require 333 lines (or filters) of the FFT. Two or three times more lines would be required to accurately measure the sidebands. But typical Dynamic Signal Analyzers only have 200 to 400 lines, not enough for accurate measurements. To increase the number of lines would greatly increase the cost of the analyzer. If we chose to pay the extra cost, we would still have trouble seeing the results. With a 4 inch (10 cm) screen, the sidebands would be only 0.01 inch (.25 mm) from the carrier.

Figure 3.20
High resolution measurements with Band Selectable Analysis.

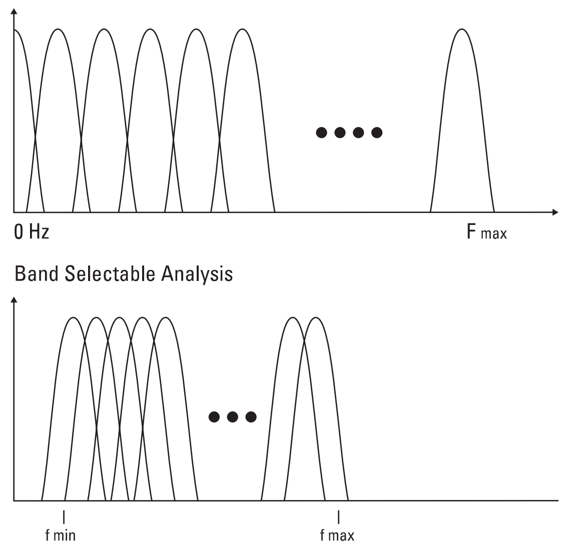
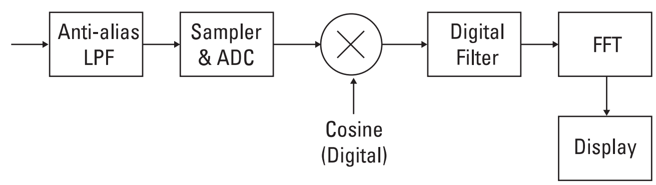


Figure 3.21
Analyzer block diagram.



A better way to solve this problem is to concentrate the filters into the frequency range of interest as in Figure 3.20. If we select the minimum frequency as well as the maximum frequency of our filters we can "zoom in" for a high resolution close-up shot of our frequency spectrum. We now have the capability of looking at the entire spectrum at once with low resolution as well as the ability to look at what interests us with much higher resolution.

This capability of increased resolution is called Band Selectable Analysis (BSA).** It is done by mixing or heterodyning the input signal down into the range of the FFT span selected. This technique, familiar to electronic engineers, is the process by which radios and televisions tune in stations.

The primary difference between the implementation of BSA in Dynamic Signal Analyzers and heterodyne radios is shown in Figure 3.21. In a radio, the sine wave used for mixing is an analog voltage. In a Dynamic Signal Analyzer, the mixing is done after the input has been digitized, so the "sine wave" is a series of digital numbers into a digital multiplier. This means that the mixing will be done with a very accurate and stable digital signal so our high resolution display will likewise be very stable and accurate.

* The shaft of an ac induction motor always runs at a rate slightly lower than a multiple of the driven frequency, an effect called slippage.

** Also sometimes called "zoom".

Section 5: Windowing

The Need for Windowing

There is another property of the Fast Fourier Transform which affects its use in frequency domain analysis. We recall that the FFT computes the frequency spectrum from a block of samples of the input called a time record. In addition, the FFT algorithm is based upon the assumption that this time record is repeated throughout time as illustrated in Figure 3.22.

This does not cause a problem with the transient case shown. But what happens if we are measuring a continuous signal like a sine wave? If the time record contains an integral number of cycles of the input sine wave, then this assumption exactly matches the actual input waveform as shown in Figure 3.23. In this case, the input waveform is said to be *periodic* in the time record.

Figure 3.24 demonstrates the difficulty with this assumption when the input is not periodic in the time record. The FFT algorithm is computed on the basis of the highly distorted waveform in Figure 3.24c.

We know from Chapter 2 that the actual sine wave input has a frequency spectrum of single line. The spectrum of the input assumed by the FFT in Figure 3.24c should be

Figure 3.22 FFT assumption - time record repeated throughout all time.

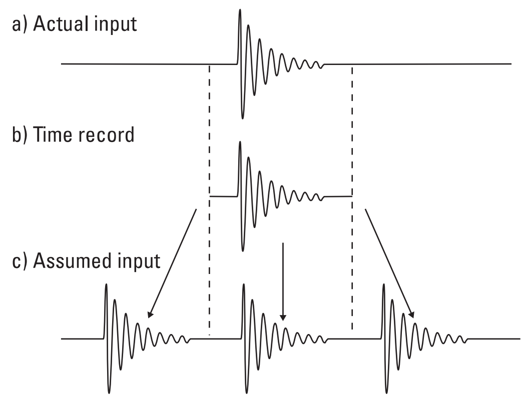


Figure 3.23 Input signal periodic in time record.

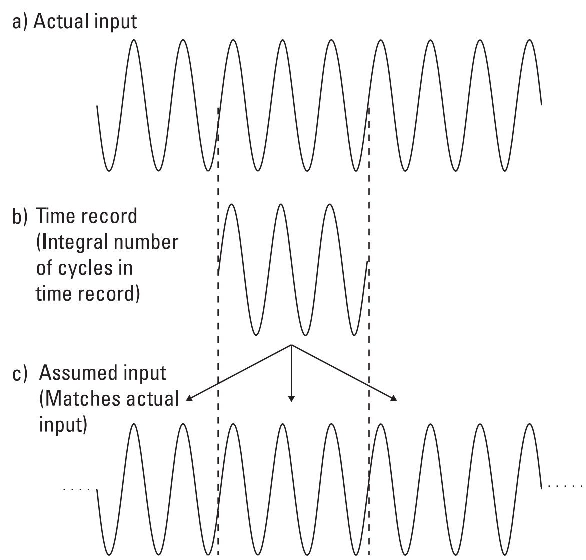
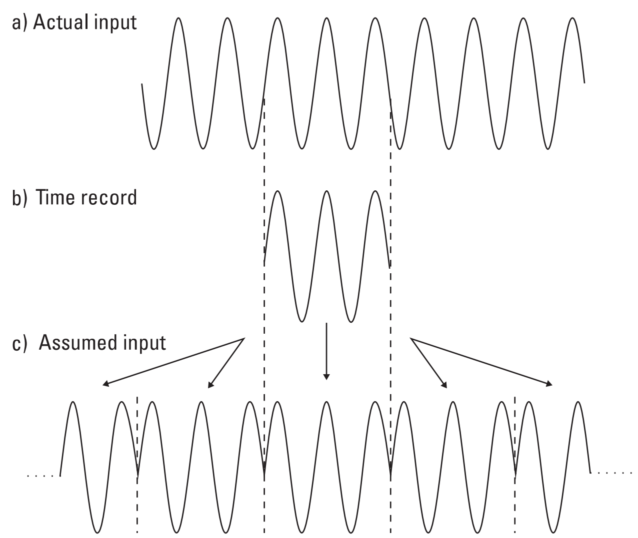


Figure 3.24 Input signal not periodic in time record.



very different. Since sharp phenomena in one domain are spread out in the other domain, we would expect the spectrum of our sine wave to be spread out through the frequency domain.

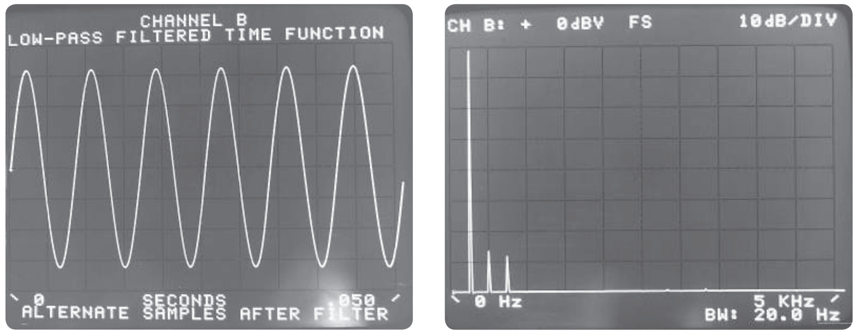
In Figure 3.25 we see in an actual measurement that our expectations are correct. In Figures 3.25 a & b, we see a sine wave that is periodic in the time record. Its frequency spectrum is a single line whose width is determined only by the resolution of our Dynamic Signal Analyzer.* On the other hand, Figures 3.25c & d show a sine wave that is not periodic in the time record. Its power has been spread throughout the spectrum as we predicted.

This smearing of energy throughout the frequency domains is a phenomena known as *leakage*. We are seeing energy leak out of one resolution line of the FFT into all the other lines.

It is important to realize that leakage is due to the fact that we have taken a finite time record. For a sine wave to have a single line spectrum, it must exist for all time, from minus infinity to plus infinity. If we were to have an infinite time record, the FFT would compute the correct single line spectrum exactly. However, since we are not willing to wait forever to measure its spectrum, we only look at a finite time record of the sine wave. This can cause leakage if the continuous input is not periodic in the time record.

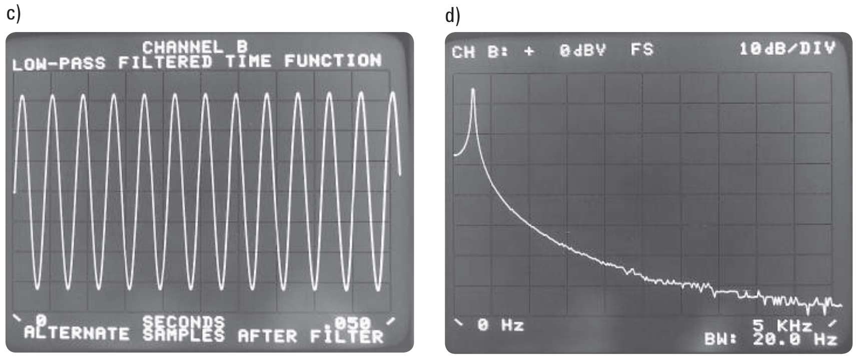
Figure 3.25
Actual FFT results.

a) b)



a) & b) Sine wave periodic in time record

c) d)



c) & d) Sine wave not periodic in time record

It is obvious from Figure 3.25 that the problem of leakage is severe enough to entirely mask small signals close to our sine waves. As such, the FFT would not be a very useful spectrum analyzer. The solution to this problem is known as windowing. The problems of leakage and how to solve them with windowing can be the most confusing concepts of Dynamic Signal Analysis. Therefore, we will now carefully develop the problem and its solution in several representative cases.

* The additional two components in the photo are the harmonic distortion of the sine wave source.

What is Windowing?

In Figure 3.26 we have again reproduced the assumed input wave form of a sine wave that is not periodic in the time record. Notice that most of the problem seems to be at the edges of the time record, the center is a good sine wave. If the FFT could be made to ignore the ends and concentrate on the middle of the time record, we would expect to get much closer to the correct single line spectrum in the frequency domain.

If we multiply our time record by a function that is zero at the ends of the time record and large in the middle, we would concentrate the FFT on the middle of the time record. One such function is shown in Figure 3.26c. Such functions are called window functions because they force us to look at data through a narrow window.

Figure 3.27 shows us the vast improvement we get by windowing data that is not periodic in the time record. However, it is important to realize that we have tampered with the input data and cannot expect perfect results. The FFT assumes the input looks like Figure 3.26d, something like an amplitude-modulated sine wave. This has a frequency spectrum which is closer to the correct single line of the input sine wave than Figure 3.26b, but it still is not correct. Figure 3.28 demonstrates that the windowed data does not have as narrow a spectrum as an unwindowed function which is periodic in the time record.

Figure 3.26
The effect of windowing in the time domain.

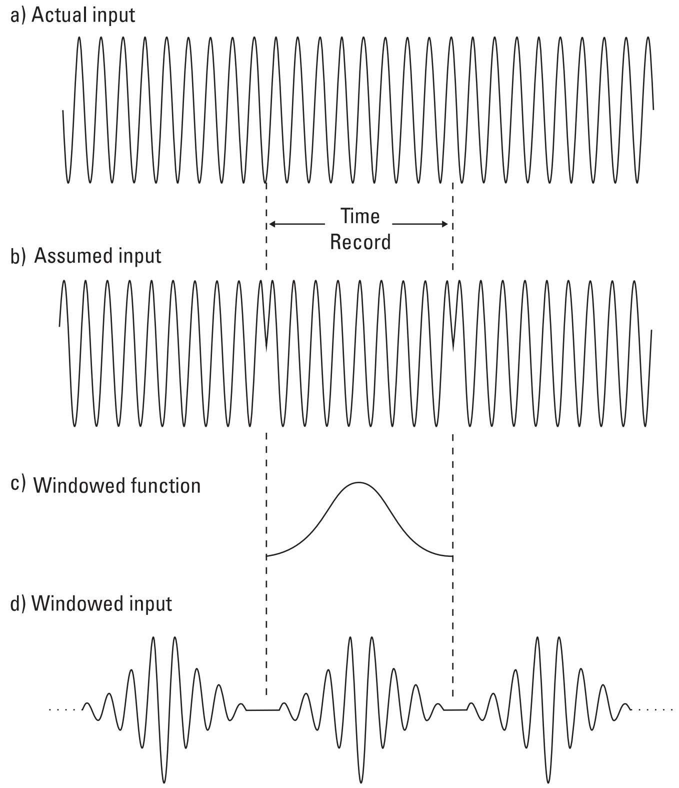
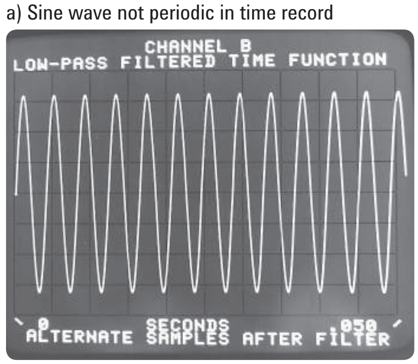
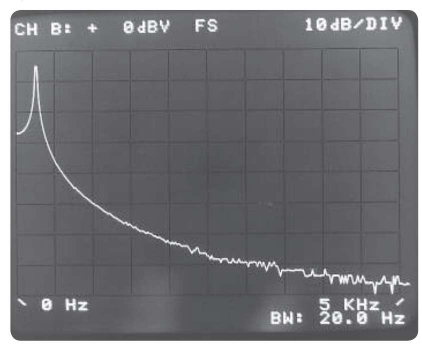


Figure 3.27
Leakage reduction with windowing.

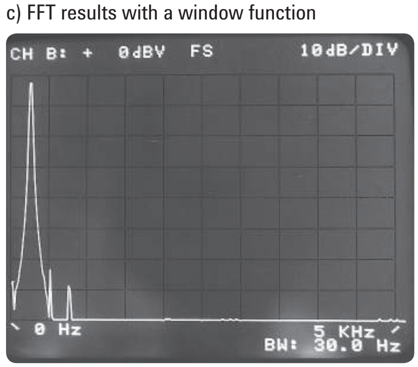
a) Sine wave not periodic in time record



b) FFT results with no window function



c) FFT results with a window function



The Hanning Window

Any number of functions can be used to window the data, but the most common one is called Hanning. We actually used the Hanning window in Figure 3.27 as our example of leakage reduction with windowing. The Hanning window is also commonly used when measuring random noise.

The Uniform Window*

We have seen that the Hanning window does an acceptably good job on our sine wave examples, both periodic and non-periodic in the time record. If this is true, why should we want any other windows?

Suppose that instead of wanting the frequency spectrum of a continuous signal, we would like the spectrum of a transient event. A typical transient is shown in Figure 3.29a. If we multiplied it by the window function in Figure 3.29b we would get the highly distorted signal shown in Figure 3.29c. The frequency spectrum of an actual transient with and without the Hanning window is shown in Figure 3.30. The Hanning window has taken our transient, which naturally has energy spread widely through the frequency domain and made it look more like a sine wave.

Therefore, we can see that for transients we do not want to use the Hanning window. We would like to use all the data in the time record equally or uniformly. Hence we will use the Uniform window which weights all of the time record uniformly.

The case we made for the Uniform window by looking at transients can be generalized. Notice that our transient has the property that it is zero at the beginning and end of the time record. Remember that we introduced windowing to force the input to be zero *at the ends of the time record*. In this case, there is no need for windowing the input. Any function like this which does not require a window because it occurs completely within the time record is called a *self-windowing function*. Self-windowing functions generate no leakage in the FFT and so need no window.

Figure 3.28
Windowing reduces leakage but does not eliminate it.

a) Leakage-free measurement - input periodic in time record

b) Windowed measurement - input not periodic in time record

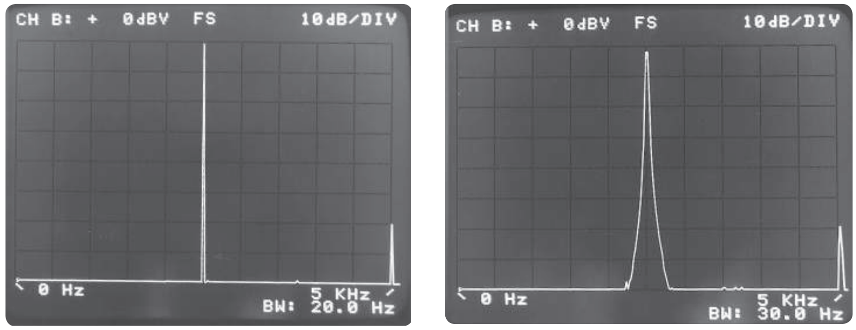


Figure 3.29
Windowing loses information from transient events.

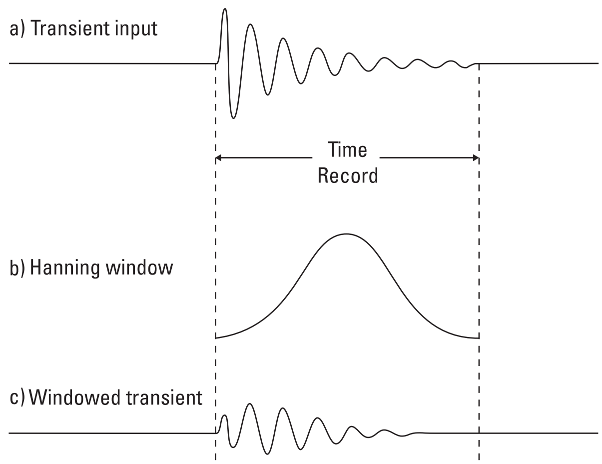
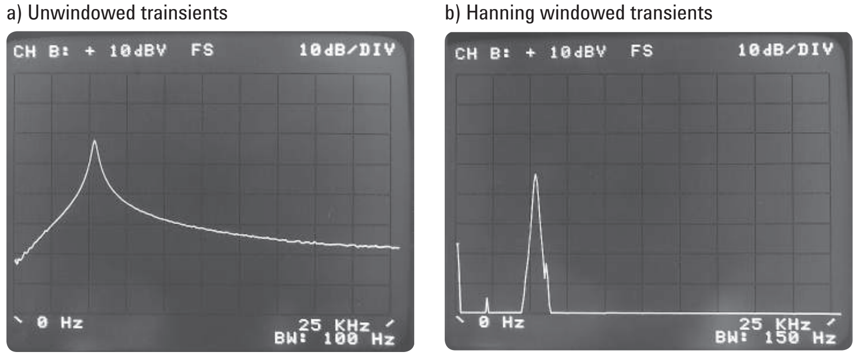


Figure 3.30
Spectrums of transients.

a) Unwindowed trainsients

b) Hanning windowed transients



* The Uniform Window is sometimes referred to as a "Rectangular Window".

There are many examples of self-windowing functions, some of which are shown in Figure 3.31. Impacts, impulses, shock responses, sine bursts, noise bursts, chirp bursts and pseudo-random noise can all be made to be self-windowing. Self-windowing functions are often used as the excitation in measuring the frequency response of networks, particularly if the network has lightly-damped resonances (high Q). This is because the self-windowing functions generate no leakage in the FFT. Recall that even with the Hanning window, some leakage was present when the signal was not periodic in the time record. This means that without a self-windowing excitation, energy could leak from a lightly damped resonance into adjacent lines (filters). The resulting spectrum would show greater damping than actually exists.*

Figure 3.31 Self-windowing function examples.

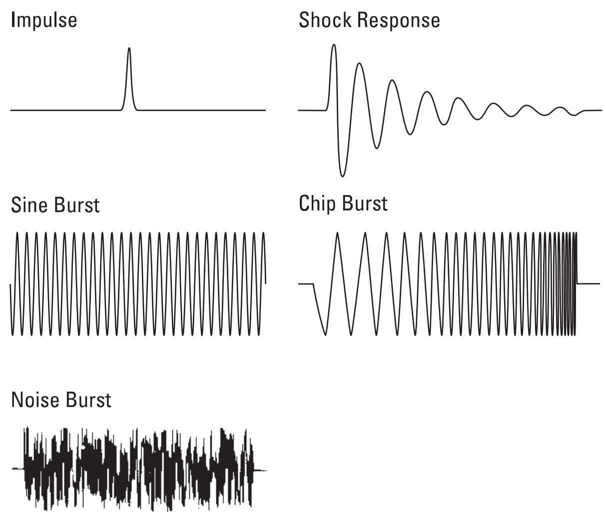
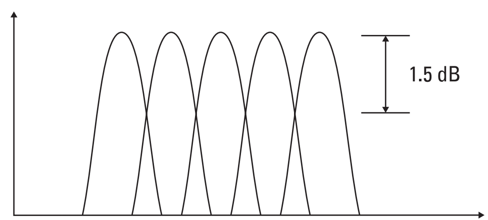


Figure 3.32 Hanning passband shapes.



The Flat-top Window

We have shown that we need a uniform window for analyzing self-windowing functions like transients. In addition, we need a Hanning window for measuring noise and periodic signals like sine waves.

We now need to introduce a third window function, the *flat-top window*, to avoid a subtle effect of the Hanning window. To understand this effect, we need to look at the Hanning window in the frequency domain. We recall that the FFT acts like a set of parallel filters. Figure 3.32 shows the shape of those filters when the Hanning window is used. Notice that the Hanning function gives the filter a very rounded top. If a component of the input signal is centered in the filter it will be measured accurately**. Otherwise, the filter shape will attenuate the component by up to 1.5 dB (16%) when it falls midway between the filters.

This error is unacceptably large if we are trying to measure a signal's amplitude accurately. The solution is to choose a window function which gives the filter a flatter passband. Such a flat-top passband shape is shown in Figure 3.33. The amplitude error from this window function does not exceed .1 dB (1%), a 1.4 dB improvement.

Figure 3.33 Flat-top passband shapes.

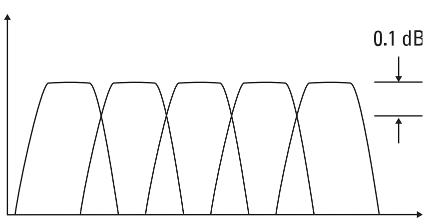
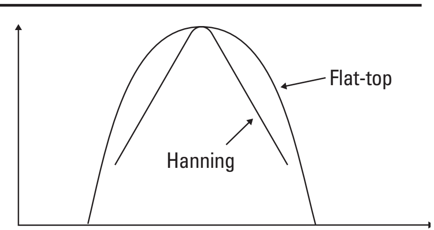


Figure 3.34 Reduced resolution of the flat-top window.



* There is another way to avoid this problem using Band Selectable Analysis. We will illustrate this in the next chapter.

** It will, in fact, be periodic in the time record

The accuracy improvement does not come without its price, however. Figure 3.34 shows that we have flattened the top of the passband at the expense of widening the skirts of the filter. We therefore lose some ability to resolve a small component, closely spaced to a large one. Some Dynamic Signal Analyzers offer both Hanning and flat-top window functions so that the operator can choose between increased accuracy or improved frequency resolution.

Other Window Functions

Many other window functions are possible but the three listed above are by far the most common for general measurements. For special measurement situations other groups of window functions may be useful. We will discuss two windows which are particularly useful when doing network analysis on mechanical structures by impact testing.

The Force and Response Windows

A hammer equipped with a force transducer is commonly used to stimulate a structure for response measurements. Typically the force input is connected to one channel of the analyzer and the response of the structure from another transducer is connected to the second channel. This force impact is obviously a self-windowing function. The response of the structure is also self-windowing if it dies out within the time record of the analyzer. To guarantee that the response does go to zero by the end of the time record, an exponential-weighted window called a response window is sometimes added. Figure 3.35 shows a response window acting on the response of a lightly damped structure which did not fully decay by the end of the time record. Notice that unlike the Hanning window, the response window is not zero at both ends of the time record. We know that the response of the structure will be zero at the beginning of the time record (before the hammer blow) so there is no need for the window function to be zero there. In addition, most of the information about the structural response is contained at the beginning of the time record so we make sure that this is weighted most heavily by our response window function.

Figure 3.35 Using the response window.

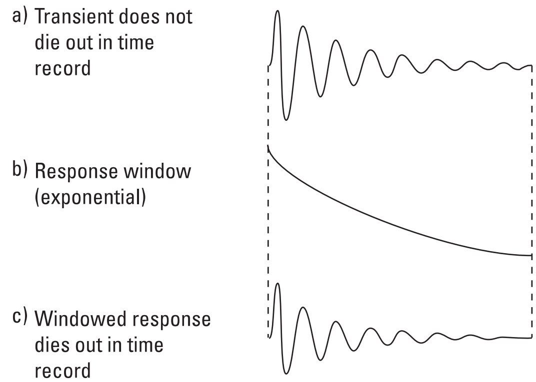
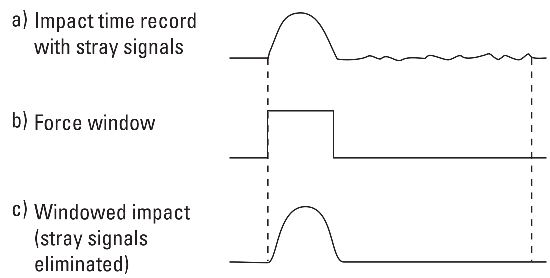


Figure 3.36 Using the force window.



The time record of the exciting force should be just the impact with the structure. However, movement of the hammer before and after hitting the structure can cause stray signals in the time record. One way to avoid this is to use a force window shown in Figure 3.36. The force window is unity where the impact data is valid and zero everywhere else so that the analyzer does not measure any stray noise that might be present.

Passband Shapes or Window Functions?

In the proceeding discussion we sometimes talked about window functions in the time domain. At other times we talked about the filter passband shape in the frequency domain caused by these windows. We change our perspective freely to whichever domain yields the simplest explanation. Likewise, some Dynamic Signal Analyzers call the uniform, Hanning and flat-top functions “windows” and other analyzers call those functions "pass-band shapes". Use whichever terminology is easier for the problem at hand as they are completely interchangeable, just as the time and frequency domains are completely equivalent.

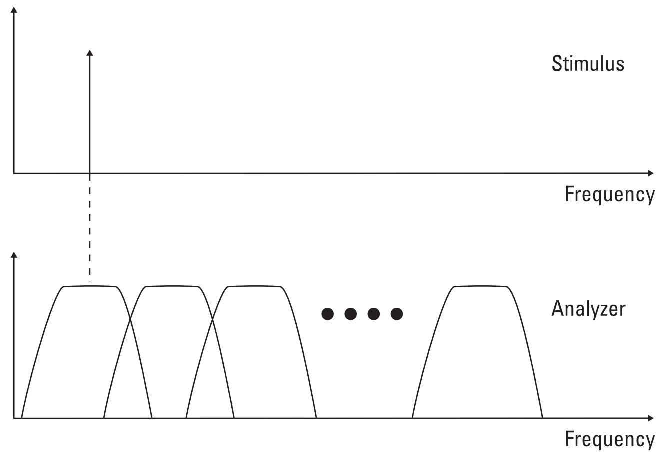


Figure 3.37 Frequency response measurements with a sine wave stimulus.

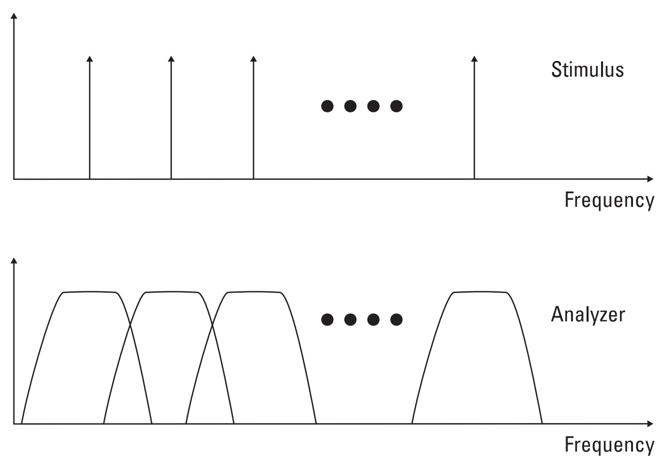


Figure 3.38 Pseudo-random noise as a stimulus.

Section 6: Network Stimulus

Recall from Chapter 2 that we can measure the frequency response at one frequency by stimulating the network with a single sine wave and measuring the gain and phase shift at that frequency. The frequency of the stimulus is then changed and the measurement repeated until all desired frequencies have been measured. Every time the frequency is changed, the network response must settle to its steady-state value before a new measurement can be taken, making this measurement process a slow task.

Many network analyzers operate in this manner and we can make the measurement this way with a two channel Dynamic Signal Analyzer. We set the sine wave source to the center of the first filter as in Figure 3.37. The analyzer then measures the gain and phase of the network at this frequency while the rest of the analyzer's filters measure only noise. We then increase the source frequency to the next filter center, wait for the network to settle and then measure the gain and phase. We continue this procedure until we have measured the gain and phase of the network at all the frequencies of the filters in our analyzer.

This procedure would, within experimental error, give us the same results as we would get with any of the network analyzers described in Chapter 2 *with any network, linear or nonlinear.*

Noise as a Stimulus

A single sine wave stimulus does not take advantage of the possible speed the parallel filters of a Dynamic Signal Analyzer provide. If we had a source that put out multiple sine waves, each one centered in a filter, then we could measure the frequency response at all frequencies at one time. Such a source, shown in Figure 3.38, acts like hundreds of sine wave generators connected together. Although this sounds very expensive, just such a source can be easily generated digitally. It is called a pseudo-random noise or periodic random noise source.

From the names used for this source it is apparent that it acts somewhat like a true noise generator, except that it has periodicity. If we add together a large number of sine waves, the result is very much like white noise. A good analogy is the sound of rain. A single drop of water makes a quite distinctive splashing sound, but a rain storm sounds like white noise. However, if we add together a large number of sine waves, our noise-like signal will periodically repeat its sequence. Hence, the name periodic random noise (PRN) source.

A truly random noise source has a spectrum shown in Figure 3.39. It is apparent that a random noise source would also stimulate all the filters at one time and so could be used as a network stimulus. Which is a better stimulus? The answer depends upon the measurement situation.

Linear Network Analysis

If the network is reasonably linear, PRN and random noise both give the same results as the swept-sine test of other analyzers. But PRN gives the frequency response much faster. PRN can be used to measure the frequency response in a single time record. Because the random source is true noise, it must be averaged for several time records before an accurate frequency response can be determined. Therefore, PRN is the best stimulus to use with fairly linear networks because it gives the fastest results*.

Figure 3.39
Random noise as a stimulus.

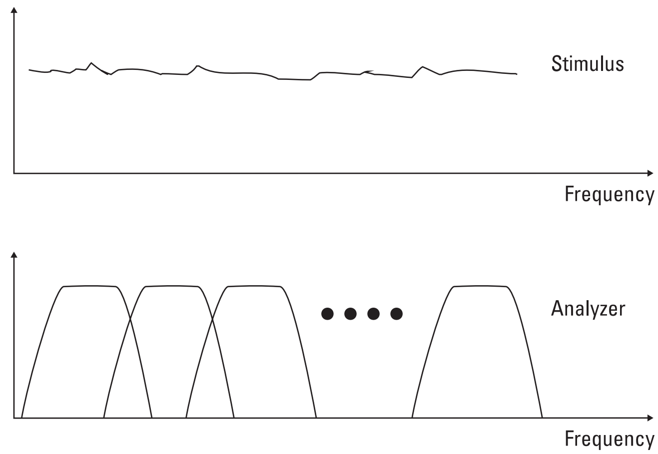
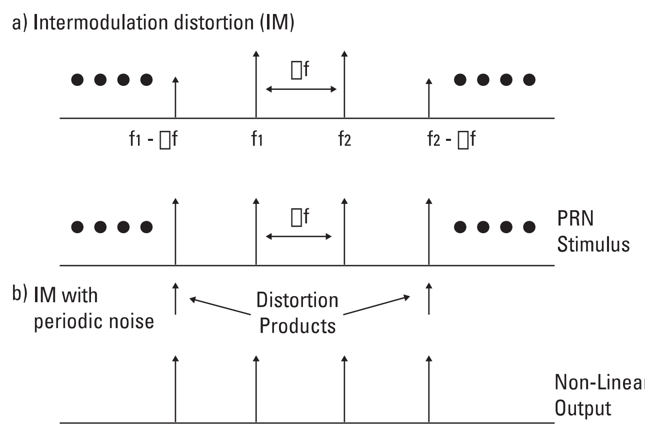


Figure 3.40
Pseudo-random noise distortion.



Non-Linear Network Analysis

If the network is severely non-linear, the situation is quite different. In this case, PRN is a very poor test signal and random noise is much better. To see why, let us look at just two of the sine waves that compose the PRN source. We see in Figure 3.40 that if two sine waves are put through a nonlinear network, distortion products will be generated equally spaced from the signals**. Unfortunately, these products will fall exactly on the frequencies of the other sine waves in the PRN. So the distortion products add to the output and therefore interfere with the measurement

* There is another reason why PRN is a better test signal than random or linear networks. Recall from the last section that PRN is self-windowing. This means that unlike random noise, pseudo-random noise has no leakage. Therefore, with PRN, we can measure lightly damped (high Q) resonances more easily than with random noise.

** This distortion is called intermodulation distortion.

of the frequency response. Figure 3.41a shows the jagged response of a nonlinear network measured with PRN. Because the PRN source repeats itself exactly every time record, this noisy looking trace never changes and will not average to the desired frequency response.

With random noise, the distortion components are also random and will average out. Therefore, the frequency response does not include the distortion and we get the more reasonable results shown in Figure 3.41b.

This points out a fundamental problem with measuring non-linear networks; *the frequency response is not a property of the network alone, it also depends on the stimulus*. Each stimulus, swept-sine, PRN and random noise will, in general, give a different result. Also, if the amplitude of the stimulus is changed, you will get a different result.

To illustrate this, consider the mass-spring system with stops that we used in Chapter 2. If the mass does not hit the stops, the system is linear and the frequency response is given by Figure 3.42a.

If the mass does hit the stops, the output is clipped and a large number of distortion components are generated. As the output approaches a square wave, the fundamental component becomes constant. Therefore, as we increase the input amplitude, the gain of the network drops. We get a frequency response like Figure 3.42b, where the gain is dependent on the input signal amplitude.

Figure 3.41
Nonlinear transfer function.

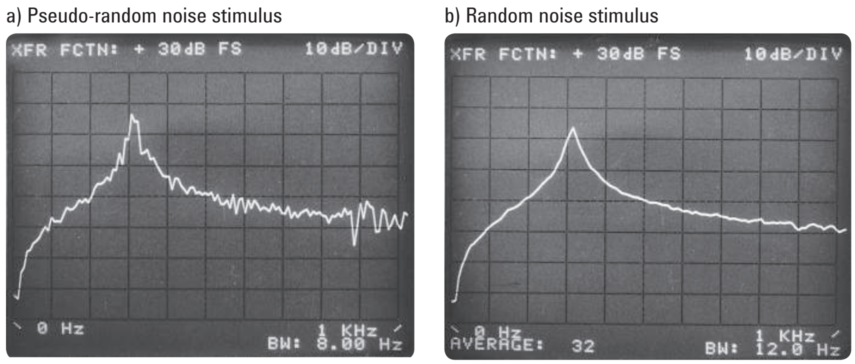
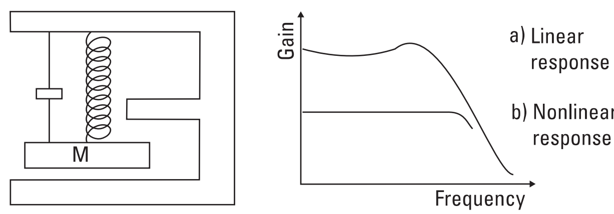


Figure 3.42
Nonlinear system.



So as we have seen, the frequency response of a nonlinear network is not well defined, i.e., it depends on the stimulus. Yet it is often used in spite of this. The frequency response of linear networks has proven to be a very powerful tool and so naturally people have tried to extend it to non-linear analysis, particularly since other nonlinear analysis tools have proved intractable.

If every stimulus yields a different frequency response, which one should we use? The "best" stimulus could be considered to be one which approximates the kind of signals you would expect to have as normal inputs to the network. Since any large collection of signals begins to look like noise, noise is a good test signal*. As we have already explained, noise is also a good test signal because it speeds the analysis by exciting all the filters of our analyzer simultaneously.

But many other test signals can be used with Dynamic Signal Analyzers and are "best" (optimum) in other senses. As explained in the beginning of this section, sine waves can be used to give the same results as other types of network analyzers although the speed advantage of the Dynamic Signal Analyzer is lost. A fast sine sweep (chirp) will give very similar results with all the speed of Dynamic Signal Analysis and so is a better test signal. An impulse is a good test signal for acoustical testing if the network is linear. It is good for acoustics because reflections from surfaces at different distances can easily be isolated or eliminated if desired. For instance, by using the "force" window described earlier, it is easy to get the free field response of a speaker by eliminating the room reflections from the windowed time record.

Band-Limited Noise

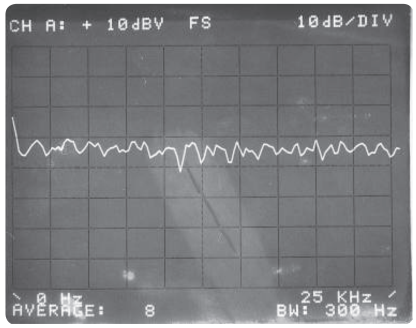
Before leaving the subject of network stimulus, it is appropriate to discuss the need to band limit the stimulus. We want all the power of the stimulus to be concentrated in the frequency region we are analyzing. Any power

* This is a consequence of the central limit theorem. As an example, the telephone companies have found that when many conversations are transmitted together, the result is like white noise. The same effect is found more commonly at a crowded cocktail party.

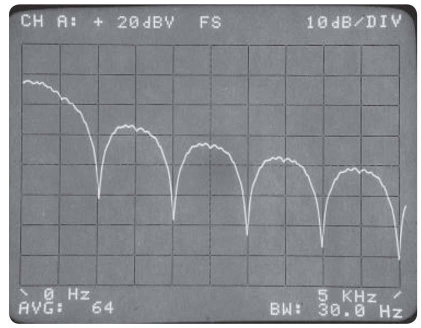
outside this region does not contribute to the measurement and could excite non-linearities. This can be a particularly severe problem when testing with random noise since it theoretically has the same power at all frequencies (white noise). To eliminate this problem, Dynamic Signal Analyzers often limit the frequency range of their built-in noise stimulus to the frequency span selected. This could be done with an external noise source and filters, but every time the analyzer span changed, the noise power and filter would have to be readjusted. This is done automatically with a built-in noise source so transfer function measurements are easier and faster.

Figure 3.43
RMS averaged spectra.

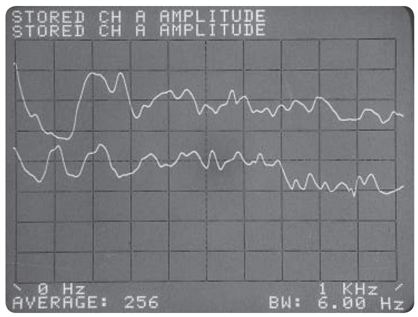
a) Random noise



b) Digital data



c) Voices



Traces were separated 30 dB for clarity

Upper trace: female speaker
Lower trace: male speaker

Section 7: Averaging

To make it as easy as possible to develop an understanding of Dynamic Signal Analyzers we have almost exclusively used examples with deterministic signals, i.e., signals with no noise. However, as the real world is rarely so obliging, the desired signal often must be measured in the presence of significant noise. At other times the "signals" we are trying to measure are more like noise themselves. Common examples that are somewhat noise-like include speech, music, digital data, seismic data and mechanical vibrations. Because of these two common conditions, we must develop techniques both to measure signals in the presence of noise and to measure the noise itself.

The standard technique in statistics to improve the estimates of a value is to average. When we watch a noisy reading on a Dynamic Signal Analyzer, we can guess the average value. But because the Dynamic Signal Analyzer contains digital computation capability we can have it compute this average value for us. Two kinds of averaging are available, RMS (or "power" averaging) and linear averaging.

RMS Averaging

When we watch the magnitude of the spectrum and attempt to guess the average value of the spectrum component, we are doing a crude RMS* average. We are trying to determine the average magnitude of the signal, ignoring any phase difference that may exist between the spectra. This averaging technique is very valuable for determining the average power in any of the filters of our Dynamic Signal Analyzers. The more averages we take, the better our estimate of the power level.

In Figure 3.43, we show RMS averaged spectra of random noise, digital data and human voices. Each of these examples is a fairly random process, but when averaged we can see the basic properties of its spectrum.

If we want to measure a small signal in the presence of noise, RMS averaging will give us a good estimate of the signal plus noise. We can not improve the signal to noise ratio with RMS averaging; we can only make more accurate estimates of the total signal plus noise power.

* RMS stands for "root-mean-square" and is calculated by squaring all the values, adding the squares together, dividing by the number of measurements (mean) and taking the square root of the result.

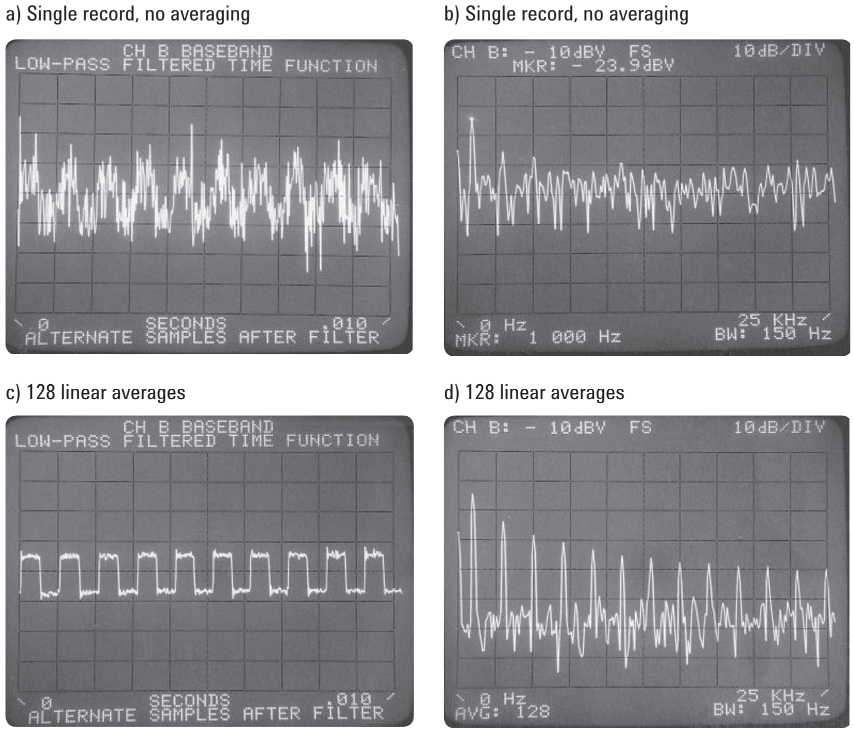
Linear Averaging

However, there is a technique for improving the signal to noise ratio of a measurement, called *linear averaging*. It can be used if a trigger signal which is synchronous with the periodic part of the spectrum is available. Of course, the need for a synchronizing signal is somewhat restrictive, although there are numerous situations in which one is available. In network analysis problems the stimulus signal itself can often be used as a synchronizing signal.

Linear averaging can be implemented many ways, but perhaps the easiest to understand is where the averaging is done in the time domain. In this case, the synchronizing signal is used to trigger the start of a time record. Therefore, the periodic part of the input will always be exactly the same in each time record we take, whereas the noise will, of course, vary. If we add together a series of these triggered time records and divide by the number of records we have taken we will compute what we call a linear average.

Since the periodic signal will have repeated itself exactly in each time record, it will average to its exact value. But since the noise is different in each time record, it will tend to average to zero. The more averages we take, the closer the noise comes to zero and we continue to improve the signal to noise ratio of our measurement. Figure 3.44 shows a time record of a square wave buried in noise. The resulting time record after 128 averages shows a marked improvement in the signal to noise ratio. Transforming both results to the frequency domain shows how many of the harmonics can now be accurately measured because of the reduced noise floor.

Figure 3.44
Linear averaging.



Section 8: Real Time Bandwidth

Until now we have ignored the fact that it will take a finite time to compute the FFT of our time record. In fact, if we could compute the transform in less time than our sampling period we could continue to ignore this computational time. Figure 3.45 shows that under this condition we could get a new frequency spectrum with every sample. As we have seen from the section on aliasing, this could result in far more spectrums every second than we could possibly comprehend. Worse, because of the complexity of the FFT algorithm, it would take a very fast and very expensive computer to generate spectrums this rapidly.

A reasonable alternative is to add a time record buffer to the block diagram of our analyzer. In Figure 3.47 we can see that this allows us to compute the frequency spectrum of the previous time record while gathering the current time record. If we can compute the transform before the time record buffer fills, then we are said to be *operating in real time*.

To see what this means, let us look at the case where the FFT computation takes longer than the time to fill the time record. The case is illustrated in Figure 3.48. Although the buffer is full, we have not finished the last transform, so we will have to stop taking data. When the transform is finished, we can transfer the time record to the FFT and begin to take another time record. This means that we missed some input data and so we are said to be *not operating in real time*.

Recall that the time record is not constant but deliberately varied to change the frequency span of the analyzer. For wide frequency spans the time record is shorter. Therefore, as we increase the frequency span of the analyzer, we eventually reach a span where the time record is equal to the FFT computation time. This frequency span is called the *real time bandwidth*. For frequency spans at and below the real time bandwidth, the analyzer does not miss any data.

Real Time Bandwidth Requirements

How wide a real time bandwidth is needed in a Dynamic Signal Analyzer? Let us examine a few typical measurements to get a feeling for the considerations involved.

Adjusting Devices

If we are measuring the spectrum or frequency response of a device which we are adjusting, we need to watch the spectrum change in what might be called *psychological real time*. A new spectrum every few tenths of a second is sufficiently fast to allow an operator to watch adjustments *in what he would consider to be real time*. However, if the response time of the device under test is long, the speed of the analyzer is immaterial. We will have to wait for the device to respond to the changes before the spectrum will be valid, *no matter how many spectrums we generate in that time*. This is what makes adjusting lightly damped (high Q) resonances tedious.

Figure 3.45 A new transform every sample.

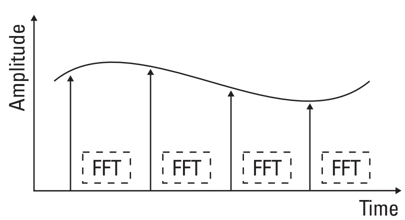


Figure 3.46 Time buffer added to block diagram.

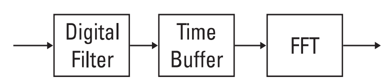


Figure 3.47 Real time operation.

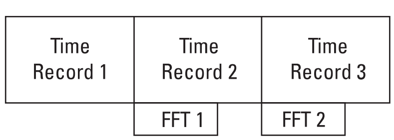
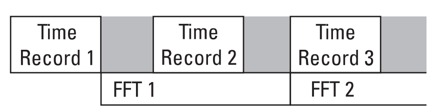


Figure 3.48 Non-real time operation.



RMS Averaging

A second case of interest in determining real time bandwidth requirements is measurements that require RMS averaging. We might be interested in determining the spectrum distribution of the noise itself or in reducing the variation of a signal contaminated by noise. There is no requirement in averaging that the records must be consecutive with no gaps*. Therefore, a small real time bandwidth will not affect the accuracy of the results.

However, the real time bandwidth will affect the speed with which an RMS averaged measurement can be made. Figure 3.49 shows that for frequency spans above the real time bandwidth, the time to complete the average of N records is dependent only on the time to compute the N transforms. Rather than continually reducing the time to compute the RMS average as we increase our span, we reach a fixed time to compute N averages.

Therefore, a small real time bandwidth is only a problem in RMS averaging when large spans are used with a large number of averages. Under these conditions we must wait longer for the answer. Since wider real time bandwidths require faster computations and therefore a more expensive processor, there is a straightforward trade-off of time versus money. In the case of RMS averaging, higher real time bandwidth gives you somewhat faster measurements at increased analyzer cost.

Figure 3.49 RMS averaging time.

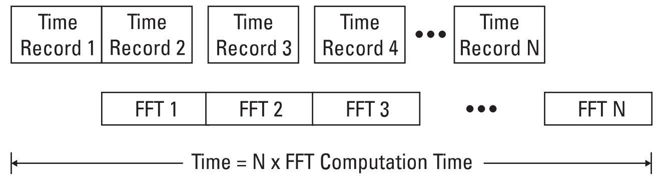
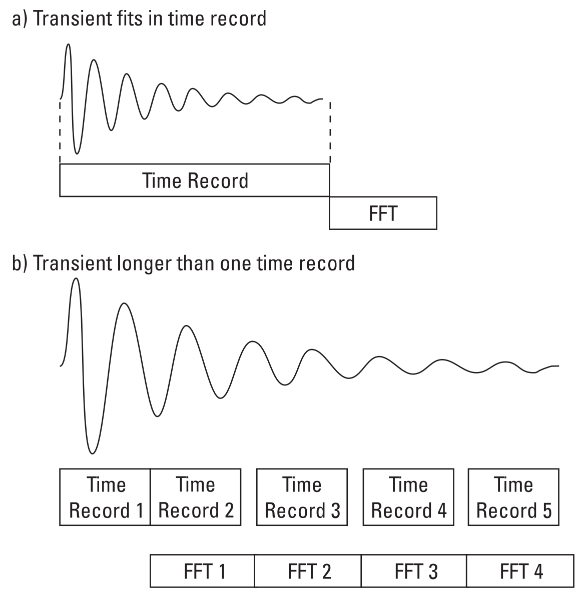


Figure 3.50 Transient analysis.



Transients

The last case of interest in determining the needed real time bandwidth is the analysis of transient events. If the entire transient fits within the time record, the FFT computation time is of little interest. The analyzer can be triggered by the transient and the event stored in the time record buffer. The time to compute its spectrum is not important.

However, if a transient event contains high frequency energy and lasts longer than the time record necessary to measure the high frequency energy, then the processing speed of the analyzer is critical. As shown in Figure 3.50b, some of the transient will not be analyzed if the computation time exceeds the time record length.

In the case of transients longer than the time record, it is also imperative that there is some way to rapidly record the spectrum. Otherwise, the

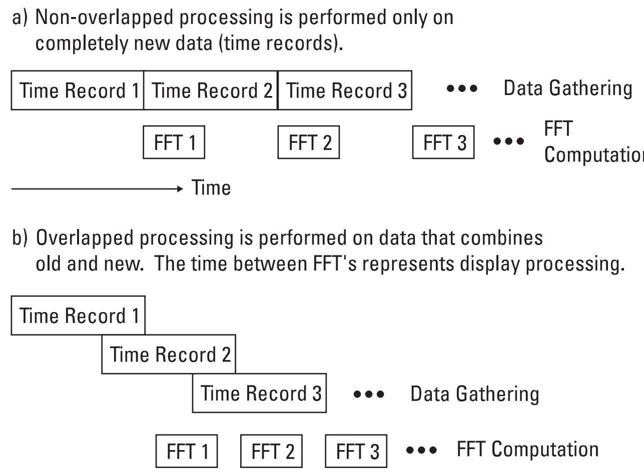
* This is because to average at all the signal must be periodic and the noise stationary.

information will be lost as the analyzer updates the display with the spectrum of the latest time record. A special display which can show more than one spectrum ("waterfall" display), mass memory, a high speed link to a computer or a high speed facsimile recorder is needed. The output device must be able to record a spectrum every time record or information will be lost.

Fortunately, there is an easy way to avoid the need for an expensive wide real time bandwidth analyzer and an expensive, fast spectrum recorder. One-time transient events like explosions and pass-by noise are usually tape recorded for later analysis because of the expense of repeating the test. If this tape is played back at reduced speed, the speed demands on the analyzer and spectrum recorder are reduced. Timing markers could also be recorded at one time record intervals. This would allow the analysis of one record at a time and plotting with a very slow (and commonly available) X-Y plotter.

So we see that there is no clear-cut answer to what real time bandwidth is necessary in a Dynamic Signal Analyzer. Except in analyzing long transient events, the added expense of a wide real time bandwidth gives little advantage. It is possible to analyze long transient events with a narrow real time bandwidth analyzer, but it does require the recording of the input signal. This method is slow and requires some operator care, but one can avoid purchasing an expensive analyzer and fast spectrum recorder. It is a clear case of speed of analysis versus dollars of capital equipment.

Figure 3.51 Understanding overlap processing.



Section 9: Overlap Processing

In Section 8 we considered the case where the computation of the FFT took longer than the collecting of the time record. In this section we will look at a technique, overlap processing, which can be used when the FFT computation takes less time than gathering the time record.

To understand overlap processing, let us look at Figure 3.51a. We see a low frequency analysis where the gathering of a time record takes much longer than the FFT computation time. Our FFT processor is sitting idle much of the time. If instead of waiting for an entirely new time record we *overlapped* the new time record with some of the old data, we would get a new spectrum as often as we computed the FFT. This *overlap processing* is illustrated in Figure 3.51b. To understand the benefits of overlap processing, let us look at the same cases we used in the last section.

Adjusting Devices

We saw in the last section that we need a new spectrum every few tenths of a second when adjusting devices. Without overlap processing this limits our resolution to a few Hertz. With overlap processing our resolution is unlimited. But we are not getting something for nothing. Because our overlapped time record contains old data from before the device adjustment, it is not completely correct. It does indicate the direction and the amount of change, but we must wait a full time record after the change for the new spectrum to be accurately displayed.

Nonetheless, by indicating the direction and magnitude of the changes every few tenths of a second, overlap processing does help in the adjustment of devices.

RMS Averaging

Overlap processing can give dramatic reductions in the time to compute RMS averages with a given variance. Recall that window functions reduce the effects of leakage by weighting the ends of the time record to zero. Overlapping eliminates most or all of the time that would be wasted taking this data. Because some overlapped data is used twice, more averages must be taken to get a given variance than in the non-overlapped case. Figure 3.52 shows the improvements that can be expected by overlapping.

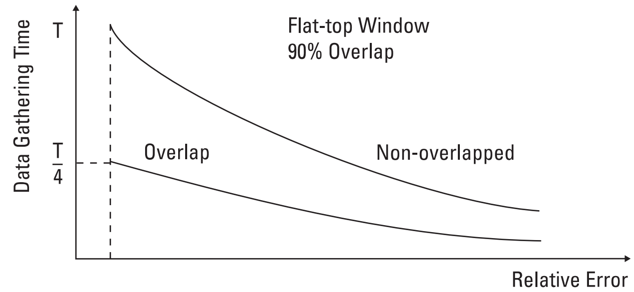


Figure 3.52 RMS averaging speed improvements with overlap processing.

Transients

For transients shorter than the time record, overlap processing is useless. For transients longer than the time record the real time bandwidth of the analyzer and spectrum recorder is usually a limitation. If it is not, overlap processing allows more spectra to be generated from the transient, usually improving resolution of resulting plots.

Section 10: Summary

In this chapter we have developed the basic properties of Dynamic Signal Analyzers. We found that many properties could be understood by considering what happens when we transform a finite, sampled time record. The length of this record determines how closely our filters can be spaced in the frequency domain and the number of samples determines the number of filters in the frequency domain. We also found that unless we filtered the input we could have errors due to aliasing and that finite time records could cause a problem called leakage which we minimized by windowing.

We then added several features to our basic Dynamic Signal Analyzer to enhance its capabilities. Band Selectable Analysis allows us to make high resolution measurements even at high frequencies. Averaging gives more accurate measurements when noise is present and even allows us to improve the signal to noise ratio when we can use linear averaging. Finally, we incorporated a noise source in our analyzer to act as a stimulus for transfer function measurements.

Chapter 4
Using Dynamic Signal Analyzers

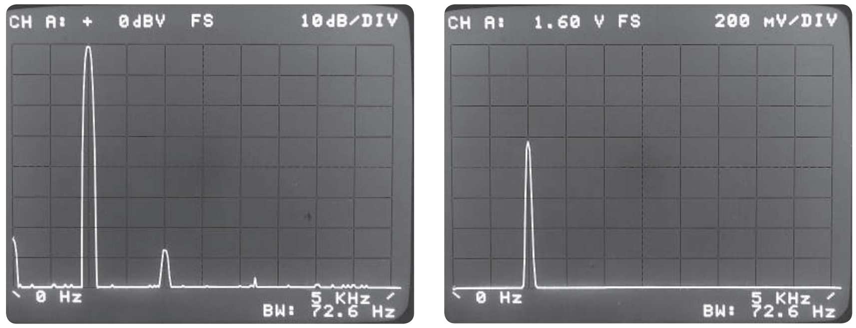
In Chapters 2 & 3, we developed an understanding of the time, frequency and modal domains and how Dynamic Signal Analyzers operate. In this chapter we show how to use Dynamic Signal Analyzers in a wide variety of measurement situations. We introduce the measurement functions of Dynamic Signal Analyzers as we need them for each measurement situation.

We begin with some common electronic and mechanical measurements in the frequency domain. Later in the chapter we introduce time and modal domain measurements.

Figure 4.1
Harmonic distortion of an Audio Oscillator - Flat-top window used.

a) Logarithmic amplitude scale

b) Linear amplitude scale



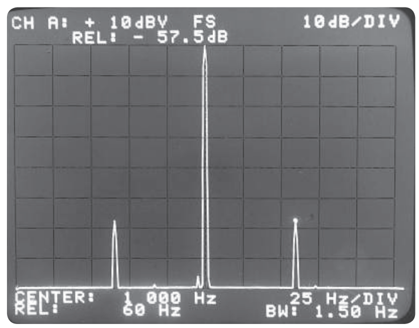
Section 1: Frequency Domain Measurements

Oscillator Characterization

Let us begin by measuring the characteristics of an electronic oscillator. An important specification of an oscillator is its harmonic distortion. In Figure 4.1, we show the fundamental through fifth harmonic of a 1 KHz oscillator. Because the frequency is not necessarily exactly 1 KHz, windowing should be used to reduce the leakage. We have chosen the flat-top window so that we can accurately measure the amplitudes.

Notice that we have selected the input sensitivity of the analyzer so that the fundamental is near the top of the display. In general, we set the input sensitivity to the most sensitive range which does not overload the analyzer. Severe distortion of the input signal will occur if its peak voltage exceeds the range of the analog to digital converter. Therefore, all dynamic signal analyzers warn the user of this condition by some kind of overload indicator.

Figure 4.2
Powerline sidebands of an Audio Oscillator - Band Selectable Analysis and Hanning window used for maximum resolution.



It is also important to make sure the analyzer is not underloaded. If the signal going into the analog to digital converter is too small, much of the useful information of the spectrum may be below the noise level of the analyzer. Therefore, setting the input sensitivity to the most sensitive range that does not cause an overload gives the best possible results.

In Figure 4.1a we chose to display the spectrum amplitude in logarithmic form to insure that we could see distortion products far below the fundamental. All signal amplitudes on this display are in dBV, decibels below 1 Volt RMS. However, since most Dynamic Signal Analyzers have very versatile display capabilities, we could also display this spectrum linearly as in Figure 4.1b. Here the units of amplitude are volts.

Power-Line Sidebands

Another important measure of an oscillator's performance is the level of its power-line sidebands. In Figure 4.2, we use Band Selectable Analysis to "zoom in" on the signal so that we can easily resolve and measure the sidebands which are only 60 Hz away from our 1 KHz signal. With some analyzers it is possible to measure signals only millihertz away from the fundamental if desired.

Phase Noise

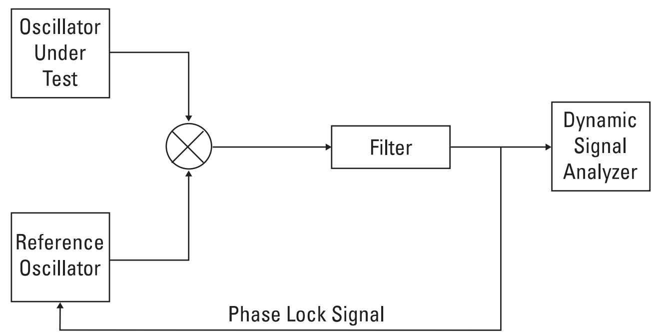
The short-term stability of a high frequency oscillator is very important in communications and radar. One measure of this is called phase noise. It is often measured by the technique shown in Figure 4.3a. This mixes down and cancels the oscillator carrier leaving only the phase noise sidebands. It is therefore possible to measure the phase noise far below the carrier level since the carrier does not limit the range of our measurement. Figure 4.3b shows the close-in phase noise of a 20 MHz synthesizer. Here, since we are measuring noise, we use RMS averaging and the Hanning window.

Dynamic Signal Analyzers offer two main advantages over swept signal analyzers in this application. First, the phase noise can be measured much closer to the carrier. This is because a good swept analyzer can only resolve signals down to about 1 Hz, while a Dynamic Signal Analyzer can resolve signals to a few millihertz. Secondly, the Dynamic Signal Analyzer can determine the complete phase noise spectrum in a few minutes whereas a swept analyzer would take hours.

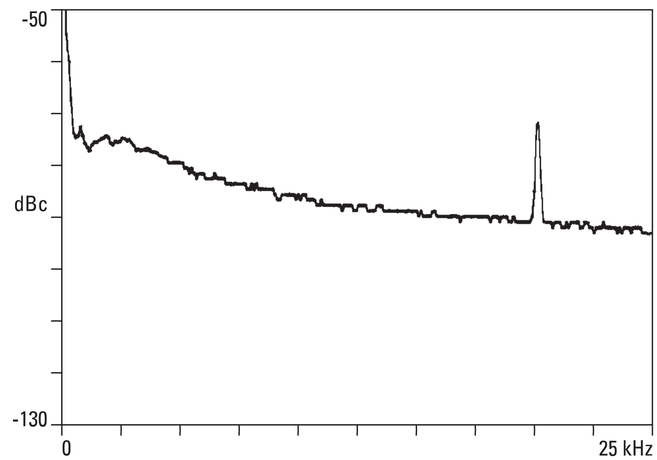
Spectra-like phase noise are usually displayed against the logarithm of frequency instead of the linear frequency scale. This is done in Figure 4.3c. Because the FFT generates linearly spaced filters, the filters are not equally spaced on the display. It is important to realize that no information is missed by these seemingly widely spaced filters. We recall on a linear frequency scale that all the filters overlapped so that no part of the spectrum was missed. All we have done here is to change the presentation of the same measurement.

Figure 4.3 Phase Noise Measurement.

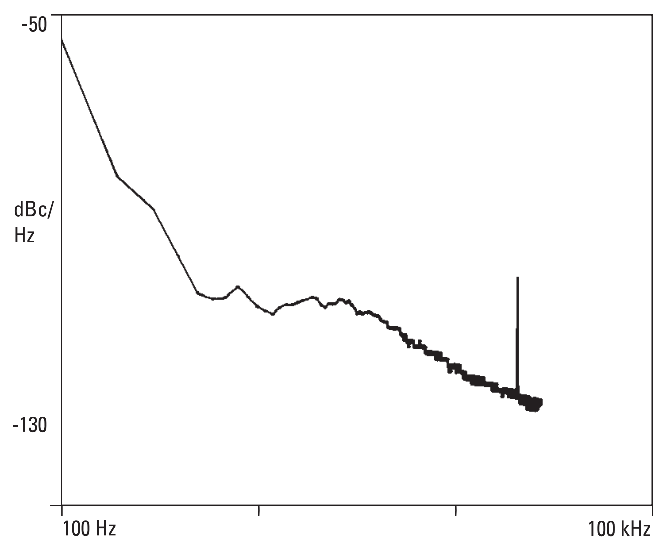
a) Block diagram of phase noise measurement



b) Phase noise of a frequency synthesizer - RMS averaging and Hanning window used for noise measurements



c) Logarithmic frequency axis presentation of phase noise normalized to a 1 Hz bandwidth (power spectral density)



In addition, phase noise and other noise measurements are often normalized to the power that would be measured in a 1 Hz wide square filter. This measurement is called a *power spectral density* and is often provided on Dynamic Signal Analyzers. It simply changes the presentation on the display to this desired form; the data is exactly the same in Figures 4.3b and 4.3c, but the latter is in the more conventional presentation.

Rotating Machinery Characterization

A rotating machine can be thought of as a mechanical oscillator.* Therefore, many of the measurements we made for an electronic oscillator are also important in characterizing rotating machinery.

To characterize a rotating machine we must first change its mechanical vibration into an electrical signal. This is often done by mounting an accelerometer on a bearing housing where the vibration generated by shaft imbalance and bearing imperfections will be the highest. A typical spectrum might look like Figure 4.4.

It is obviously much more complicated than the relatively clean spectrum of the electronic oscillator we looked at previously. There is also a great deal of random noise; stray vibrations from sources other than our motor that the accelerometer picks up. The effects of this stray vibration have been minimized in Figure 4.4b RMS averaging.

In Figure 4.5, we have used the Band Selectable Analysis capability of our analyzer to "zoom-in" and separate the vibration of the stator at 120 Hz from the vibration caused by the rotor imbalance only a few tenths of a Hertz lower in frequency.** This ability to resolve closely spaced spectrum lines is crucial to our capability to diagnose why the vibration levels of a rotating machine are excessive. The actions we would take to correct an excessive vibration at 120 Hz are quite different if it is caused by a loose stator pole rather than an imbalanced rotor.

Since the bearings are the most unreliable part of most rotating machines, we would also like to check our spectrum for indications of bearing failure. Any defect in a bearing, say a spalling on the outer face of a ball bearing, will cause a small vibration to occur each time a ball passes it. This will produce a characteristic frequency in the vibration called the passing frequency. The frequency domain is ideal for

Figure 4.4
Spectrum of electrical motor vibration.

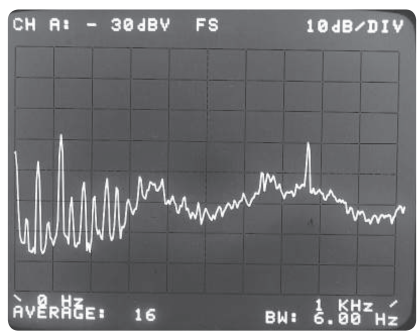


Figure 4.5
Stator vibration and rotor imbalance measurement with Band Selectable Analysis.

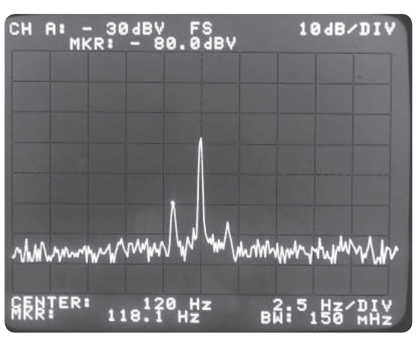
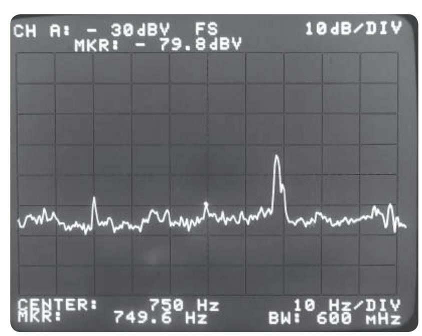


Figure 4.6
Vibration caused by small defect in the bearing.



* Or, if you prefer, electronic oscillators can be viewed as rotating machines which can go at millions of RPM's.

** The rotor in an AC induction motor always runs at a slightly lower frequency than the excitation, an effect called slippage.

separating this small vibration from all the other frequencies present. This means that we can detect impending bearing failures and schedule a shutdown long before they become the loudly squealing problem that signals an immediate shutdown is necessary.

In most rotating machinery monitoring situations, the absolute level of each vibration component is not of interest, just how they change with time. The machine is measured when new and throughout its life and these successive spectra are compared. If no catastrophic failures develop, the spectrum components will increase gradually as the machine wears out. However, if an impending bearing failure develops, the passing frequency component corresponding to the defect will increase suddenly and dramatically.

An excellent way to store and compare these spectra is by using a small desktop computer. The spectra can be easily entered into the computer by an instrument interface like GPIB* and compared with previous results by a trend analysis program. This avoids the tedious and error-prone task of generating trend graphs by hand. In addition, the computer can easily check the trends against limits, pointing out where vibration limits are exceeded or where the trend is for the limit to be exceeded in the near future.

Desktop computers are also useful when analyzing machinery that normally operates over a wide range of speeds. Severe vibration modes can be excited when the machine runs at critical speeds. A quick way to determine if these vibrations are a problem is to take a succession of spectra as the machine runs up to speed or coasts down. Each spectrum shows the vibration components of the machine as it passes through an rpm range. If each spectrum is transferred to the computer via GPIB, the results can be processed and displayed as in Figure 4.8. From such a display it is easy to see shaft imbalances, constant frequency vibrations (from sources other than the variable speed shaft) and structural vibrations excited by the rotating shaft. The computer gives the capability of changing the display presentation to other forms for greater clarity. Because all the values of the spectra are stored in memory, precise values of the vibration components can easily be determined. In addition, signal processing can be used to clarify the display. For instance, in Figure 4.8 all signals below -70 dB were ignored. This eliminates meaningless noise from the plot, clarifying the presentation.

So far in this chapter we have been discussing only single channel frequency domain measurements. Let us now look at some measurements we can make with a two channel Dynamic Signal Analyzer.

Figure 4.7 Desktop computer system for monitoring rotating machinery vibration.

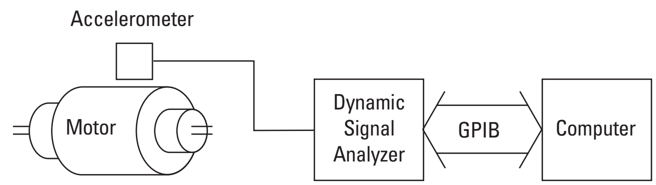
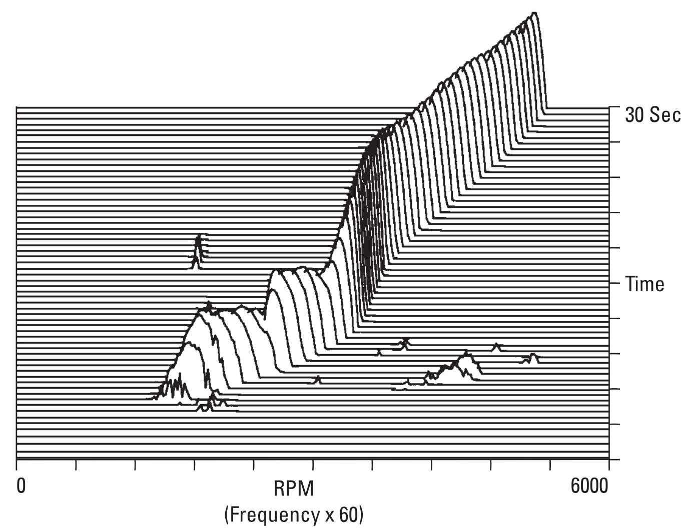


Figure 4.8 Run up test from the system in Figure 4.7.



* General Purpose Interface Bus, Keysight's implementation of IEEE-488-1975.

Electronic Filter Characterization

In Section 6 of the last chapter, we developed most of the principles we need to characterize a low frequency electronic filter. We show the test setup we might use in Figure 4.9. Because the filter is linear we can use pseudo-random noise as the stimulus for very fast test times. The uniform window is used because the pseudo-random noise is periodic in the time record.* No averaging is needed since the signal is periodic and reasonably large. We should be careful, as in the single channel case, to set the input sensitivity for both channels to the most sensitive position which does not overload the analog to digital converters.

With these considerations in mind, we get a frequency response magnitude shown in Figure 4.10a and the phase shown in Figure 4.10b. The primary advantage of this measurement over traditional swept analysis techniques is speed. This measurement can be made in 1/8 second with a Dynamic Signal Analyzer, but would take over 30 seconds with a swept network analyzer. This speed improvement is particularly important when the filter under test is being adjusted or when large volumes are tested on a production line.

Structural Frequency Response

The network under test does not have to be electronic. In Figure 4.11, we are measuring the frequency response of a single structure, in this case a printed circuit board. Because this structure behaves in a linear fashion,

Figure 4.9 Test setup to measure frequency response of filter.

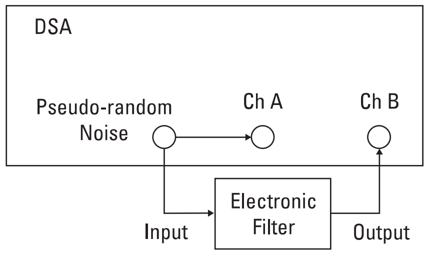
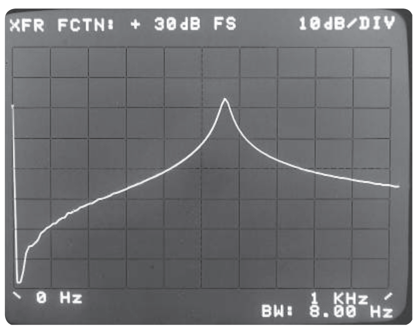


Figure 4.10 Frequency response of electronic filter using PRN and uniform window.

a) Frequency response magnitude



b) Frequency response magnitude and phase

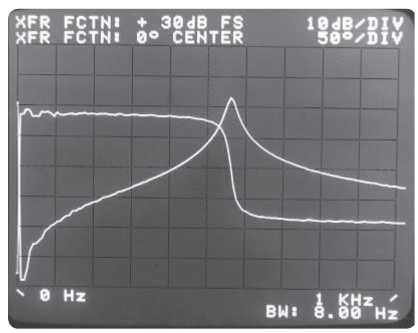
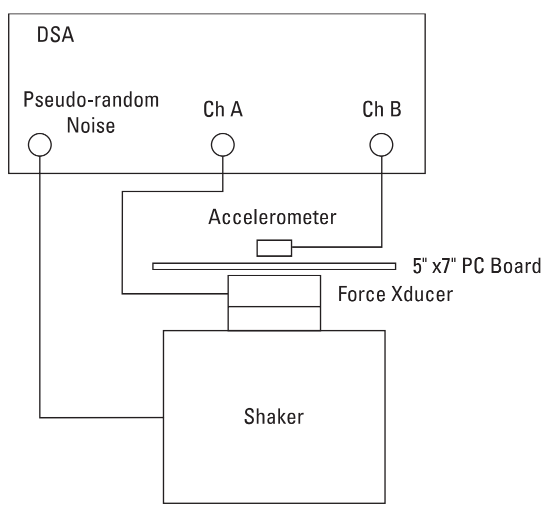


Figure 4.11 Frequency response test of a mechanical structure.



* See the uniform window discussion in Section 6 of the previous chapter for details.

we can use pseudo-random noise as a test stimulus. But we might also desire to use true random noise, swept-sine or an impulse (hammer blow) as the stimulus. In Figure 4.12 we show each of these measurements and the frequency responses. As we can see, the results are all the same.

The frequency response of a linear network is a property solely of the network, independent of the stimulus used.

Since all the stimulus techniques in Figure 4.12 give the same results, we can use whichever one is fastest and easiest. Usually this is the impact stimulus, since a shaker is not required.

In Figure 4.11 and 4.12, we have been measuring the acceleration of the structure divided by the force applied. This quality is called *mechanical accelerance*. To properly scale the displays to the required g's/lb, we have entered the sensitivities of each transducer into the analyzer by a feature called engineering units. Engineering units simply changes the gain of each channel of the analyzer so that the display corresponds to the physical parameter that the transducer is measuring.

Other frequency response measurements besides mechanical accelerance are often made on mechanical structures. Figure 4.14 lists these measurements. By changing transducers we could measure any of these parameters. Or we can use the computational capability of the Dynamic Signal Analyzer to compute these measurements from the mechanical impedance measurement we have already made.

Figure 4.12 Frequency response of a linear network is independent of the stimulus used.

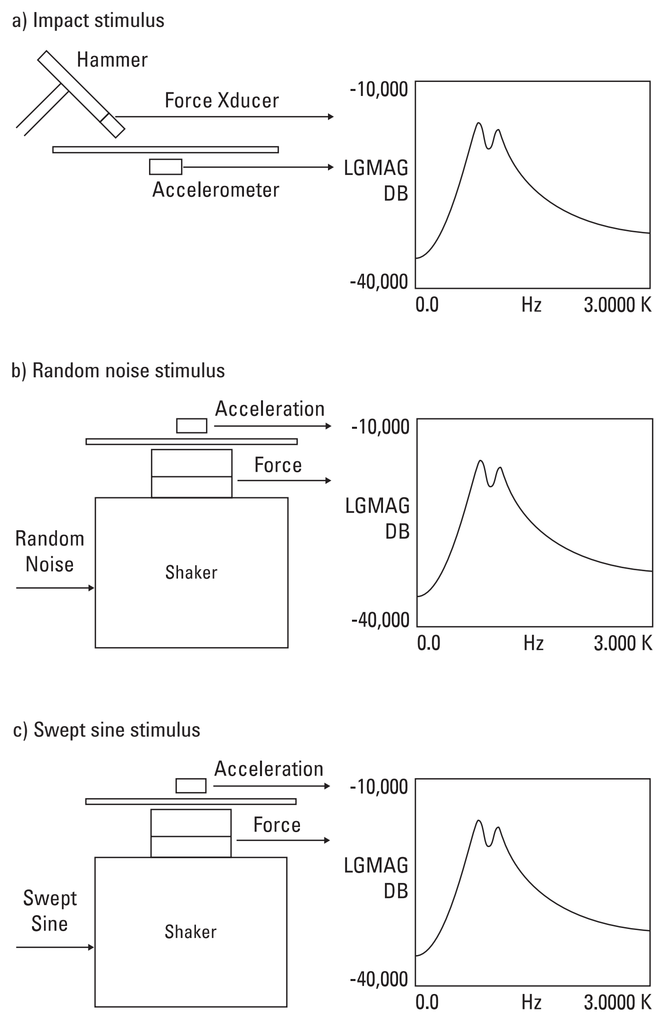


Figure 4.13 Engineering units set input sensitivities to properly scale results.

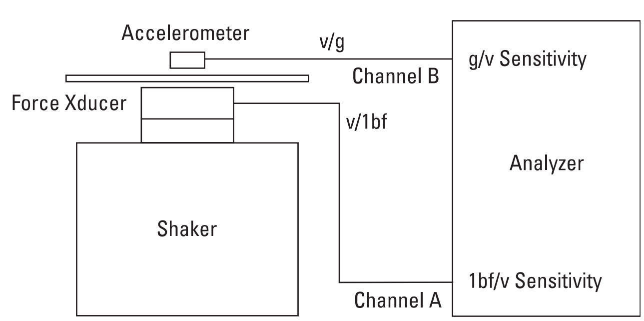
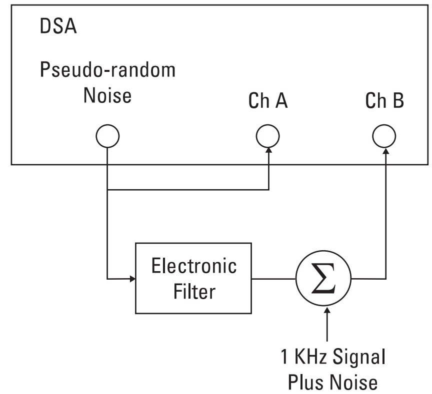


Figure 4.14
Mechanical frequency response measurements.

Accelerance	Acceleration / Force
Effective Mass	Force / Acceleration
Mobility	Velocity / Force
Impedance	Force / Velocity
Dynamic Compliance	Displacement / Force
Dynamic Stiffness	Force / Displacement

Figure 4.15
Simulation of frequency response measurement in the presence of noise.



For instance, we can compute velocity by integrating our acceleration measurement. Displacement is a double integration of acceleration. Many Dynamic Signal Analyzers have the capability of integrating a trace by simply pushing a button. Therefore, we can easily generate all the common mechanical measurements without the need of many expensive transducers.

Coherence

Up to this point, we have been measuring networks which we have been able to isolate from the rest of the world. That is, the only stimulus to the network is what we apply and the only response is that caused by this controlled stimulus. This situation is often encountered in testing components, e.g., electric filters or parts of a mechanical structure. However, there are times when the components we wish to test can not be isolated from other disturbances. For instance, in electronics we might be trying to measure the frequency response of a switching power supply which has a very large component at the switching frequency. Or we might try to measure the frequency response of part of a machine while other machines are creating severe vibration.

In Figure 4.15 we have simulated these situations by adding noise and a 1 KHz signal to the output of an electronic filter. The measured frequency response is shown in Figure 4.16. RMS averaging has reduced the noise contribution, but has not completely eliminated the 1 KHz interference.* If we did not know of the interference, we would think that this filter has an additional resonance at 1 KHz. But Dynamic Signal Analyzers can often make an additional measurement that is not available with traditional network analyzers called *coherence*. Coherence measures the power in the response channel that is *caused* by the power in the reference channel. It is the output power that is coherent with the input power.

Figure 4.17 shows the same frequency response magnitude from Figure 4.16 and its coherence. The coherence goes from 1 (all the output power at

Figure 4.16
Magnitude of frequency response.

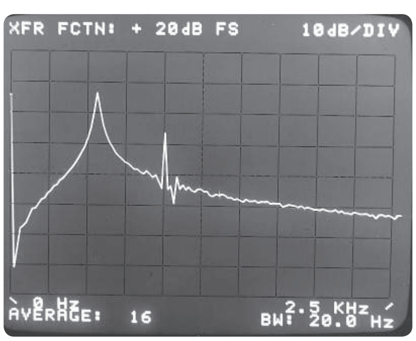
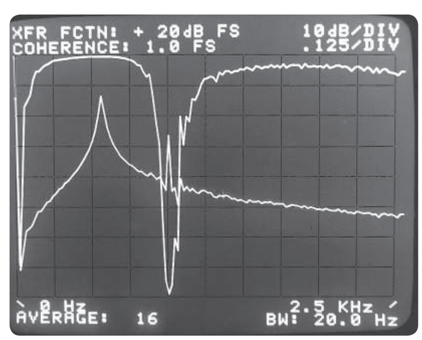


Figure 4.17
Magnitude and coherence of frequency response.



* Additional averaging would further reduce this interference.

that frequency is caused by the input) to 0 (none of the output power at that frequency is caused by the input). We can easily see from the coherence function that the response at 1 KHz is not caused by the input but by interference. However, our filter response near 500 Hz has excellent coherence and so the measurement here is good.

Section 2: Time Domain Measurements

A Dynamic Signal Analyzer usually has the capability of displaying the time record on its screen. This is the same waveform we would see with an oscilloscope, a time domain view of the input. For very low frequency or single-shot phenomena the digital time record storage eliminates the need for storage oscilloscope. But there are other time domain measurements that a Dynamic Signal Analyzer can make as well. These are called correlation measurements. We will begin this section by defining correlation and then we will show how to make these measurements with a Dynamic Signal Analyzer.

Correlation is a measure of the similarity between two *quantities*. To understand the correlation between two *waveforms*, let us start by multiplying these waveforms together at each instant in time and adding up all the products. If, as in Figure 4.18, the waveforms are identical, every product is positive and the resulting sum is large. If however, as in Figure 4.19, the two records are dissimilar, then some of the products would be positive and some would be negative.

Figure 4.18 Correlation of two identical signals.

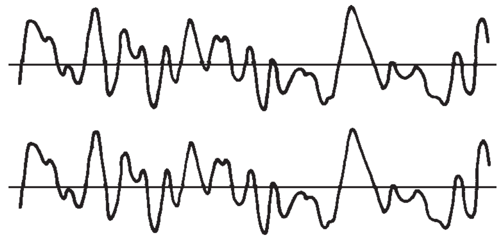


Figure 4.19 Correlation of two different signals.

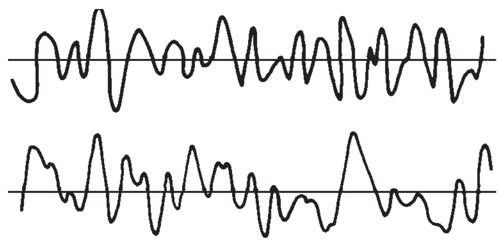
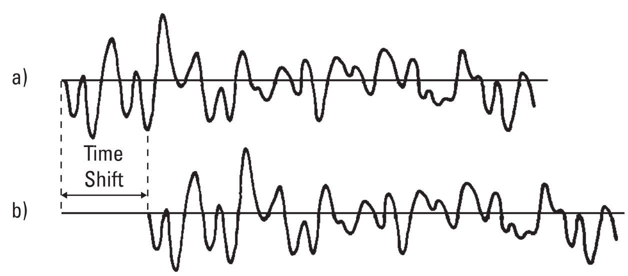


Figure 4.20 Correlation of time displaced signals.



There would be a tendency for the products to cancel, so the final sum would be smaller.

Now consider the waveform in Figure 4.20a, and the same waveform shifted in time, Figure 4.20b. If the time shift were zero, then we would have the same conditions as before, that is, the waveforms would be in phase and the final sum of the products would be large. If the time shift between the two waveforms is made large however, the waveforms appear dissimilar and the final sum is small.

Going one step farther, we can find the average product for each time shift by dividing each final sum by the number of products contributing to it. If we now plot the average product as a function of time shift, the resulting curve will be largest when the time shift is zero and will diminish to zero as the time shift increases. This curve is called the *auto-correlation function* of the waveform. It is a graph of the similarity (or correlation) between a waveform and itself, as a function of the time shift.

The auto-correlation function is easiest to understand if we look at a few examples. The random noise shown in Figure 4.21 is not similar to itself with any amount of time shift (after all, it is random) so its auto-correlation has only a single spike at the point of 0 time shift. Pseudo-random noise, however, repeats itself periodically, so when the time shift equals a multiple of the period, the auto-correlation repeats itself exactly as in Figure 4.22. These are both special cases of a more general statement; the auto-correlation of any periodic waveform is periodic and has the same period as the waveform itself.

Figure 4.21
Auto correlation of random noise.

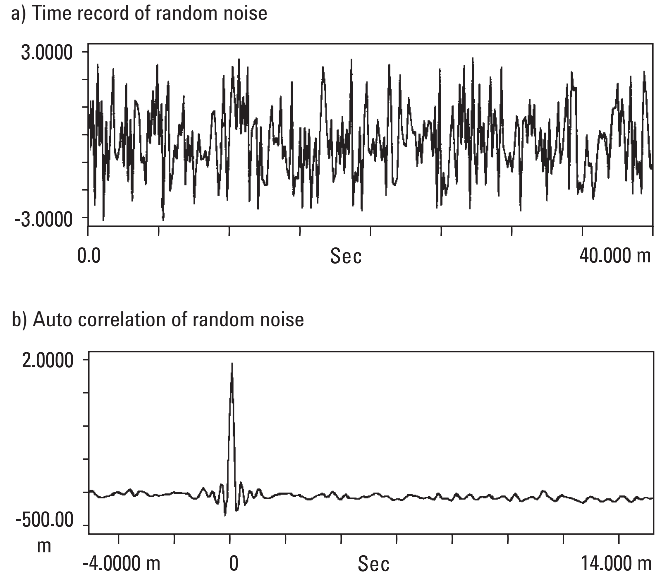
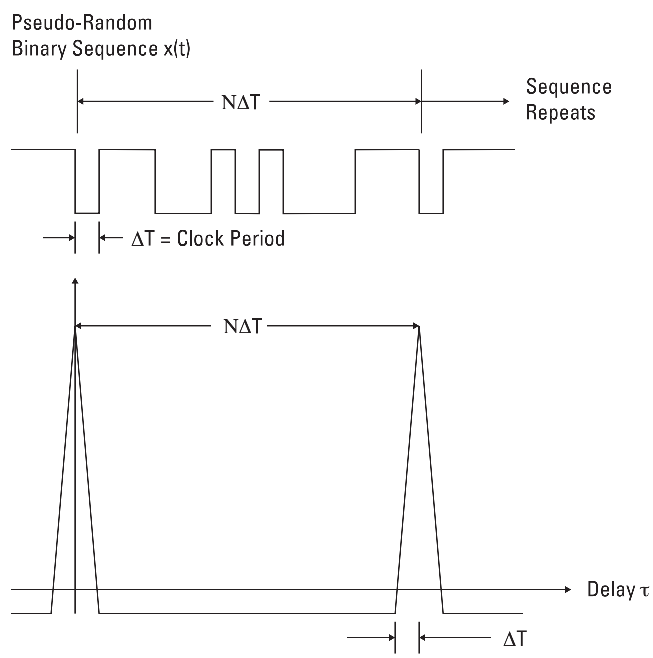


Figure 4.22
Auto correlation of pseudo-random noise.



This can be useful when trying to extract a signal hidden by noise. Figure 4.24a shows what looks like random noise, but there is actually a low level sine wave buried in it. We can see this in Figure 4.24b where we have taken 100 averages of the auto-correlation of this signal. The noise has become the spike around a time shift of zero whereas the auto-correlation of the sine wave is clearly visible, repeating itself with the period of the sine wave.

If a trigger signal that is synchronous with the sine wave is available, we can extract the signal from the noise by linear averaging as in the last section. But the important point about the auto-correlation function is that *no synchronizing trigger is needed*. In signal identification problems like radio astronomy and passive sonar, a synchronizing signal is not available and so auto-correlation is an important tool. The disadvantage of auto-correlation is that the input waveform is not preserved as it is in linear averaging.

Since we can transform any time domain waveform into the frequency domain, the reader may wonder what is the frequency transform of the auto-correlation function? It turns out to be the magnitude squared of the spectrum of the input. Thus, there is really no new information in the auto-correlation function, we had the same

Figure 4.23 Auto-correlation of periodic waveforms.

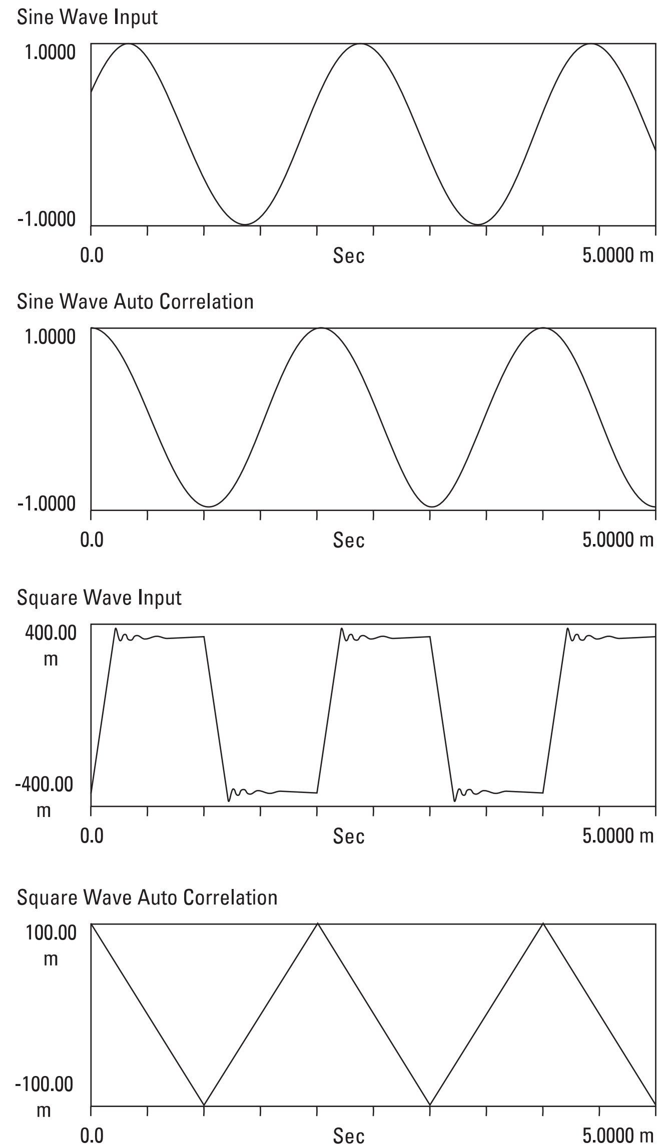
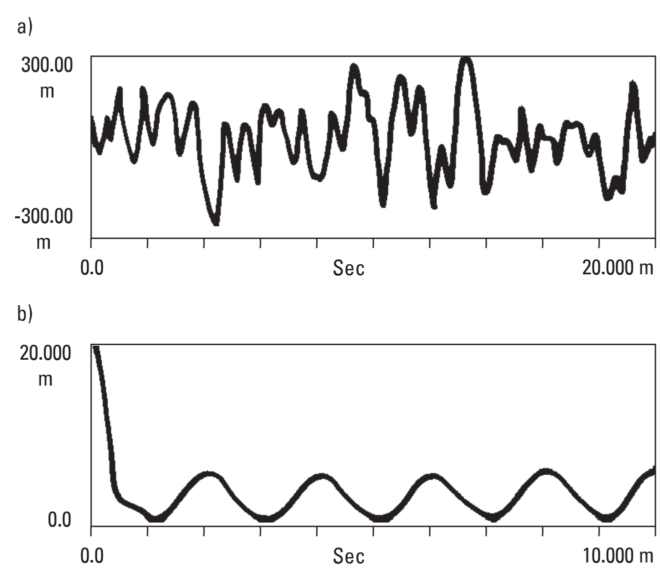


Figure 4.24 Auto-correlation of a sine wave buried by noise.



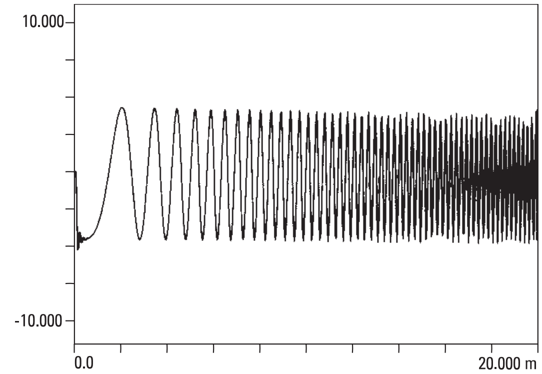
information in the spectrum of the signal. But as always, a change in perspective between these two domains often clarifies problems. In general, impulsive type signals like pulse trains, bearing ping or gear chatter show up better in correlation measurements, while signals with several sine waves of different frequencies like structural vibrations and rotating machinery are clearer in the frequency domain.

Cross Correlation

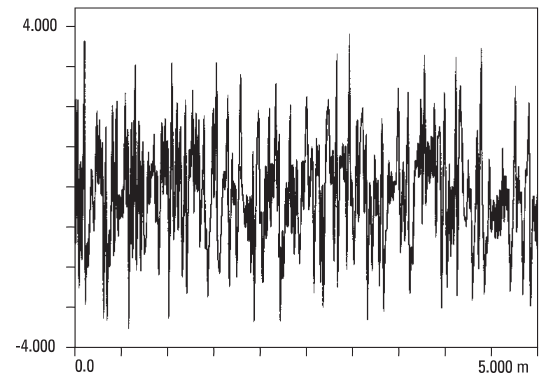
If auto-correlation is concerned with the similarity between a signal and a time shifted version of itself, then it is reasonable to suppose that the same technique could be used to measure the similarity between two *non-identical* waveforms. This is called the *cross correlation function*. If the same signal is present in both waveforms, it will be reinforced in the cross correlation function, while any uncorrelated noise will be reduced. In many network analysis problems, the stimulus can be cross correlated with the response to reduce the effects of noise. Radar, active sonar, room acoustics and transmission path delays all are network analysis problems where the stimulus can be measured and used to remove contaminating noise from the response by cross correlation.*

Figure 4.25
Simulated radar cross correlation.

a) 'Transmitted' signal, a swept-frequency sine wave



b) 'Received' signal, the swept sine wave plus noise



c) Result of cross correlation of the transmitted and received signals. Distance from left edge to peak represents transmission delay.

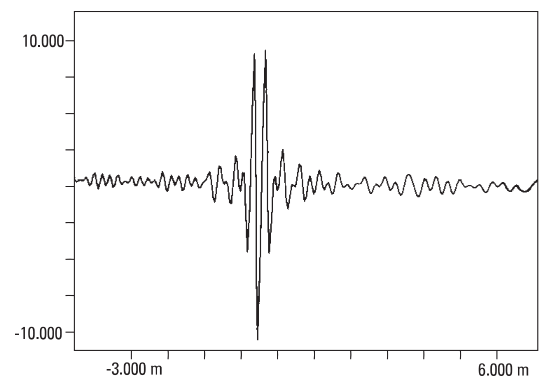
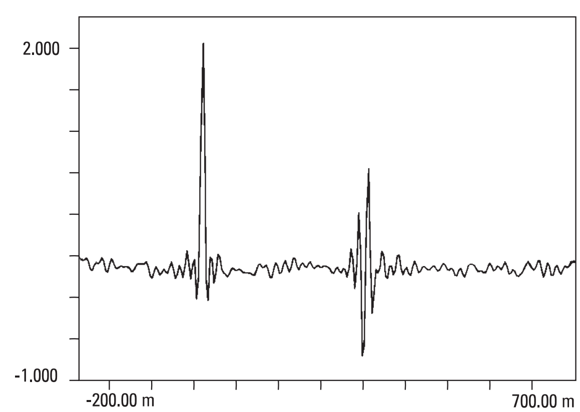


Figure 4.26
Cross correlation shows multiple transmission paths.



* The frequency transform of the cross correlation function is the cross power spectrum, a function discussed in Appendix A.

Section 3: Modal Domain Measurements

In Section 1 we learned how to make frequency domain measurements of mechanical structures with Dynamic Signal Analyzers. Let us now analyze the behavior of a simple mechanical structure to understand how to make measurements in the modal domain. We will test a simple metal plate shown in Figure 4.27. The plate is freely suspended using rubber cords in order to isolate it from any object which would alter its properties.

The first decision we must make in analyzing this structure is how many measurements to make and where to make them on the structure. There are no firm rules for this decision; good engineering judgment must be exercised instead. Measuring too many points make the calculations unnecessarily complex and time consuming. Measuring too few points can cause spatial aliasing; i.e., the measurement points are so far apart that high frequency bending modes in the structure can not be measured accurately. To decide on a reasonable number of measurement points, take a few trial frequency response measurements of the structure to determine the highest significant resonant frequencies present. The wave length can be determined empirically by changing the distance between the stimulus and the sensor until a full 360° phase shift has occurred from the original measurement point. Measurement point spacing should be approximately one-quarter or less of this wavelength.

Measurement points can be spaced uniformly over the structure using this guideline, but it may be desirable to modify this procedure slightly. Few structures are as uniform as this simple plate example,* but complicated structures are made of simpler, more uniform parts. The behavior of the structure at the junction of these parts is often of great interest, so measurements should be made in these critical areas as well.

Once we have decided on where the measurements should be taken, we number these measurement points (the order can be arbitrary) and enter the coordinates of each point into our modal analyzer. This is necessary so that the analyzer can correlate the measurements we make with a position on the structure to compute the mode shapes.

The next decision we must make is what signal we should use for a stimulus. Our plate example is a linear structure as it has no loose rivet joints, non-linear damping materials, or other non-linearities. Therefore, we know that we can use any of the stimuli described in Chapter 3, Section 6. In this case, an impulse would be a particularly good test signal. We could supply the impulse by hitting the structure with a ham-

Figure 4.27 Modal analysis example - Determine the modes in this simple plate.

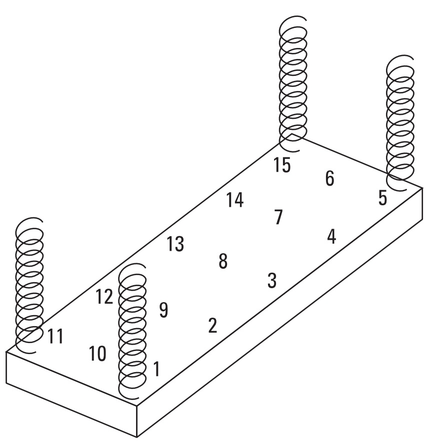
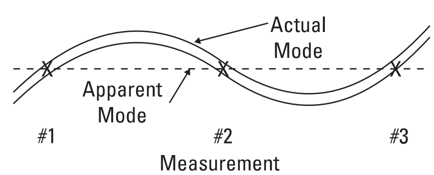


Figure 4.28 Spacial Aliasing - Too few measurement points lead to inaccurate analysis of high frequency bending mode.



* If all structures were this simple, there would be no need for modal analysis.

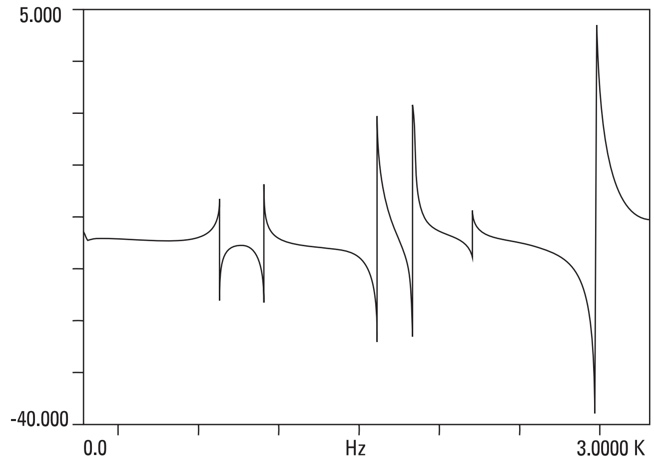
mer equipped with a force transducer. This is probably the easiest way to excite the structure as a shaker and its associated driver are not required. As we saw in the last chapter, however, if the structure were non-linear, then random noise would be a good test signal. To supply random noise to the structure we would need to use a shaker. To keep our example more general, we will use random noise as a stimulus.

The shaker is connected firmly to the plate via a load cell (force transducer) and excited by the band-limited noise source of the analyzer. Since this force is the network stimulus, the load cell output is connected through a suitable amplifier to the reference channel of the analyzer. To begin the experiment, we connect an accelerometer* to the plate at the same point as the load cell. The accelerometer measures the structure's response and its output is connected to the other analyzer channel.

Because we are using random noise, we will use a Hanning window and RMS averaging just as we did in the previous section.

The resulting frequency response of this measurement is shown in Figure 4.29. The ratio of acceleration to force in g's/lb is plotted on the vertical axis by the use of engineering units, and the data shows a number of distinct peaks and valleys at particular frequencies. We conclude that the plate moves more freely when subjected to energy at certain specific frequencies than it does in response to energy at other frequencies. We recall that each of the resonant peaks correspond to a mode of vibration of the structure.

Figure 4.29
A frequency response of the plate.



Our simple plate supports a number of different modes of vibration, all of which are well separated in frequency. Structures with widely separated modes of vibration are relatively straightforward to analyze since each mode can be treated as if it is the only one present. Tightly spaced, but lightly damped vibration modes can also be easily analyzed if the Band Selectable Analysis capability is used to narrow the analyzer's filter sufficiently to resolve these resonances. Tightly spaced modes whose damping is high enough to cause the responses to overlap create computational difficulties in trying to separate the effects of the vibration modes. Fortunately, many structures fall into the first two categories and so can be easily analyzed.

Having inspected the measurement and deciding that it met all the above criteria, we can store it away. We store similar measurements at each point by moving our accelerometer to each numbered point. We will then have all the measurement data we need to fully characterize the structure in the modal domain.

Recall from Chapter 2 that each frequency response will have the same number of peaks, with the same resonant frequencies and dampings. The next task is to determine these resonant frequency and damping values for each resonance of interest. We do this by retrieving our stored frequency responses and, using a curve-fitting routine, we calculate the frequency and damping of each resonance of interest.

With the structural information we entered earlier, and the frequency and damping of each vibration mode which we have just determined, the

* Displacement, velocity or strain transducers could also be used, but accelerometers are often used because they are small and light, and therefore do not affect the response of the structure. In addition, they are easy to mount on the structure, reducing the total measurement time.

analyzer can calculate the mode shapes by curve fitting the responses of each point with the measured resonances. In Figure 4.30 we show several mode shapes of our simple rectangular plate. These mode shapes can be animated on the display to show the relative motion of the various parts of the structure. The graphs in Figure 4.30, however, only show the maximum deflection.

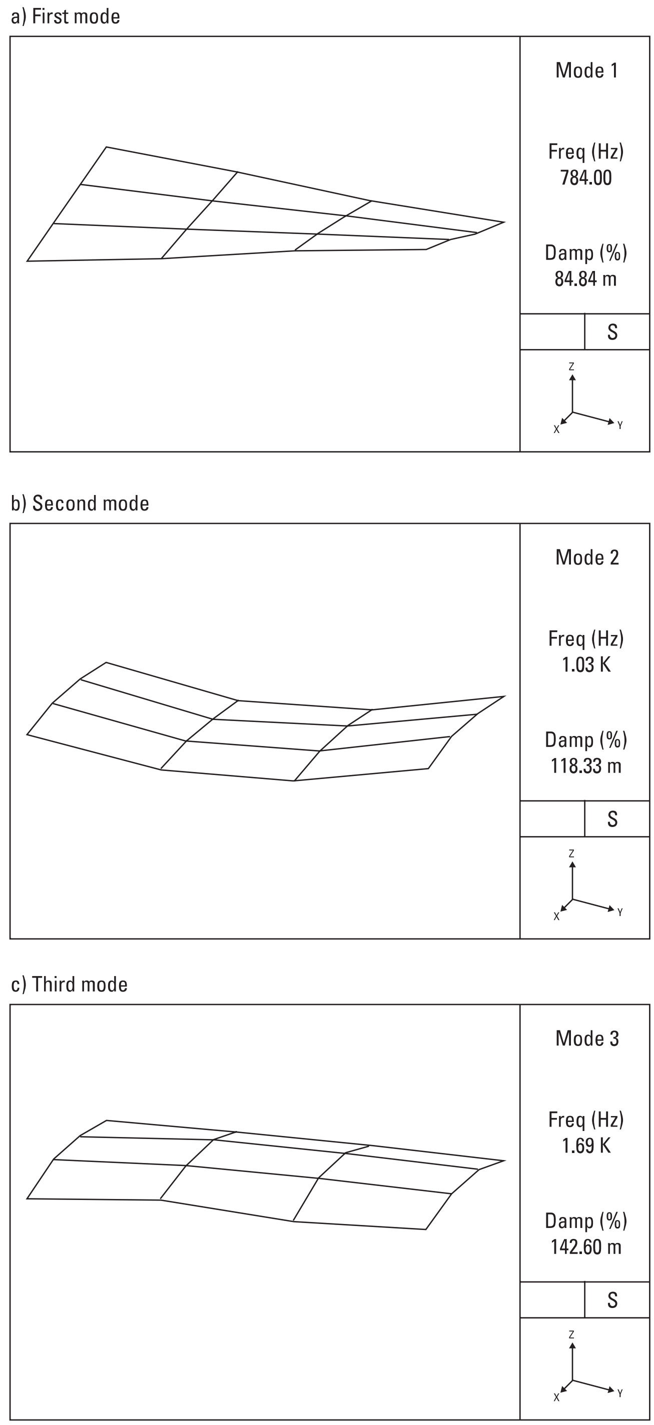


Figure 4.30 Mode shapes of a rectangular plate.

Section 4: Summary

This note has attempted to demonstrate the advantages of expanding one's analysis capabilities from the time domain to the frequency and modal domains. Problems that are difficult in one domain are often clarified by a change in perspective to another domain. The Dynamic Signal Analyzer is a particularly good analysis tool at low frequencies. Not only can it work in all three domains, it is also very fast.

We have developed heuristic arguments as to why Dynamic Signal Analyzers have certain properties because understanding the principles of these analyzers is important in making good measurements. Finally, we have shown how Dynamic Signal Analyzers can be used in a wide range of measurement situations using relatively simple examples. We have used simple examples throughout this text to develop understanding of the analyzer and its measurements, but it is by no means limited to such cases. It is a powerful instrument that, in the hands of an operator who understands the principles developed in this note, can lead to new insights and analysis of problems.

Appendix A

The Fourier Transform: A Mathematical Background

The Fourier Transform

The transformation from the time domain to the frequency domain and back again is based on the Fourrier Transform and its inverse. This Fourier Transform pair is defined as:

$$S_x(f) = \int_{-\infty}^{\infty} x(t)\, e^{-j2\pi ft} dt \qquad \text{(Forward Transform)} \qquad \text{A.1}$$

$$x(t) = \int_{-\infty}^{\infty} S_x(f)\, e^{j2\pi ft} df \qquad \text{(Inverse Transform)} \qquad \text{A.2}$$

where

$x(t)$ = time domain representation of the signal x

$S_x(f)$ = frequency domain representation of the signal x

$j = \sqrt{-1}$

The Fourier Transform is valid for both periodic* and non-periodic x(t) that satisfy certain minimum conditions. All signals encountered in the real world easily satisfy these requirements.

The Discrete Fourier Transform

To compute the Fourier Transform digitally, we must perform a numerical integration. This will give us an approximation to a true Fourier Transform called the Discrete Fourier Transform.

There are three distinct difficulties with computing the Fourier Transform. First, the desired result is a continuous function. We will only be able to calculate its value at discrete points. With this constraint our transform becomes,

$$S_x(m\Delta f) = \int_{-\infty}^{\infty} x(t)\, e^{-j2\pi m\Delta ft} dt \qquad \text{A.3}$$

where m = 0, ±1, ±2

and Δf = frequency spacing of our lines

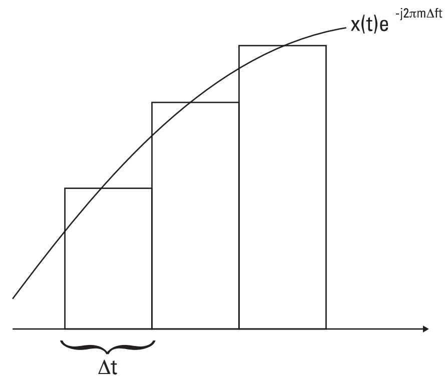


Figure A.1 Numerical integration used in the Fourier Transform

The second problem is that we must evaluate an integral. This is equivalent to computing the area under a curve. We will do this by adding together the areas of narrow rectangles under the curve as in Figure A.1.

Our transform now becomes:

$$S_x(m\Delta f) \approx \Delta t \sum_{-\infty}^{\infty} x(n\Delta t)\, e^{-j2\pi m\Delta fn\Delta t} \qquad \text{A.4}$$

where Δt = time interval between samples

The last problem is that even with this summation approximation to the integral, we must sum samples over all time from minus to plus infinity. We would have to wait forever to get a result. Clearly then, we must limit the transform to a finite time interval.

$$S_x(m\Delta f) \approx \Delta t \sum_{n-0}^{n-1} x(n\Delta t)\, e^{-j2\pi m\Delta fn\Delta t} \qquad \text{A.5}$$

As developed in Chapter 3, the frequency spacing between the lines must be the reciprocal of the time record length. Therefore, we can simplify A.5 to our formula for the *Discrete Fourier Transform*, $S'_{x \cdot n-1}$

$$S'_x(m\Delta f) \approx \frac{T}{N} \sum_{n-0} x(n\Delta t)\, e^{-j2\pi mn/N} \qquad \text{A.6}$$

* The Fourier Series is a special case of the Fourier Transform.

The Fast Fourier Transform

The Fast Fourier Transform (FFT) is an algorithm for computing this Discrete Fourier Transform (DFT). Before the development of the FFT the DFT required excessive amounts of computation time, particularly when high resolution was required (large N). The FFT forces one further assumption, that N is a multiple of 2. This allows certain symmetries to occur reducing the number of calculations (specifically multiplications) which have to be done.

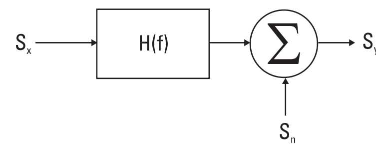
It is important to recall here that the Fast Fourier Transform is only an approximation to the desired Fourier Transform. First, the FFT only gives samples of the Fourier Transform. Second and more important, it is only a transform of a finite time record of the input.

Two Channel Frequency Domain Measurements

As was pointed out in the main text, two channel measurements are often needed with a Dynamic Signal Analyzer. In this section we will mathematically define the two channel transfer function and coherence measurements introduced in Chapter 4 and prove their more important properties.

However, before we do this, we wish to introduce one other function, the *Cross Power Spectrum*, G_{xy} . This function is not often used in measurement situations, but is used internally by Dynamic Signal Analyzers to compute transfer functions and coherence.

Figure A.2 Transfer function measurments with noise present.



The *Cross Power Spectrum*, G_{xy}, is defined as taking the Fourier Transform of two signals separately and multiplying the result together as follows:

$$G_{xy}(f) = S_x(f)\, S^*_y(f)$$

where * indicates the complex conjugate of the function.

With this function, we can define *the Transfer Function*, H(f), using the cross power spectrum and the spectrum of the input channel as follows:

$$H(f) = \frac{\overline{G_{yy}(f)}}{\overline{G_{xx}(f)}}$$

where $\overline{\quad}$ denotes the average of the function.

At first glance it may seem more appropriate to compute the transfer function as follows:

$$|H(f)|^2 = \frac{\overline{G_{yy}}}{\overline{G_{xx}}}$$

This is the ratio of two single channel, averaged measurements. Not only does this measurement not give any phase information, it also will be in error when there is noise in the measurement. To see why let us solve the equations for the special case where noise is injected into the output as in Figure A.2. The output is:

$$S_y(f) = S_x(f)H(f) + S_n(f)$$

So

$$G_{yy}=S_yS_y^*= G_{xx}|H|^2+S_xHS_n+S_x^*H^*S_n+|S_n|^2$$

If we RMS average this result to try to eliminate the noise, we find the S_xS_n terms approach zero because S_x, and S_n, are uncorrelated. However, the $\overline{|S_n|^2}$ term remains as an error and so we get

$$\frac{\overline{G_{yy}}}{\overline{G_{xx}}} = |H|^2 + \frac{\overline{|S_n|^2}}{\overline{G_{xx}}}$$

Therefore if we try to measure $|H|^2$ by this single channel techniques, our value will be high by the noise to signal ratio.

If instead we average the cross power spectrum we will eliminate this noise error. Using the same example,

$$\overline{G_{yx}}=\overline{S_yS_x}^*=\overline{(S_xH+S_n)S_x}^*= \overline{G_{xx}}H +\overline{S_nS_x}^*$$

so

$$\frac{\overline{G_{yx}}}{\overline{G_{xx}}} =H(f)+S_nS_x^*$$

Because S_n, and S_x, are uncorrelated, the second term will average to zero, making this function a much better estimate of the transfer function.

The *Coherence Function*, γ^2, is also derived from the cross power spectrum by:

$$\gamma^2(f) = \frac{\overline{G_{yx}(f)}\ \overline{G_{xy}^*(f)}}{\overline{G_{xx}(f)}\ \overline{G_{yy}(f)}}$$

As stated in the main text, the coherence function is a measure of the power in the output signal caused by the input. If the coherence is 1, then all the output power is caused by the input. If the coherence is 0, then none of the output is caused by the input. Let us now look at the mathematics of the coherence function to see why this is so.

As before, we will assume a measurement condition like Figure A.2. Then, as we have shown before,

$$\overline{G_{xy}}= \overline{G_{xx}}|H|^2+\overline{S_xHS_n^*}+ \overline{S_x^*H^*S_n}+\overline{|S|^2}$$

$$\overline{G_{xy}}=\overline{G_{xx}}H+\overline{S_nS_x^*}$$

As we average, the cross terms S_nS_x, approach zero, assuming that the signal and the noise are not related. So the coherence becomes

$$\gamma^2(f) = \frac{(\overline{HG_{yx}})^2}{\overline{G_{xx}}(|H|^2\overline{G_{xx}}+\overline{|S_n|^2})}$$

$$\gamma^2(f) = \frac{|H|^2\overline{G_{xx}}}{|H|^2\overline{G_{xx}}+\overline{S_n}^2}$$

We see that if there is no noise, the coherence function is unity. If there is noise, then the coherence will be reduced. Note also that the coherence is a function of frequency. The coherence can be unity at frequencies where there is no interference and low where the noise is high.

Time Domain Measurements

Because it is sometimes easier to understand measurement problems from the perspective of the time domain, Dynamic Signal Analyzers often include several time domain measurements. These include auto and cross correlation and impulse response.

Auto Correlation, $R_{xx}(\tau)$, is a comparison of a signal with itself as a function of time shift. It is defined as;

$$R_{xx}(\tau)= \lim_{T\to\infty} \frac{1}{T}\int_T x(t)x(t+\tau)dt$$

That is, the auto correlation can be found by taking a signal and multiplying it by the same signal displaced by a time τ and averaging the product over all time. However, most Dynamic Signal Analyzers compute this quantity by taking advantage of its dual in the frequency domain. It can be shown that

$$R_{xx}(\tau)=F^{-1}[S_x(f)S_x^*(f)]$$

where F^{-1} is the inverse Fourier Transform and S_x is the Fourier Transform of x(t)

Since both techniques yield the same answer, the latter is usually chosen for Dynamic Signal Anlyzer since the Frequency Transform algorithm is already in the instrument and the results can be computed faster because fewer multiplications are required.

Cross Correlation, $Rxy(\tau)$, is a comparison of two signals as a function of a time shift between them. It is defined as:

$$R_{xy}(\tau)=\lim_{T\to\infty}\frac{1}{T}\int_T x(t)y(t+\tau)dt$$

As in auto correlation, a Dynamic Signal Analyzer computes this quantity indirectly, in this case from the cross power spectrum.

$$R_{xy}(\tau)=F^{-1}[G_{xy}]$$

Lastly, the *Impulse Response*, h(t), is the dual of the transfer function,

$$h(t) = F^{-1}[H(f)]$$

Note that because the transfer function normalized the stimulus, the impulse response can be computed no matter what stimulus is actually used on the network.

Appendix B Bibliography

Bendat, Julius S. and Piersol, Allan G., *"Random Data: Analysis and Measurement Procedures"*, Wiley-Interscience, New York, 1971.

Bendat, Julius S. and Piersol, Allan G., *"Engineering Applications of Correlation and Spectral Analysis"*, Wiley-lnterscience, New York, 1980.

Bracewell, R., *"The Fourier Transform and its Applications"*, McGraw-Hill, 1965.

Cooley, J.W. and Tukey, J.W., *"An Algorithm for the Machine Calculation of Complex Fourier Series"*, Mathematics of Computation, Vo. 19, No. 90, p. 297, April 1965.

McKinney, W., *"Band Selectable Fourier Analysis"*, Hewlett-Packard Journal, April 1975, pp. 20-24.

Otnes, R.K. and Enochson, L., *"Digital Time Series Analysis"*, John Wiley, 1972.

Potter, R. and Lortscher, J., *"What in the World is Overlap Processing"*, Hewlett-Packard Santa Clara DSA/Laser Division *"Update"*, Sept. 1978.

Ramse , K.A., *"Effective Measurements for Structural Dynamics Testing"*, Part 1, Sound and Vibration Magazine, November 1975, pp. 24-35.

Roth, P., *"Effective Measurements Using Digital Signal Analysis"*, IEEE Spectrum, April 1971, pp. 62-70.

Roth, P., *"Digital Fourier Analysis"*, Hewlett-Packard Journal, June 1970.

Welch, Peter D., *"The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms"*, IEEE Transactions on Audio and Electro-acoustics, Vol. AU-15, No. 2, June 1967, pp. 70-73.

myKeysight

myKeysight

www.keysight.com/find/mykeysight
A personalized view into the information most relevant to you.

DEKRA Certified ISO 9001:2008

www.keysight.com/go/quality
Keysight Technologies, Inc.
DEKRA Certified ISO 9001:2008
Quality Management System

Keysight Channel Partners
www.keysight.com/find/channelpartners
Get the best of both worlds: Keysight's measurement expertise and product breadth, combined with channel partner convenience.

This document was formerly known as application note 243.

www.keysight.com/find/assist

For more information on Keysight Technologies' products, applications or services, please contact your local Keysight office. The complete list is available at:
www.keysight.com/find/contactus

Americas

Canada	(877) 894 4414
Brazil	55 11 3351 7010
Mexico	001 800 254 2440
United States	(800) 829 4444

Asia Pacific

Australia	1 800 629 485
China	800 810 0189
Hong Kong	800 938 693
India	1 800 112 929
Japan	0120 (421) 345
Korea	080 769 0800
Malaysia	1 800 888 848
Singapore	1 800 375 8100
Taiwan	0800 047 866
Other AP Countries	(65) 6375 8100

Europe & Middle East

Austria	0800 001122
Belgium	0800 58580
Finland	0800 523252
France	0805 980333
Germany	0800 6270999
Ireland	1800 832700
Israel	1 809 343051
Italy	800 599100
Luxembourg	+32 800 58580
Netherlands	0800 0233200
Russia	8800 5009286
Spain	800 000154
Sweden	0200 882255
Switzerland	0800 805353
	Opt. 1 (DE)
	Opt. 2 (FR)
	Opt. 3 (IT)
United Kingdom	0800 0260637

For other unlisted countries:
www.keysight.com/find/contactus
(BP-09-23-14)

KEYSIGHT TECHNOLOGIES

Published in USA, July 31, 2014
5952-8898E
www.keysight.com