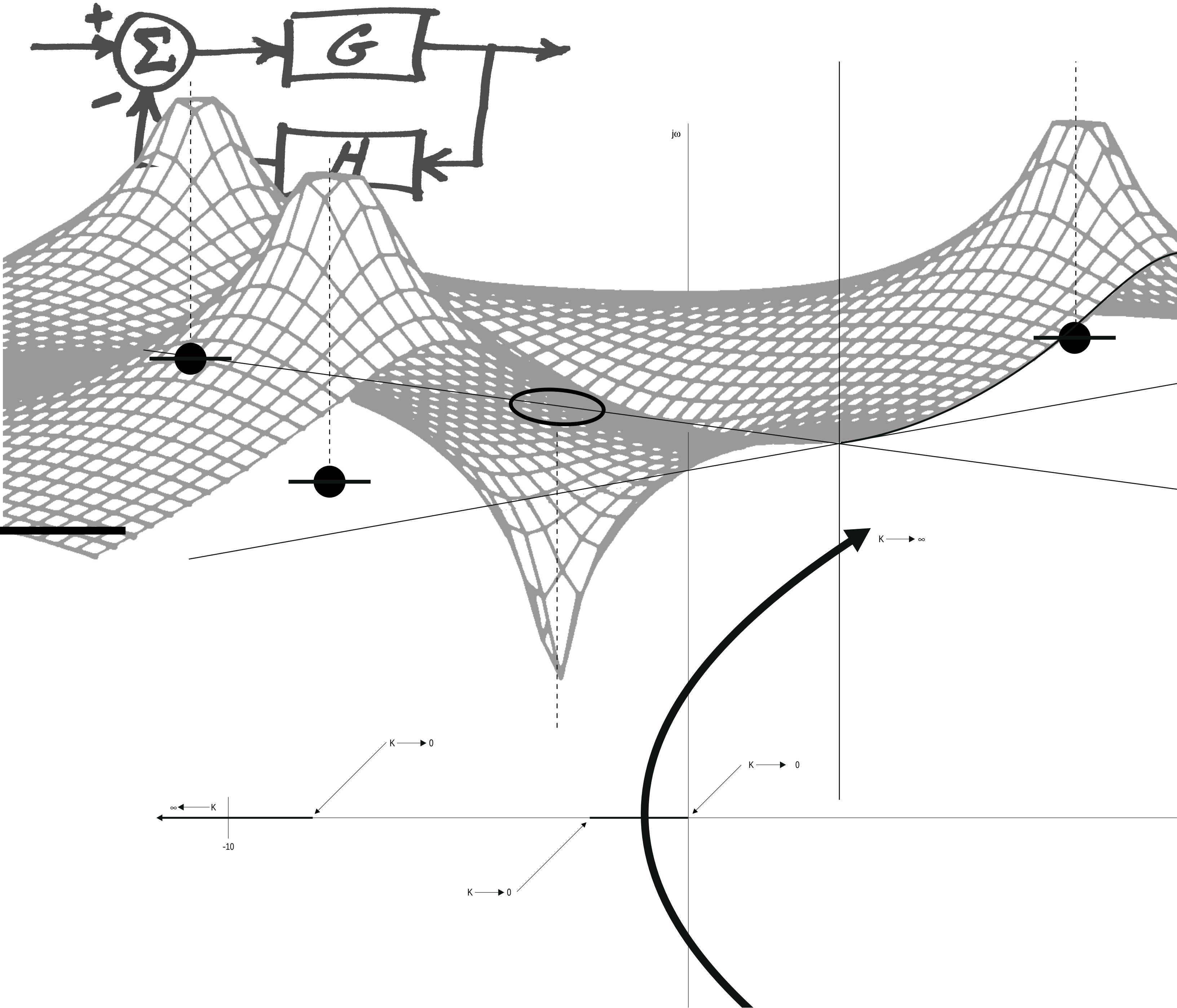
HEWLETT® PACKARD

Control System Development using Dynamic Signal Analyzers

Application Note 243-2

$$\frac{G(s)}{1 + G(s)H(s)}$$



Introduction

Dynamic Signal Analyzers (DSAs) represent a new generation of microprocessor-based test instruments designed to support the development of control systems. By combining the computational resources of microprocessors with the accuracy of precision measurement hardware, DSAs combine high-performance measurements and powerful computer-aided-engineering. By consolidating this much power into a single instrument, DSAs have expanded the role of test instruments beyond traditional testing functions to include contributions in the areas of modeling, design and analysis.

The purpose of this application note is to examine how the advanced measurement and analysis capabilities of a DSA can be applied to the development and production of control systems to reduce testing time, reduce analysis time, provide more information from measurements and, in general, enhance the overall development and production process.

Using This Application Note

This application note is designed for both the experienced control systems engineer who may be unfamiliar with DSAs and the experienced DSA user entering the field of control systems. To accommodate this broad range of readers, the note is divided into two parts.

Part 1 is a review of the basic concepts associated with control systems and linear control theory. This section serves as a general resource and may be considered optional reading for the experienced control system engineer.

Part 2 is an introduction to the features and functions of DSA's which directly contribute to the development of control systems. Each feature or function is briefly described with example applications provided.

A glossary of control system terms is provided in Appendix A.

Contents

Part 1: An Introduction to Control Systems and Classical Control Theory **4**

Chapter 1: Basic Terms and Definitions **4**

Chapter 2: Modeling **5**
2-1: The Open-Loop Model 5
2-2: The Closed-Loop Model 6

Chapter 3: Measuring Performance **8**
3-1: Time Domain Performance — Step Response 8
3-2: Frequency Domain Performance 9
3-2.1: Frequency Domain Terms and Definitions 9
3-2.2: Nyquist's Stability Criterion (s-plane) 12
3-2.3: Nyquist Diagrams 13
3-2.4: Nyquist's Stability Criterion (Nyquist diagram) 14
3-2.5: Magnitude and Phase Contours 15
3-2.6: Bode on Stability 16
3-2.7: Gain Margin and Phase Margin 18

Chapter 4: More Tools for Design and Analysis **20**
4-1: The Bode Diagram 20
4-2: Stability and the Bode Diagram 22
4-3: The Nichols Diagram 22
4-4: The Root Locus Diagram 24

Chapter 5: Nonlinear Systems **26**

Part 2: Measurement and Analysis Tools Applied to the Development Process **28**

Chapter 1: Modeling the Development Process **28**

Chapter 2: Test **29**
2-1: Time Domain Measurements 29
2-1.1: Time Capture 30
2-1.2: Time Throughput 30
2-2: Frequency Domain Measurements 31
2-2.1: Swept Fourier Analysis 31
2-2.2: Fast Fourier Transform Analysis 32

Chapter 3: Analyze **34**
3-1: Waveform Math 34
3-2: Curve Fitting 39
3-3: Coherence 43

Chapter 4: Model **44**
4-1: Curve Fitting Applied to the Modelling Process 44
4-2: Frequency Response Synthesis Applied to the Modeling Process 48

Chapter 5: Design **49**
5-1: Applying Frequency Response Synthesis, Waveform Math and Curve Fitting to the Design Process 49
5-2: Using Display Formats Other Than the Bode Plot 55

Chapter 6: Summary **56**

Appendicies **57**

Appendix A: Glossary 57
Appendix B: Bibliography 60
Appendix C: Acknowledgments 60

Part 1:
An Introduction to Control Systems and Classical Control Theory.

Chapter 1: Basic Terms and Definitions

A control system has been formally described as, "A system in which deliberate guidance or manipulation is used to achieve a prescribed value of a variable."[1] With a variable further defined as, "a quantity or condition which is subject to change," it becomes apparent that the components of a control system may be virtually any definable entity, be it electrical, mechanical, biological, organizational or otherwise.

The human circulatory system, pacemakers, motor speed controls, clothes dryers, automobile cruise controls and voltage regulators are a few examples of the vast number of control systems in existence. The diversity of control systems may at first seem a barrier against the development of a common analysis and design strategy. Fortunately, if the components of a system can be represented through a common mathematical symbolism, then there exists a collection of concepts and methods for studying the physical properties of control systems known as control theory.

While a thorough study of control theory is far beyond the scope of this application note, the following paragraphs present the basic concepts associated with classic control theory as applied to continuous linear control systems.[2]

To categorize control systems with common traits or functions, several subclasses of control systems have been defined. One of the basic categories of control systems are those systems which operate without human intervention. Control systems in this category are called automatic control systems. An example of an automatic control system is an automobile cruise control which maintains the speed of the vehicle without attention from the driver. If the driver disengages the cruise control, he then becomes part of the control system regulating the speed of the car and, therefore, part of a nonautomatic control system.

Another category involves those automatic control systems which involve mechanical motion as the controlled variable. These control systems are called servomechanisms (commonly referred to as servos) and are defined as, "An automatic feedback control system in which the controlled variable is mechanical position or any of its time derivatives." While this definition seems straightforward, general usage has diluted the literal meaning to include virtually any electronic, electromechanical or mechanical control system.

Control systems are also categorized as being either open-loop or closed-loop. The difference between these two categories, the use of feedback, becomes easier to understand when viewing the basic model of a control system. Formal definitions of open-loop and closed-loop control systems have therefore been incorporated into the following chapter on control system modeling.

[1] American National Standards Institute specification MC85.1M-1981, *Terminology for Automatic Control.*

[2] References for further study of modern or classic control theory as applied to linear, nonlinear, continuous and discrete control systems are listed at the end of this note.

Chapter 2: Modeling

The first step in the design or analysis of a control system is to develop an analytical model of the system. This is done by dividing the control system into functional blocks. Each block may represent any portion of the control system from an individual component to a group of components which perform an identifiable function.

2-1: The Open-Loop Model

Figure 1-1A is a block diagram which represents a very basic control system. The letters *r* and *c* represent the directly controlled variable and the reference input respectively. The letter *g* represents an equation which describes the influence of the elements within the functional block on a signal or action compared at the input and output of the functional block. All lower case letters generally denote functions in the time domain unless otherwise specified (for example, $c = c(t)$). The upper case variables *R* and *C* in Figure 1-1A represent the Laplace transform of *r* and *c* expressed as functions of the complex variable *s*[1]. The upper case *G* represents the Laplace transform of *g* and is generally referred to as a transfer function[2].

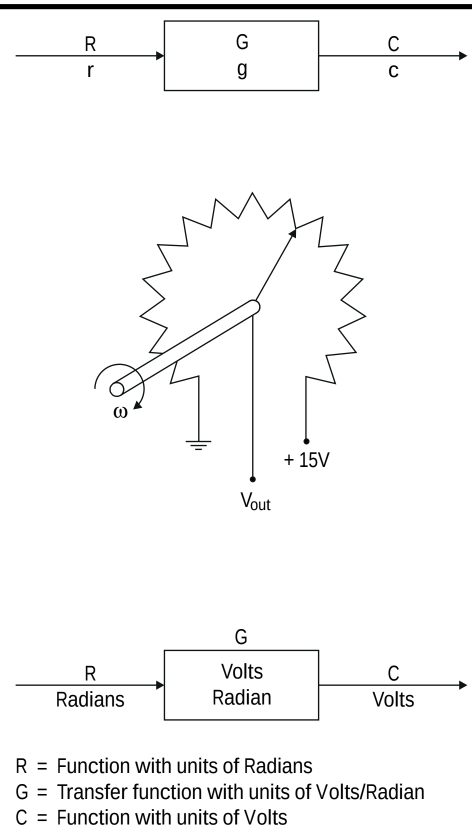
A simple example of the type of control system shown in Figure 1-1A is a potentiometer connected as a voltage divider, as shown in Figure 1-1B. For this example the reference input *R* would have units of radians, the directly controlled variable *C* units of volts, and the transfer function *G* would be a constant with units of volts per radian (as shown in Figure 1-1C). A drawback of this type of a control system is its inability to respond to dynamic changes in the system. For example, if a load resistance was connected to the output, there would be an undesirable change in the output voltage. This type of control system, which cannot take corrective action to alleviate undesirable changes of the directly controlled variable, is called an open-loop control system.

Figure 1-1:

A. Control system block diagram (open-loop).

B. Control system corresponding to block diagram of Figure 1A.

C. Detailed block diagram of control system shown in Figure 1B.



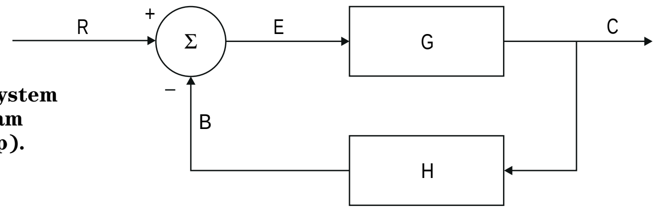
[1] In general, capital letters denote transformed quantities. The quantities may be either Laplace transformed as a function of the complex variable *s*, (e.g., *G(s)*), or Fourier transformed as a function of the frequency variable *jω*, (e.g., *G(jω)*). Functions of *s* are generally abbreviated to their capital letter only (i e., *G(s)* is abbreviated to *G*). Functions of *jω*, however, are never abbreviated.

[2] A transfer function is defined as the ratio of the Laplace transform of the output to the Laplace transform of the input in the absence of all other signals, and with all initial conditions zero. Input and output refer to the signals or variables applied to and delivered from a system or element, respectively.

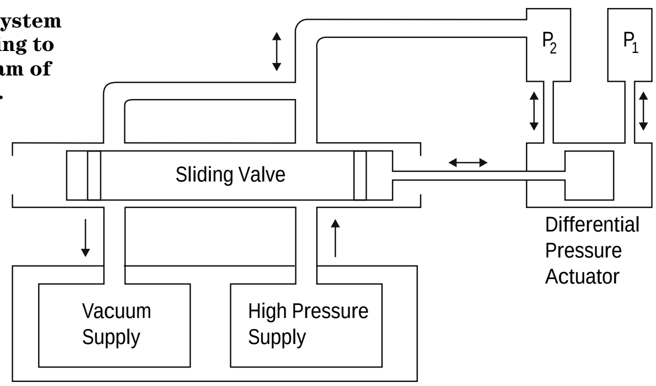
2-2: The Closed-Loop Model

Another basic form of control system is shown in Figure 1-2A. In this system the output *C* is fed back through a functional block with a feedback transfer function *H* and compared to the reference signal *R* via a summing junction. The signal resulting from the difference between *R* and the feedback signal *B* is called the error or actuating signal *E*. The principal advantage of this form of system is that any change in *C*, with *R* remaining constant, causes a change in *E*, ($E = R - B = R - CH$). If the system is operating properly, the change in *E* forces *C* to return to the point where the value of *B* approaches the value of *R*. The effect is that the output is maintained at a desired value despite disturbances to the system. This type of control system is called a closed-loop control system and is defined as any control system in which the directly controlled variable has an effect upon the input quantity in such a manner as to maintain the desired output level.

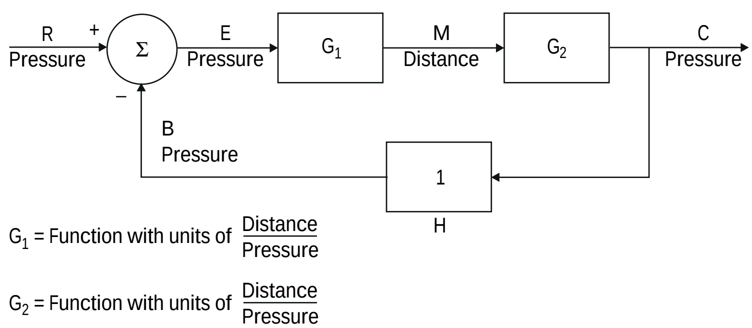
Figure 1-2:



A. control system block diagram (closed-loop).



B. Control system corresponding to block diagram of Figure 1-2A.



C. Detailed block diagram of control system shown in Figure 1-2B.

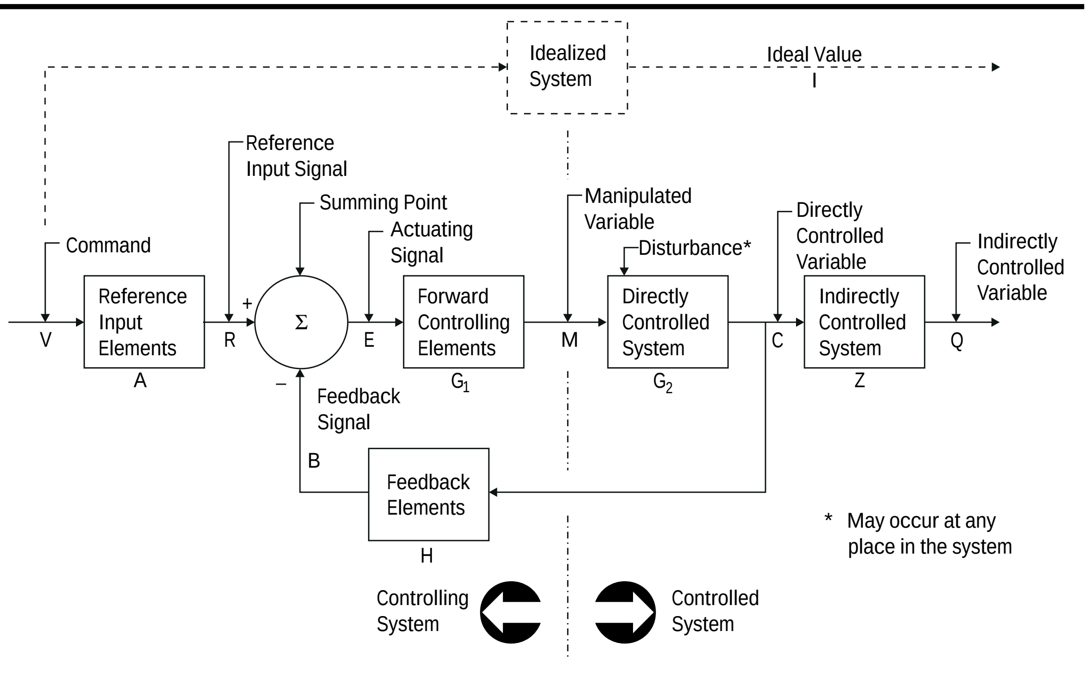
An example of this second form of control system is illustrated by the pressure regulator shown in Figure 1-2B. The objective of this system is to adjust the pressure in Tank 2 (P_2) until it is equal to the pressure in Tank 1 (P_1).

Figure 1-2C is one possible block diagram for this system. In this block diagram the function of the differential pressure actuator is represented by a summing junction and a forward transfer function G_1. The action of the sliding valve is then represented by the forward transfer function G_2. A perfectly valid alternative would be to combine G_1 and G_2 into a single forward transfer function G. The resultant block diagram would then have the same form as Figure 1-2A.

This control system is also an example of a system in which the controlled variable is fed back to the summing junction without any modification; the transfer function, H, is simply equal to 1. This type of control system is called a unity feedback control system.

A general block diagram illustrating most of the elements of an automatic closed-loop control system is shown in Figure 1-3[1].

Figure 1-3: Block diagram of automatic control system.



[1] Figure 1-3 is adapted from the American National Standard ANSI MC85.1M1981, *Terminology for Automatic Control.*

Chapter 3:
Measuring Performance

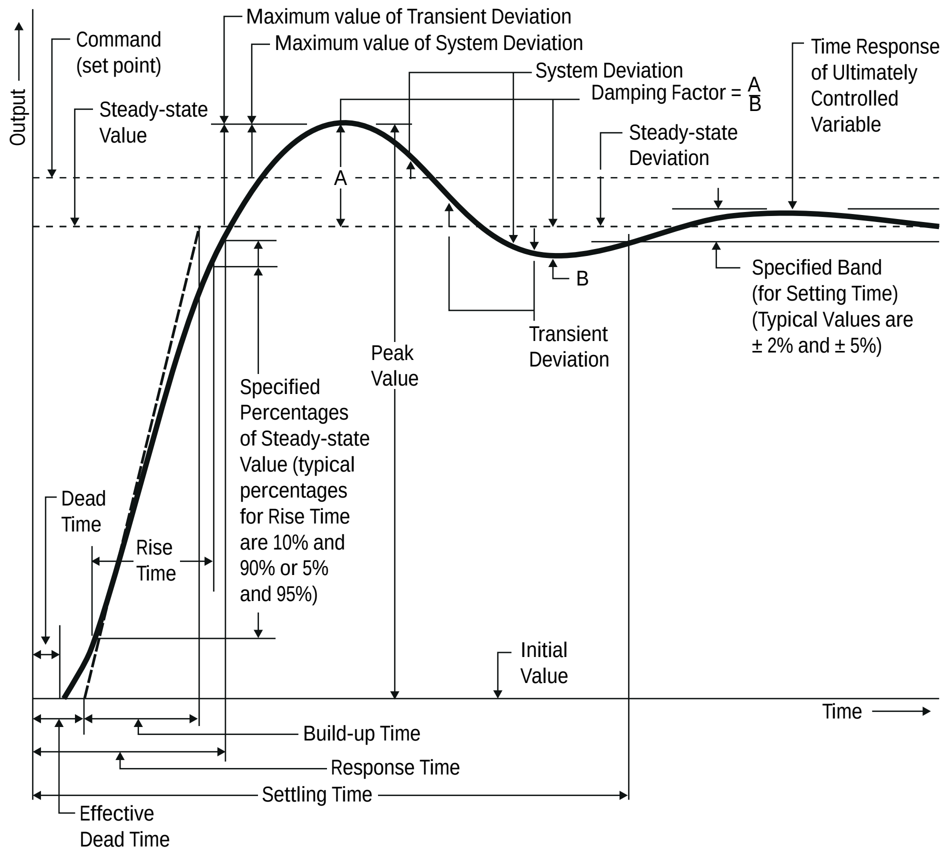
The primary objective in designing a control system is to construct a system that achieves the desired output level as fast as possible and maintains that output with little or no variation. One of the first techniques developed to measure a control system's compliance with these design goals was the step response.

3-1: Time Domain Performance Step Response

The step response is the measured reaction of the control system to a step change in the input. A typical step response and its associated parameters are illustrated in Figure 1-4[1]. The step response has several favorable characteristics which have maintained its universal acceptance and popularity:

- the step stimulus is easy to generate
- the stimulus is easily modeled [$u(t)$] making the solution to the differential equation (used to predict the system's time domain response) much less complicated
- several measurement techniques are available for recording the time domain response to the step input
- key aspects of the control system's performance can be derived from the step response.

Figure 1-4: Typical time response of a system to a step increase of input.



[1] Figure 1-4 is adapted from the American National Standard ANSI MC85.1M-1981, *Terminology for Automatic Control.*

There are several measures of performance which can be derived from the step response. The rise time of the step response provides a measure of how fast a system can initially achieve the desired output level. The maximum overshoot (shown in Figure 1-4 in terms of either peak value or maximum value of transient deviation) provides a relative measure of the maximum output level resulting from a specific input. The steady-state deviation indicates a constant error in achieving a desired output. Settling time, perhaps the most significant parameter, is a measure of how long it takes the system to settle to its steady-state value.

If the system never settles to its steady-state value (for example, it constantly oscillates about a desired output), the system is considered unstable. Taken one step further, the settling time can be interpreted as a relative measure of stability, with a short settling time considered more stable than a long settling time.

In addition to the step response, there were two other early stimulus signals: the ramp function [$tu(t)$] and the parabolic function [$t^2u(t)$]. These signals provided the same simplicity in modeling as the step response and also provided a means of measuring a control systems ability to track dynamic signals.

3-2: Frequency Domain Performance

The time domain responses to the step, ramp and parabolic forcing functions were the only universally accepted techniques for measuring the performance of a control system until the early 1930s. It was during this period that three Bell Laboratories scientists, H.S. Black, H.W. Bode and H. Nyquist, were doing pioneering work on the characterization of control systems in the frequency domain. In an attempt to provide amplifiers with better linearity, Black began a rigorous study of the effects of negative feedback on electronic amplifiers (a basic form of automatic closed-loop control system). Early experiments resulted in several observations including improved linearity and, in some cases, unexpected oscillations in the amplifier's output. It was the unexpected oscillations which inspired Nyquist to study the cause of such instabilities in closed-loop control systems. From his studies, Nyquist discovered that the stability of a closed-loop system could be determined from a simple frequency response plot. Before discussing Nyquist's discovery, it is helpful to review a few of the basic definitions and concepts associated with the frequency domain aspects of a control system.

3-2.1: Frequency Domain Terms and Definitions

One of the most important transfer functions associated with a closed-loop control system relates the directly controlled variable C to the reference input. The ratio C/R is referred to as either the control ratio or the closed-loop transfer function; this note refers to it as the latter. By solving for C/R in terms of G and H we have: $C/R = G/(1 + GH)$, as shown in Figure 1-5. As previously mentioned, capital letters with no subscripts represent transformed quantities expressed as a function of s. The closed-loop transfer function can therefore be expressed as:

$$\frac{C(s)}{R(s)} = \frac{G(s)}{1 + G(s)\,H(s)} = \frac{G(s)}{1 + GH(s)}$$

Important values of s are those values which set the numerator and/or denominator of the closed-loop transfer function equal to zero. Values of s which set the numerator to zero are called zeros of the closed-loop transfer function or closed-loop zeros. Values of s which set the denominator equal to zero (i.e., s such that $1 + GH(s) = 0$) are called poles of the closed-loop transfer function or closed-loop poles.

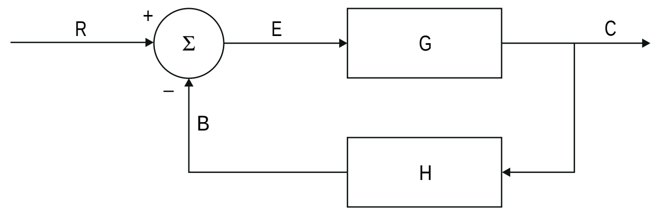
At this point it is important to note that the complex variable s can be further expressed in terms of the variables σ and $j\omega$. That is, $s = \sigma + j\omega$ where σ represents the real or damping component of s, and $j\omega$ represents the imaginary or frequency component of s.

A common tool used to study control systems is a graph called the s-plane. The s-plane is a two-dimensional Cartesian graph which represents values of *s*. The ordinate of the s-plane represents the imaginary part of *s* (i.e., $j\omega$), and the abscissa represents the real part of *s* (i,e., σ). If values of *s* which constitute the closed-loop poles are plotted with X's on the s-plane and the values which constitute closed-loop zeros are plotted with 0's, the result is a pole/zero plot of the closed-loop transfer function as shown in Figure 1-6.

When the magnitude of the closed-loop transfer function is plotted as a third axis of the s-plane, the effects of the poles and zeros on the magnitude of the closed-loop transfer function at any value of s can be quickly realized as shown in Figure 1-7.

Figure 1-7 shows only the left half of the s-plane to illustrate the contour of $|C/R|$ for values of *s* along the $j\omega$ axis (i.e., for values of *s* equal to $0 + j\omega$). This contour is significant in that it represents the same curve produced by evaluating the magnitude of the Fourier transform of *c* divided by the Fourier transform of *r* for positive values of (i.e., $|C(jw)/R(j\omega)|$ for values of $\omega \geq 0$). Therefore, this contour also represents the gain-versus-frequency plot obtained by physically measuring the gain of a control system between its input and output.

Figure 1-5: Solving for *C/R* in terms of *G* and *H*.



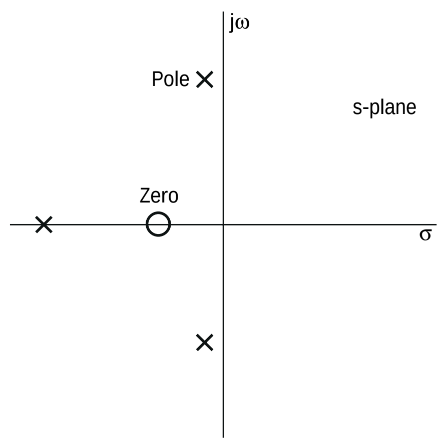
C = EG
E = R - B
B = CH
Solving for C in terms of C, G, H and R we have:
C = RG - BG
C = RG - CHG

Solving for $\frac{C}{R}$ we have:

$$C = \frac{RG}{1 + GH}$$

$$\frac{C}{R} = \frac{G}{1 + GH}$$

Figure 1-6: Pole/zero plot for the closed loop transfer function $\frac{G(s)}{1 + G(s)H(s)}$.



A similar diagram can be drawn for the phase of $C(s)/R(s)$ as shown in Figure 1-8. Again the contour presented by the values of $\angle C(s)/R(s)$ along the $s = 0 + j\omega$ axis represent $\angle C(j\omega)/R(j\omega)$ for positive values of ω. This contour also represents the phase-versus-frequency plot obtained by physically measuring the phase shift of a control system between its input and output.

The information provided by the highlighted contours in Figures 1-7 and 1-8 represents the frequency-dependent relation between steady-state sinusoidal input signals ($R(j\omega)$) and the resulting steady-state sinusoidal output signals ($C(j\omega)$), that is, they represent the frequency response of the device characterized by C/R.

For transfer functions in general, the information produced by evaluating the Fourier transform for all values of $j\omega$ can be regarded as a subset of the overall contour produced by evaluating the Laplace transform for all values of s. The Fourier transform of a transfer function evaluated for positive values of ω also represents the physically measured gain and phase relationship (i.e., frequency response) between the input and output of the device modeled by the transfer function.

Figure 1-7: Magnitude plot of $\frac{G(s)}{1 + GH(s)}$ versus values of s.

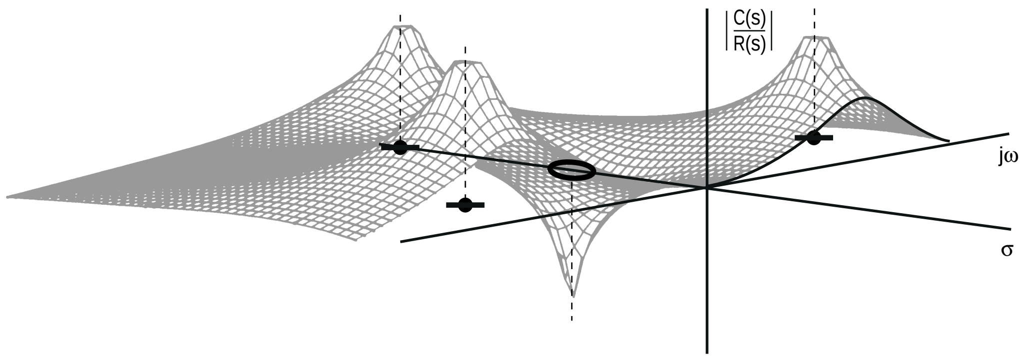
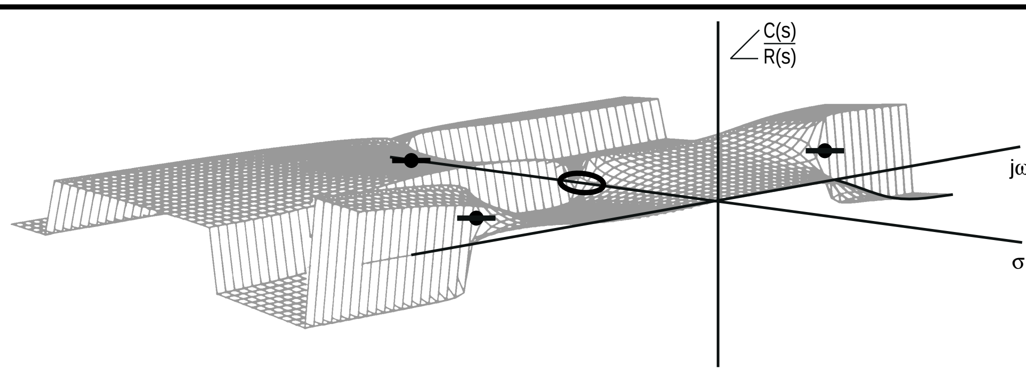


Figure 1-8: Phase plot of $\frac{G(s)}{1 + GH(s)}$ versus values of s.



3-2.2: Nyquist's Stability Criterion (s-plane)

With the evaluation of transfer functions over the s-plane well established, the fundamental condition for stability discovered by Nyquist, can now be presented. Simply stated, for a control system to be stable, there can be no closed-loop poles in the right half of the s-plane. (Poles on the $j\omega$ axis are not directly addressed but are generally considered to represent instability.) This relationship between closed-loop pole locations and system stability constitutes Nyquist's Stability Criterion as applied to the s-plane. This relationship can be extremely useful in predicting the stability of a system if the position of each closed-loop pole is known. Trying to determine the exact location of closed-loop poles from measured data without a computer, however, can often be a difficult task. Fortunately, Nyquist's original work included a very useful technique for evaluating the presence of closed-loop poles in the right-half plane without necessarily knowing their exact locations. To examine this technique closely, however, we will need a few more terms and definitions.

In the preceding paragraphs it was established that the roots of the equation $1 + GH(s) = 0$ (i.e. values of s for which $GH(s) = -1$) were the closed-loop poles and the sole factor in determining if the system would be stable. Because of its influence on the stability of the system and, ultimately, the character of the time domain response, the equation $1 + GH(s) = 0$ is known as the characteristic equation.

From the characteristic equation it is apparent that the term $GH(s)$ contains all the information concerning the location of the closed-loop poles ($GH(s)$ is understood to represent the transfer function of all of the elements in the loop between the error signal (E) and the feedback signal (B). The function $GH(s)$ is called the loop transfer function or open-loop transfer function and is denoted by either $GH(s)$ or $B(s)/E(s)$, as shown in Figure 1-9. This note uses the notation $GH(s)$ and refers to it as the open-loop transfer function.

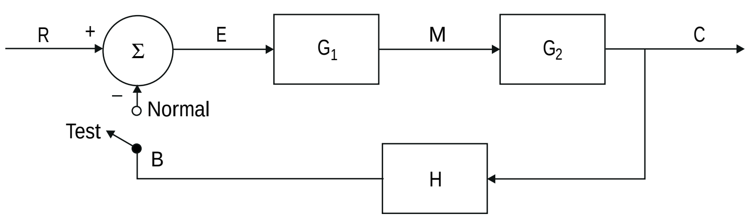


Figure 1-9: Open-loop transfer function of a closed-loop control system.

$GH(s) = G_1G_2H(s) = \frac{B(s)}{E(s)}$ for switch in "test" or open position (therefore the name "open-loop transfer function")

Also: $B(s) = EG_1G_2H(s)$ for switch in either "test" or "normal" position

$$\frac{B(s)}{E(s)} = G_1G_2H(s)$$

At this point it is worthwhile to recognize that *G(s)* and *H(s)* are themselves generally ratios of polynomials in s. *G(s)* and *H(s)* can therefore be represented by:

$$G(S) = \frac{G_n(s)}{G_d(s)} \quad and\ H(s) = \frac{H_n(s)}{H_d(S)}$$

where the subscripts *n* and *d* indicate the numerator and denominator portions of *G(s)* and *H(s)*, respectively. If the closed-loop transfer function is reformulated in terms of the numerator and denominator of *G(s)* and *H(s)* we have:

$$\frac{C(s)}{R(s)} = \frac{G(s)}{1 + GH(s)} = \frac{\dfrac{G_n(s)}{G_d(s)}}{1 + \dfrac{G_n(s)H_n(s)}{G_d(s)H_d(s)}} = \frac{\dfrac{G_n(s)H_d(s)}{G_d(s)H_d(s)}}{\dfrac{G_d(s)H_d(s) + G_n(s)H_n(s)}{G_d(s)H_d(s)}}$$

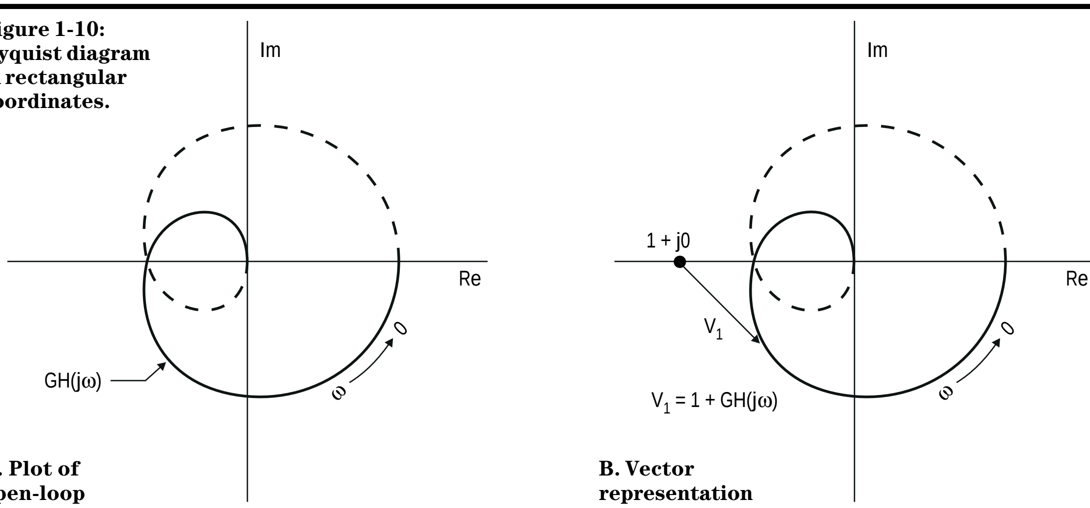
The objective of expressing the closed-loop transfer function in this manner is to illustrate that the term *1+ GH(s)* itself has poles and zeros, and that it is the zeros of this term that determine the poles of the closed-loop transfer function. It is also worth noticing that the zeros of the closed-loop transfer function are the roots of the equation $G_n(s)H_d(s) = 0$.

3-2.3: Nyquist Diagrams

It was Nyquist's observation that the frequency response of the open-loop transfer function (i.e. *GH(jω)*) can be used to determine if there are any zeros of the term *1 + GH(s)* (and therefore poles of the closed-loop transfer function) in the right half of the s-plane. To make this determination, *GH(jω)* is first plotted on a two-dimensional Cartesian coordinate system whose ordinate is the imaginary part of *GH(jω)* and abscissa is the real part of *GH(jω)*. The complex conjugate of the frequency response curve is then plotted on the same graph, as shown by the dashed line in Figure 1-10A.

The next step is to establish a vector V_1 whose tail is affixed to the point $-1 + j\phi$. If the head of the vector is then placed anywhere along the curve of *GH(jω)*, the vector then represents the quantity *1 + GH(jω)*, as illustrated in Figure 1-10B.

Figure 1-10: Nyquist diagram in rectangular coordinates.



A. Plot of open-loop frequency response and its complex conjugate.

B. Vector representation of the quantity *1 + GH(jω)*.

3-2.4: Nyquist's Stability Criterion (Nyquist diagram)

At this point, Nyquist's Stability Criterion states that as the head of the vector traces the *GH(jω)* curve in the direction of increasing positive frequency, the net number of complete rotations *N* is equal to the number of poles P_r of the term $1 + GH(s)$ in the right half of the s-plane minus the number of zeros Z_r of the term $1 + GH(s)$ in the right half of the s-plane. That is:

$$N = Z_r - P_r$$

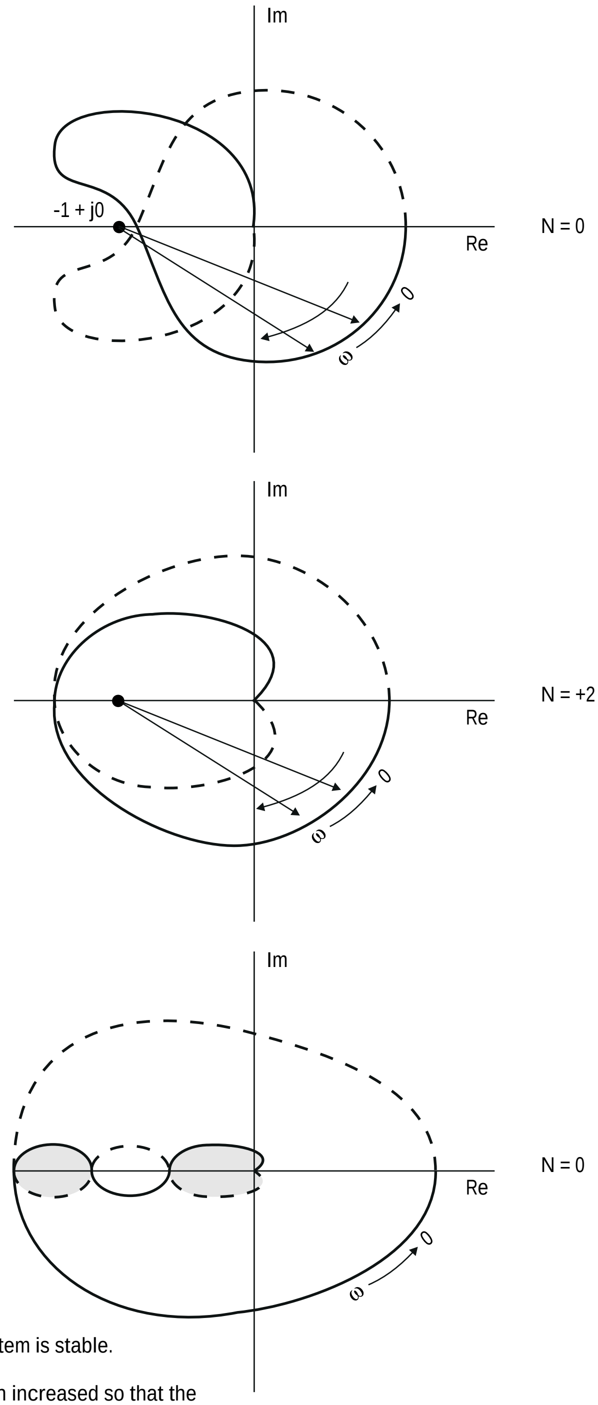
where *N* is positive for clockwise rotations and negative for counterclockwise rotations. We therefore know that a system is stable only if $N = -P_r$. It is a general consensus that for most real systems $P_r = 0$ and, therefore, $N = Z_r$. When this assumption is true, the condition for stability can be restated as: a system is stable if and only if $N = 0$. Figure 1-11 illustrates examples of systems which are stable, conditionally stable, and unstable.

Figure 1-11: Nyquist diagrams

A. Stability

B. Instability

C. Conditional stability



As illustrated, the system is stable.

However, if the gain in increased so that the area shaded encloses the -1 + j0 point, then N = 2, and the system is again unstable.

Also, if the gain is decreased so that the area shaded in gray encloses the -1 + J0 point, then N = 2, and the system is again unstable.

Therefore by either increasing or decreasing the gain, the system becomes unstable (i.e., the system is conditionally stable).

3-2.5: Magnitude and Phase Contours

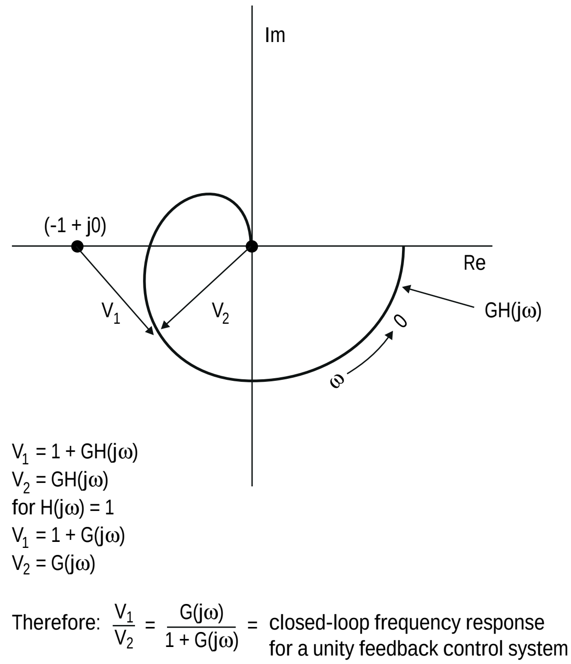
The Nyquist diagram can also be used to evaluate the closed-loop frequency response from the open-loop frequency response if the system being analyzed has unity feedback. For $H(j\omega) = 1$ the closed-loop transfer function for real frequencies becomes:

$$\frac{C(jw)}{R(jw)} = \frac{G(jw)}{1 + G(jw)}$$

If another vector V_2 is added to the Nyquist diagram so that it projects from the origin and meets with the vector V_1 at the curve of $G(j\omega)$, then the closed-loop transfer function can be represented by the ratio of V_2/V_1, as shown in Figure 1-12.

Useful tools for evaluating the performance of a unity-feedback control system are magnitude contours (often referred to as M-contours). A magnitude contour is a locus of points for which the ratio of the magnitudes of V_1 and V_2 is a constant. When plotted on the Nyquist diagram, a magnitude contour will appear as a circle (except when $|V_2/V_1| = 1.0$), as shown in Figure 1-13. When the open-loop transfer function is plotted on a Nyquist diagram with magnitude contours, the maximum gain of the closed-loop transfer function can be identified as the value of the magnitude contour which is tangent to the plotted curve, as shown in Figure 1-4. A similar diagram can also be constructed for constant values of phase difference between V_1 and V_2. Plots of constant phase are called phase contours or N-contours.

Figure 1-12: Evaluation of closed-loop response from open-loop frequency response (unity feedback).



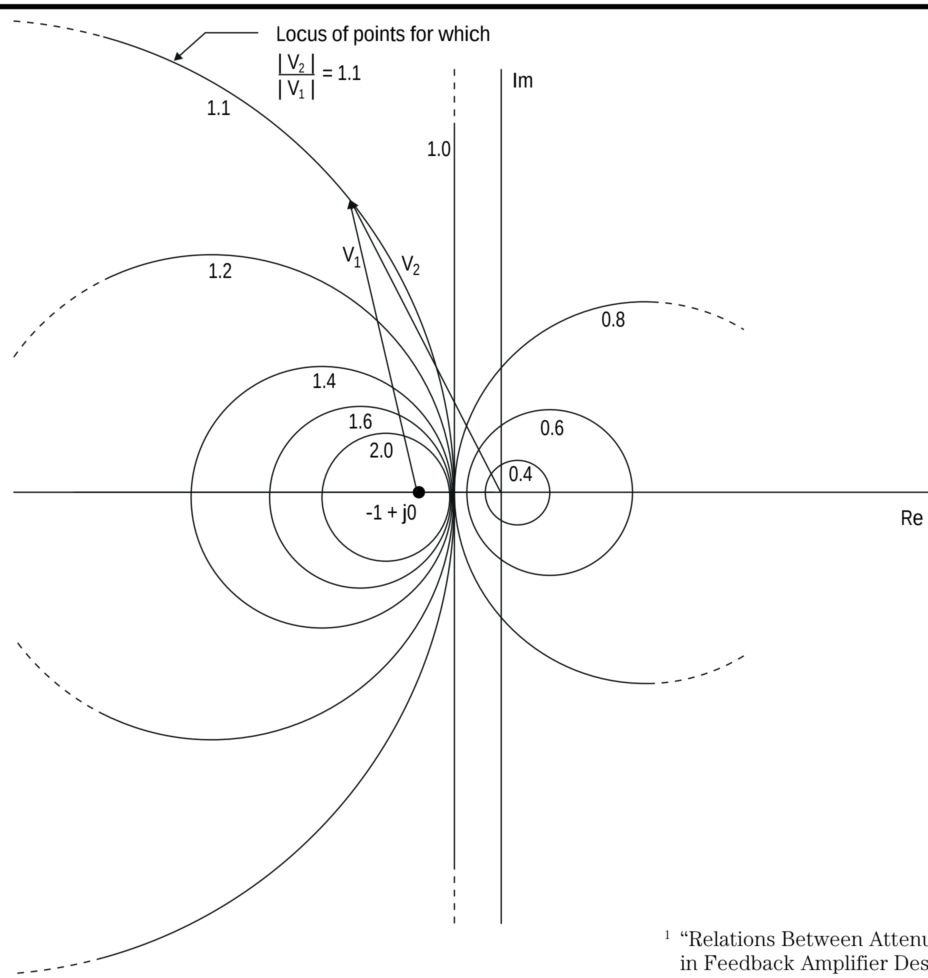
From Figure 1-14 it can be seen that as the curve of the open-loop frequency response approaches *-1 + j0*, the maximum gain of the closed-loop frequency response approaches infinity (for either unity or nonunity feedback). It can be shown that the movement of the open-loop frequency response toward *-1 + j0* is directly related to the movement of the closed-loop poles toward the right half of the s-plane, therefore causing the system to become less stable.

3-2.6: Bode on Stability

When H. Bode published his paper, "Relations Between Attenuation and Phase in Feedback Amplifier Design" in 1940, he noted that for a system to be absolutely stable it can only cross the negative real axis between the origin and *-1 + j0*. According to Bode, crossing the negative real axis anywhere else produces a system which is either unstable or conditionally stable; neither of which is generally desirable.

Bode's statement is much easier to interpret if the scale of the Nyquist diagram is changed from rectangular coordinates to polar coordinates, as shown in Figure 1-15. The *-1 + j0* point then represents a magnitude of 1 and a phase of -180 degrees. Using a polar Nyquist diagram, Bode's observation can be restated as: for a closed-loop system to be absolutely stable, the phase of the open-loop frequency response should not exceed 180 degrees until its magnitude becomes less than one.

Figure1-13: Loci of constant magnitude contours.



[1] "Relations Between Attenuation and Phase in Feedback Amplifier Design," *Bell System Tech. J.,* 19, 421-454 (July 1940).

Figure 1-14:
Determining the maximum gain of the closed-loop frequency response of a unity-feedback control system.

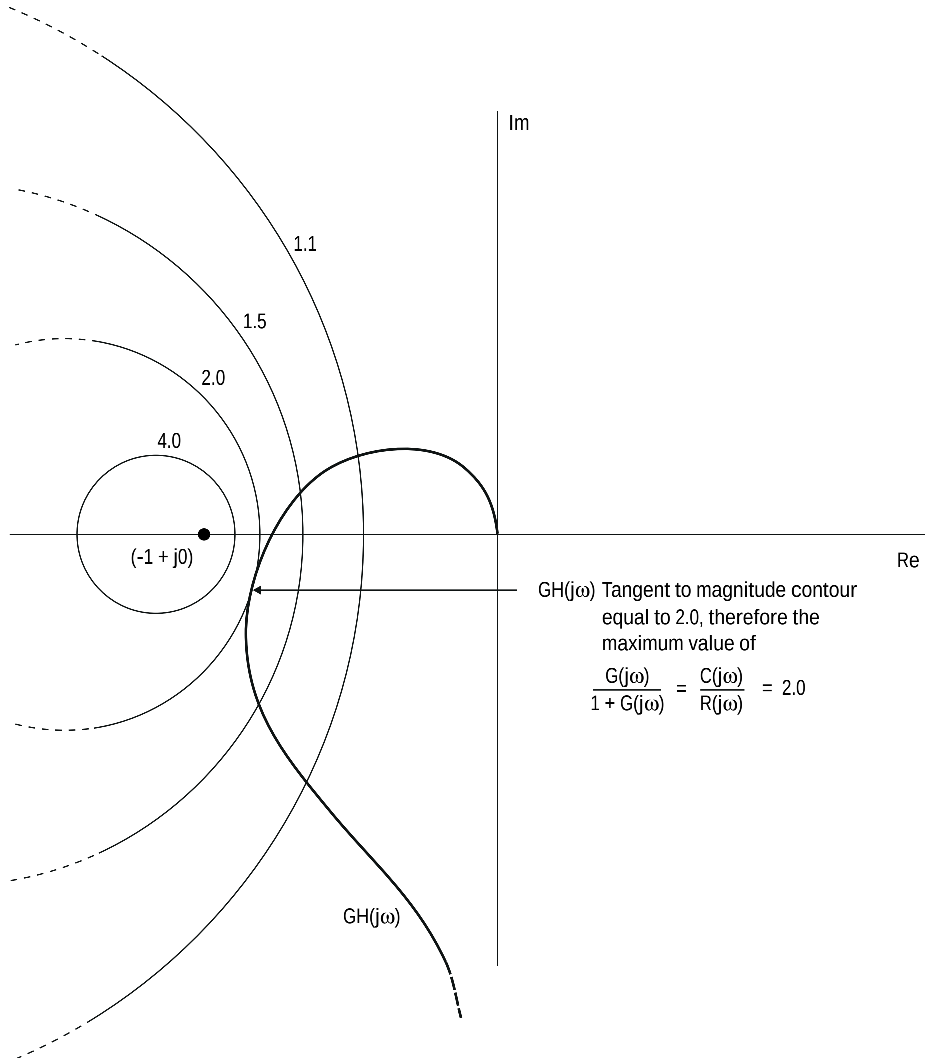
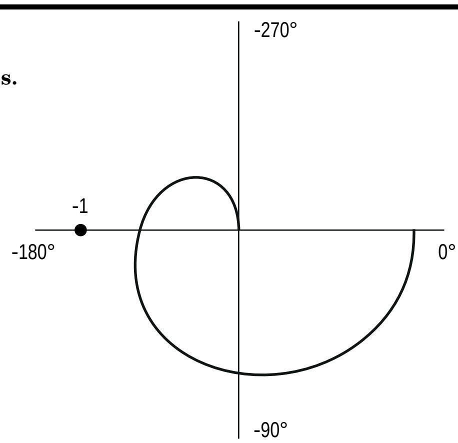


Figure 1-15:
Nyquist diagram in polar coordinates.



3-2.7: Gain Margin and Phase Margin

Bode also explained that an open-loop frequency response curve which just met this criterion would rarely produce a stable system since any small variations in the system's performance would place the response in an unstable region. He therefore suggested that a certain amount of margin should be allotted for both the phase and gain values as they approached the point representing a magnitude of 1 and a phase shift of -180 degrees. These margins are now standard performance parameters known as the phase margin and gain margin.

Phase margin is defined as 180 degrees minus the absolute value of the phase of the open-loop frequency response at the point where the magnitude of the open-loop frequency response (i.e., the open-loop gain) is equal to one. That is:

phase margin = $180 - |\angle GH(j\omega)|$
where $|GH(j\omega)| = 1$

Gain margin is defined as the reciprocal of the open-loop frequency response gain at the point where the phase of the open-loop frequency response is equal to minus 180 degrees. That is:

gain margin = $\dfrac{1}{|GH(j\omega)|}$

where $GH(j\omega) = -180$ degrees

The gain margin therefore represents the amount the open-loop gain can be increased before it reaches a magnitude of 1. Examples of gain margin and phase margin are shown in Figure 1-16.

The importance of the gain and phase margin can be fully appreciated when they are compared with, and shown to correlate with, the time domain parameters of the step response. For example, for a system whose response characteristics are dominated by a pair of complex poles (a very common case), the following relationships can be observed. An increase or decrease in the system's frequency independent gain[1] will cause both the gain margin and phase margin to decrease or increase, respectively. For the case in which the gain is increased, the following events will occur:

- the gain margin and phase margin will decrease;
- the maximum overshoot will increase;
- the rise time will decrease;
- and, in some cases, the steady-state deviation will decrease.

From this series of interactions it can be seen that the development of a control system is generally a trade-off between the desired performance characteristics. Although each control system has unique requirements, minimum acceptable levels of gain margin and phase margin are typically 2 (or 6 dB)[2] and 30 degrees, respectively.

In addition to gain margin and phase margin, there are several other performance quantities such as the system type and steady-state error coefficients which can be extracted from a Nyquist diagram. Unfortunately, a complete description of these quantities is beyond the scope of this document (several references for further study are listed at the end of this note). It can be assumed, however, that the key performance characteristics of a control system can be adequately characterized with a Nyquist diagram.

One shortfall of the Nyquist diagram is the difficulty encountered when attempting to predict the effects of changes to a control system. Most alterations (other than a change in frequency independent gain) require a significant number of calculations, or a new measurement, to accurately obtain the correct Nyquist diagram. As a result, several other analysis techniques were developed to make the design and analysis of a control system easier. These are discussed in detail in the following chapter.

[1] Frequency independent gain is also referred to as proportional amplification and is represented by the variable *K*. A more detailed explanation is provided in the discussion of the root locus diagram, Section 4-4.

[2] dB represents a unit of comparison known as the decibel. It is calculated for both voltage and power rations with respective formulas for each being: dB = 10 log (power ratio) and dB = 20 log (voltage ration). See Hewlett-Packard Application Note 243, *The Fundamentals of Signal Analysis*, p. 5, for further details.

Figure 1-16: Measuring gain margin and phase margin on a Nyquist diagram[1].

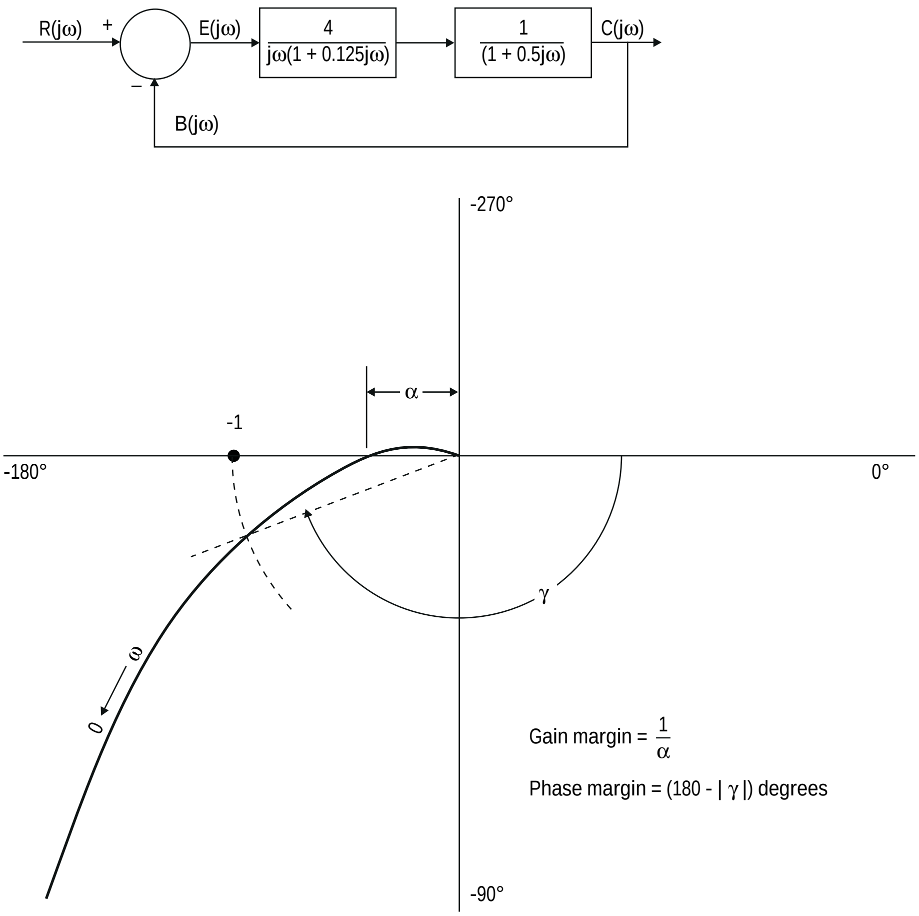


Table 1-1: Generalized relationships between time domain and frequency domain performance parameters relative to an increase in the frequency independent gain.

	Frequency Domain		Time Domain			
Frequency Independent Gain	Gain Margin	Phase Margin	Settling Time	Rise Time	Maximum Overshoot	Steady-state Deviation
⇧	↓	↓	↕ (dashed)	↓	↑	↓

[1] Figure 1-16 is adapted from the American National Standard ANSI MC85 1M-1981, *Terminology for Automatic Control*.

Chapter 4:
More Tools for Design and Analysis

There is perhaps no design tool which has gained as much popularity as the diagram which Bode presented in his 1940 paper, “Relations Between Attenuation and Phase in Feedback Amplifiers.” This chapter looks at the famous Bode diagram and two other popular design and analysis tools; the Nichols diagram and root locus diagram.

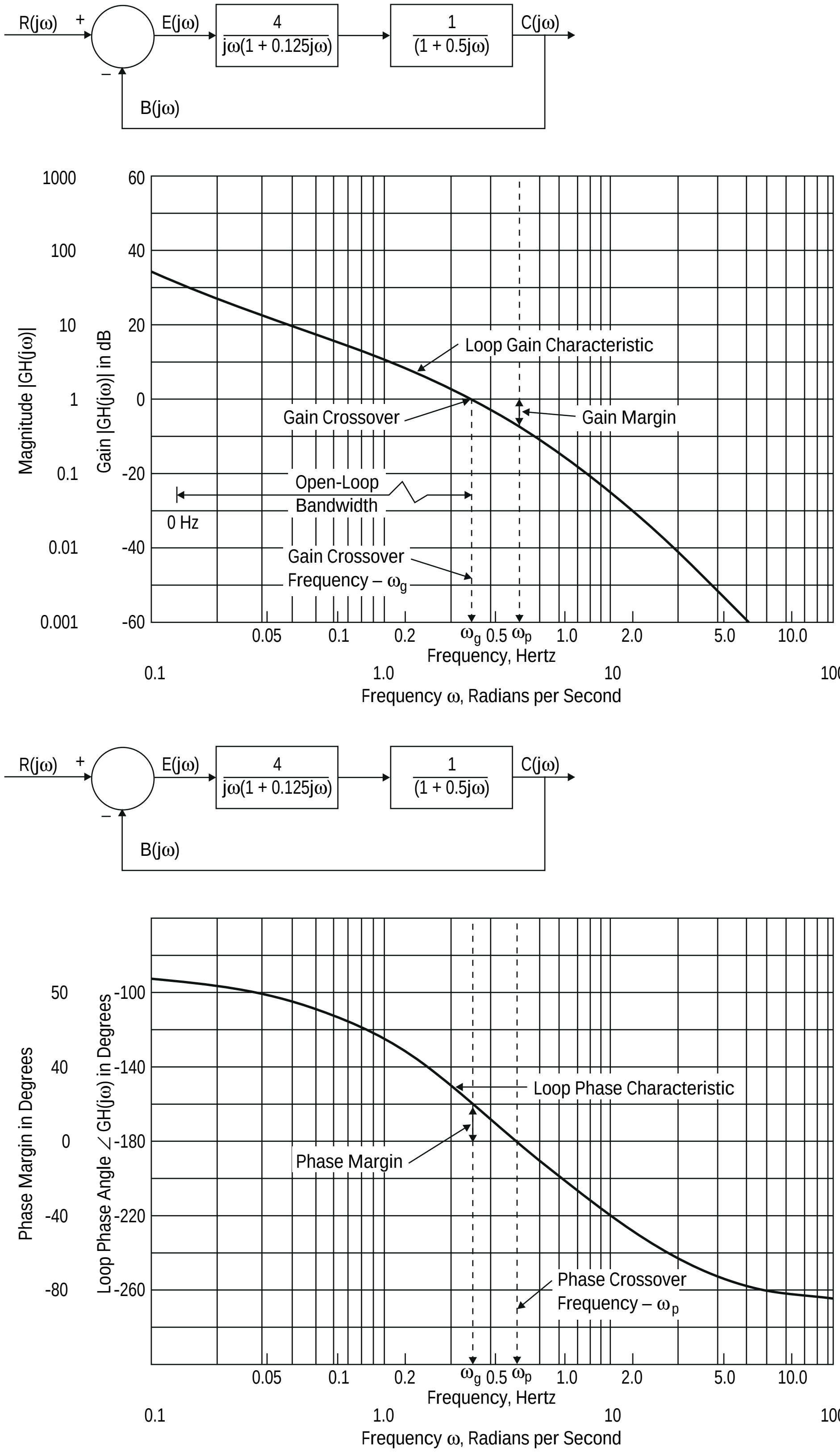
4-1: The Bode Diagram

The Bode diagram is similar to the Nyquist diagram in that it also represents a plot of the open-loop frequency response. However, the Bode diagram considers the gain and phase of the response separately by providing a plot of each versus frequency. The plot of open-loop gain versus frequency is called the loop gain characteristic and the plot of open-loop phase versus frequency is called the loop phase characteristic, as shown in Figure 1-17.

Bode diagrams use logarithmic units (i.e., dB) for gain and logarithmic scales for frequency; phase is the only parameter represented linearly. The use of logarithmic scales and units provides the Bode diagram with three key advantages. First, by displaying gain in units of dB, a much wider range of gain levels can be displayed on a single plot. Second, the effect on the open-loop frequency response of adding a new component in a control loop can be calculated through simple addition rather than multiplication. That is, by plotting the frequency response of a new component on the same Bode diagram as the original response, the frequency response of the new system can be calculated by graphically adding the two plots. Third, the logarithmic scales and units facilitate a technique for quickly estimating the frequency response of an analytic transfer function. This last point is a major topic of Bode’s paper. In his paper, Bode presented a relatively simple set of procedures for constructing a set of curves which would closely estimate the actual frequency response of a transfer function without ever actually calculating or measuring the response.

An equally powerful tool was the ability to apply Bode’s construction procedures in reverse. That is, to obtain information about the analytic transfer function from the measured frequency response.

Figure 1-17: Bode diagrams showing frequency response for a typical open-loop transfer function[1]**.**



[1] Figure 1-17 is adapted from the American National Standard ANSI MC85. 1M-1981, *Terminology for Automatic Control*.

4-2: Stability and the Bode Diagram

The Bode diagram also provides a simple check for stability. According to Bode's interpretation of Nyquist's Stability Criterion, for a system to be absolutely stable, the loop gain characteristic must be less than one before the loop phase characteristic exceeds (becomes more negative than) 180 degrees. On a Bode diagram, this means the frequency at which the loop gain characteristic becomes equal to 0 dB (i.e., the gain crossover frequency) must be lower than the frequency at which the loop phase characteristic becomes equal to – 180 degrees (i.e., the phase crossover frequency).

The phase margin, gain margin, and open-loop bandwidth[1] of a system can also be read directly from the Bode diagram, as shown in Figure 1-17.

One of two disadvantages of the Bode diagram is that there is no technique for directly relating the open-loop frequency response to the closed-loop frequency response (as was possible with the magnitude and phase contours of the Nyquist diagram). However, the frequency response information from a Bode diagram can be directly transferred to a Nyquist or Nichols diagram to evaluate the closed-loop frequency response. (It is important to note that a reverse exchange of information, that is, from a Nichols or Nyquist diagram to a Bode diagram, may not be possible due to the loss of frequency information in both the Nichols and Nyquist diagrams.)

A second disadvantage of the Bode diagram is its limited ability to verify the stability of control systems which are conditionally stable. Fortunately, conditionally stable systems are rarely designed intentionally and can be analyzed by transferring the frequency response data to a Nyquist diagram if necessary.

4-3: The Nichols Diagram

The Nichols diagram (also known as the log magnitude-angle diagram) is essentially a combination of the Nyquist and Bode diagrams. It is conceptually similar to the Nyquist diagram in that it plots the magnitude of $GH(j\omega)$ versus the angle of $GH(j\omega)$ as a function of frequency (ω) on a single graph, as shown in Figure 1-18. Its structure, however, more closely resembles a Bode diagram in that it uses a rectangular coordinate system and scales gain in units of dB.

The Nichols diagram incorporates some of the advantages provided by the Bode and Nyquist diagrams into a single graph. By plotting gain versus phase, the Nichols diagram allows the construction of magnitude and phase contours similar to those used on the Nyquist diagram. However, by scaling the gain in units of dB, a single set of contours can be applied over a much broader range of gain levels. A single Nichols diagram can therefore provide a direct readout of the closed-loop frequency response (of a unity feedback control system) for a much broader range of open-loop gains. Nichols diagrams which have a large set of magnitude and phase contours drawn on them are often called Nichols charts.

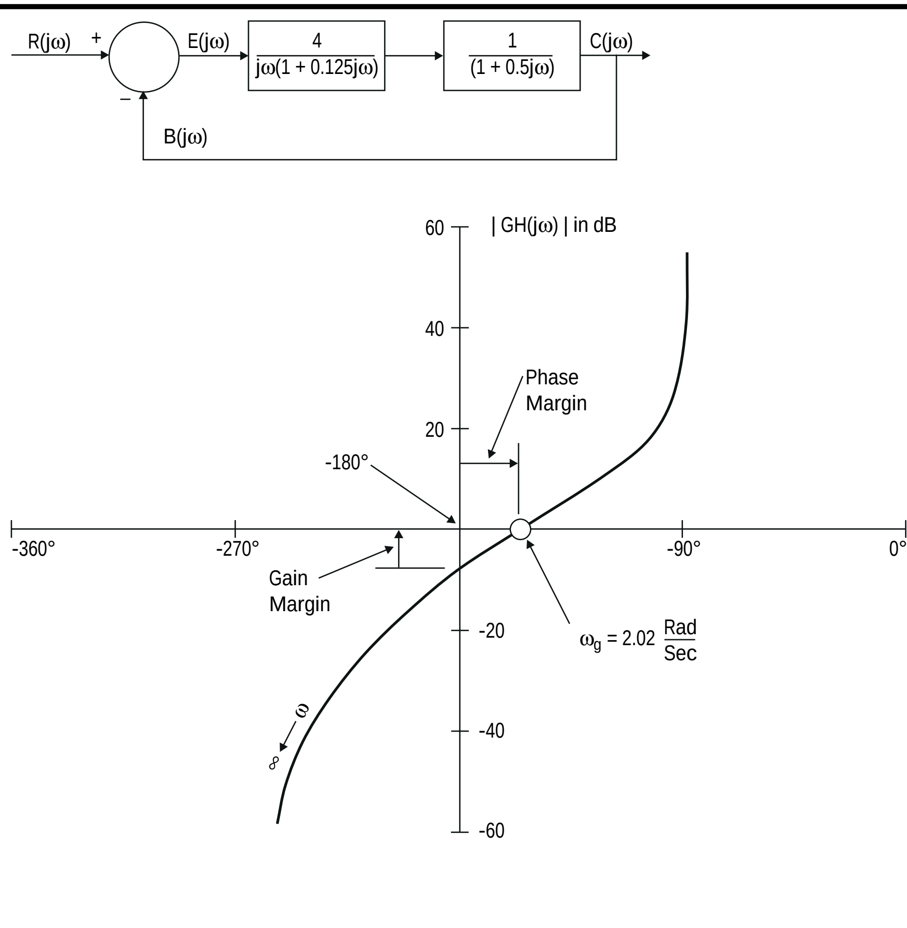
[1] Open-loop bandwidth is defined as the frequency span between 0 Hz and frequency at which the gain of the open-loop frequency response is equal to 1.

Gain margin and phase margin can also be read directly from the Nichols diagram. However, to obtain the open-loop bandwidth, the gain crossover frequency must be evaluated while the plot is being constructed, and then marked on the graph, as shown in Figure 1-18.

The main disadvantage of the Nichols diagram is the difficulty in plotting *GH(jω)* directly from the transfer function. Unlike the Bode diagram, there is no simple set of rules which provides a quick estimation of a transfer function's frequency response. It is therefore difficult to predict the effect of a compensation circuit on the system's performance.

The Nichols diagram is also limited in its ability to verify the stability of conditionally stable systems. However, like the Bode diagram, the frequency response information can be transferred to the Nyquist diagram for analysis.

Figure 1-18: Nichols diagram (log magnitude versus angle diagram) for a typical open-loop transfer function.



4-4: The Root Locus Diagram

The root locus diagram (or root locus plot) was developed by W.R. Evans and presented in his 1950 paper, "Control System Synthesis by Root Locus Method"[1]. The root locus diagram is a departure from the frequency response plotting techniques used by the Bode, Nichols and Nyquist diagrams. All three of the latter techniques use the frequency response of the open-loop transfer function, $GH(j\omega)$, to gain information about the relative location of the closed-loop poles in the s-plane. The root locus diagram, however, uses the location of the open-loop poles and zeros in the s-plane to predict the actual location of the closed-loop poles. Before discussing the root locus diagram further, it is again necessary to introduce another concept.

The symbol G was previously defined as a transfer function whose gain and phase characteristics change with respect to the variable s or $j\omega$. It can, however, be divided into two factors: 1) a proportional amplification often denoted as K, which is independent of s or $j\omega$ and associated with a dimensioned scale factor relating the units of input and output; 2) a dimensionless factor often denoted as G which is dependent on s or $j\omega$. Therefore, if K is used as a prefix when expressing a transfer function, it is understood that K represents a gain value extracted from the transfer function which is independent of s or $j\omega$. For example, if the open-loop transfer function is expressed as $KGH(j\omega)$, it is understood that K is the gain portion of $GH(j\omega)$ which is independent of $j\omega$.

The objective of the root locus diagram is to graphically locate values of s which set the open-loop frequency response equal to -1, that is s such that $GH(s) = -1$. These values of s will therefore also represent roots of the characteristic equation $1 + GH(s) = 0$ and, further, represent the location of the closed-loop poles.

The power of the root locus technique is its recognition of the frequency independent gain of the open-loop transfer function, K of $KGH(s)$. The root locus technique recognizes that for each value of K there is a unique set of values for s which satisfy the equation $KGH(s) = -1$. For example, if K is set equal to 3 in the open-loop transfer function:

$$KGH(s) = \frac{K}{s(1 + 0.125s)(1 + 0.5s)}$$

then there exists a unique set of values of s, in this case those shown in Figure 1-19A, for which $3GH(s) = -1$(or alternatively, $GH(s) = -1/3$). If K is set equal to 4, then there exists another set of values of s for which $GH(s) = -1/4$, as illustrated in Figure 1-19B. This new set of values for s represents the new locations of the closed-loop poles when K is increased from 3 to 4.

If the unique set of values for s were calculated for each value of K from zero to infinity and plotted on the same graph, the result would be a set of lines which represent a locus of roots to the equation $1 + KGH(s) = 0$ for all possible values of K, as shown in Figure 1-19C. This plot is called a root locus diagram.

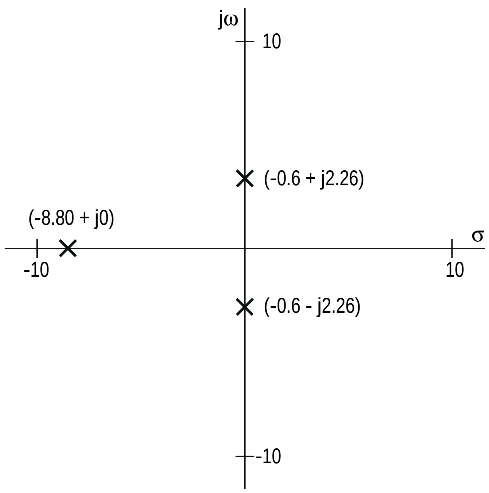
If root locus diagrams were constructed in this fashion, it would require many calculations and make the construction of the diagram much too involved to be of practical value, at least without the aid of a computer. Fortunately, Evans also presented a technique for graphically estimating the root locus diagram based on the location of the open-loop poles and zeros in the s-plane. The procedure is relatively simple and it is not uncommon for people who have mastered the root locus technique to quickly sketch the root locus diagram based solely on the location of the open-loop poles and zeros (i.e., with virtually no calculations). A root locus diagram will therefore generally include designators indicating the position of the open-loop poles and zeros as shown in Figure 1-20.

The root locus diagram is a very powerful design tool since it works directly with the location of the closed-loop poles in the s-plane. However, the root locus technique can only be used if the number and location of the open-loop poles and zeros are known. It is therefore less flexible than the Nyquist or Bode diagrams which need only the measured open-loop frequency response to predict performance and provide design information. It does, however, provide more information during the initial design process and is better suited for the design of complex compensation networks.

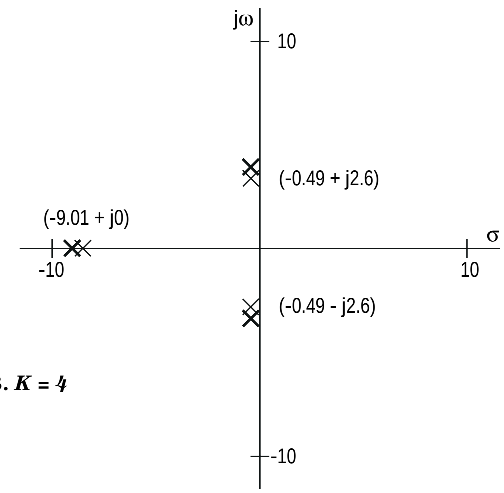
[1] "Control System Synthesis by Root Locus Method," *Trans, AIEE,* 69, 1-4 (Mar 10, 1950).

Figure 1-19: Value of *s* which satisfy the equation

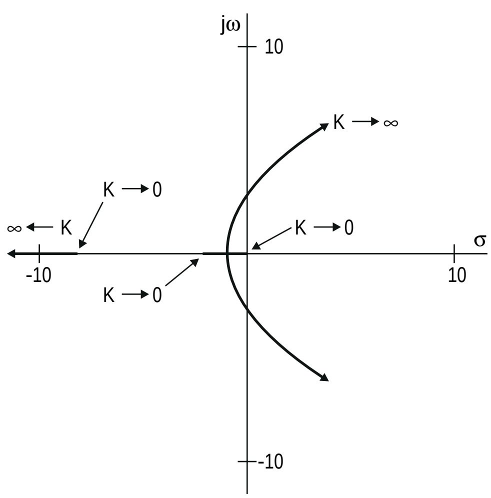
$$GH(s) = \frac{K}{s(1 + 0.125s)(1 + 0.5s)} = -1$$



A. $K = 3$



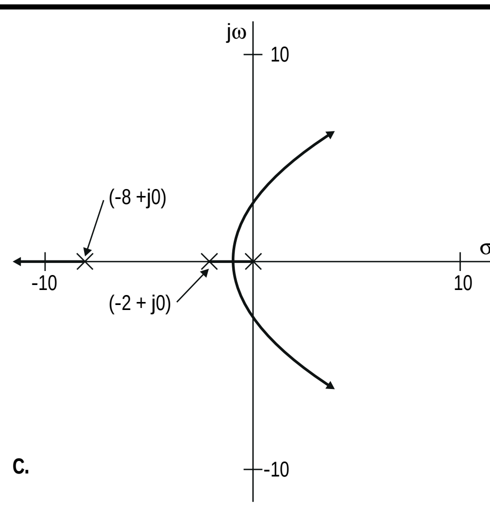
B. $K = 4$



C. $0 \leq K \leq \infty$

Figure 1-20: Root locus diagram of the equation

$$GH(s) = \frac{K}{s(1 + 0.125s)(1 + 0.5s)} = -1$$



C.

Chapter 5: Nonlinear Systems

The design and analysis tools presented so far have all assumed that the control system or subsystem being analyzed is linear. Unfortunately, the vast majority of control systems are actually nonlinear, either by design or by virtue of the components within the system.

There are some very complex analysis tools which deal directly with nonlinearities; however, a very common practice is to obtain an approximation of the system's nonlinear operation which best conforms to a linear response. The approximation can then be used with the tools presented in the previous chapters.

For example, Figure 1-21A shows a typical gain curve (V_{out}/V_{in}) which is essentially linear for input voltages less than V_L and nonlinear for input voltages greater than V_L.

If the system characterized by Figure 1-21A is operated within a narrow range of voltages centered about a voltage V_1, as shown in Figure 1-21B, then the system will operate over a linear region of the curve and can be modeled with the linear equation:

$$V_{out} = aV_{in}$$

or $$\frac{V_{out}}{V_{in}} = a$$

where a is a constant.

If, however, the system operates under the same conditions except at a higher average voltage V_2, as shown in Figure 1-21C, then the system is not operating in a linear region and a linear approximation is required.

Graphically, a linear approximation could be obtained by simply drawing a straight line through the operating region which best fits the gain curve. This approximation, however, would not address the distribution of energy throughout the response spectrum due to the distortion of the output waveform, as shown in Figure 1-21C.

For this type of nonlinearity, a better technique for obtaining a linear approximation of the system's gain is to measure only that part of the response spectrum which is at the same frequency as the input. That is, measure the system gain at the fundamental frequency of the stimulus and ignore all the other frequency components, including those created by system nonlinearities. If a series of both gain and phase measurements are made over a range of frequencies, the results can be plotted to produce a graph of the system's frequency response. The resulting frequency response can then be used to generate a transfer function based solely on the fundamental. Such a transfer function is often called a describing function and is generally considered a good linearized approximation of a system with nonlinearities such as harmonic distortion and intermodulation distortion.

A common technique used to make the measurement described is to stimulate the system with a swept sine wave source and measure both the stimulus and the response with narrow bandpass filters which track the frequency of the source. Test instruments capable of making this type of measurement include network analyzers, frequency response analyzers, and properly equipped Dynamic Signal Analyzers (DSAs).

It is important to note that for any small change in either the mean voltage V_2, or the amplitude of the sine wave itself, the measured frequency response will also change. This change in measurement result due to changes in the testing conditions is a common phenomenon associated with most nonlinear devices.

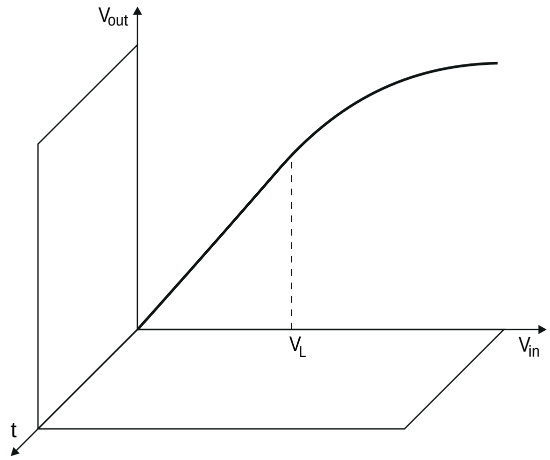
If a nonlinear system is both sensitive to changes in the stimulus signal (as described above) and operated over a wide range of stimulus levels, then there is typically no one unique frequency response or describing function which can accurately model the operation of the system.

As a practical solution to this problem, a nonlinear device is typically tested under conditions which closely approximate the actual operating conditions of the system. If the operating conditions themselves do not vary widely, and they can be adequately simulated during testing, then the resulting measurements are generally assumed to be a linearized estimation of the device's operation.

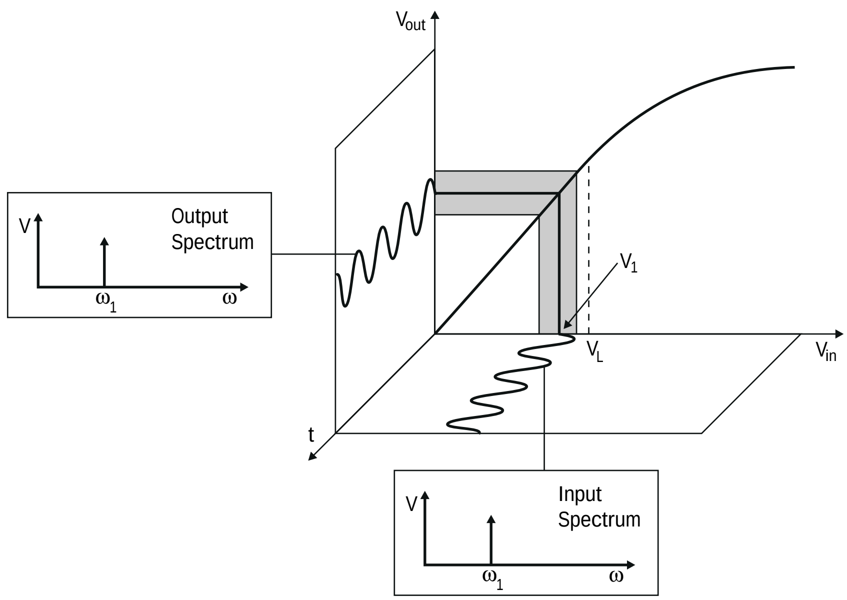
To provide maximum flexibility in obtaining a linearized estimation of a device's operation, advanced DSAs provide two separate analysis functions for measuring the frequency response of both linear and nonlinear devices: Swept Fourier Analysis (SFA) and Fast Fourier Transform (FFT) analysis. More information concerning SFA and FFT analysis as well as many of the other measurement capabilities provided by DSAs are presented in Part 2 of this application note.

Figure 1-21:

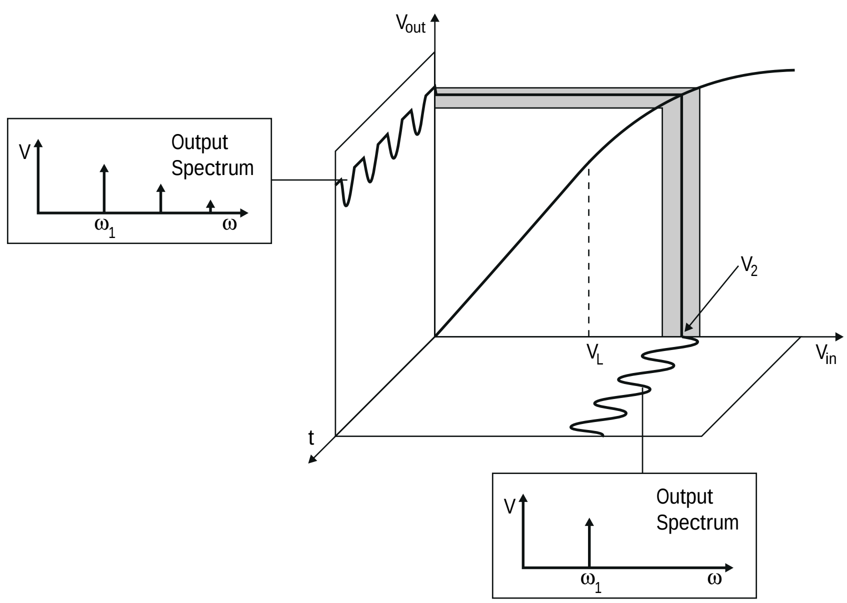
A. Typical gain curve with additional third axis representing time.



B. Operation within linear region.



C. Operation in non-linear region.



Part 2:
Measurement and Analysis Tools Applied to the Development Process

Historically, a test instrument's primary contribution to the development of a control system has been the collection of stimulus and response data. While this is still true, microprocessor-based Dynamic Signal Analyzers (DSAs) have expanded the role of the test instrument to include significant contributions in other areas of control system development, such as modeling and design.

The purpose of the following chapters is to provide a basic introduction to the measurement and analysis capabilities provided by high performance DSAs, and to suggest how these tools can be used in the various phases of control system development.

Chapter 1: Modeling the Development Process

In general, it is recognized that the development of a control system typically involves some unique combination of five distinct processes: model, design, build, test and analyze. For the purpose of this application note, these five processes are defined as follows:

Design: determining the combination of physical or theoretical components or parameters that will produce a desired action or result.

Model: the process of transforming the observed characteristics of some device or process into theoretical representations consistent with the analysis/design technique being used.

Build: the physical construction of a system and/or its components.

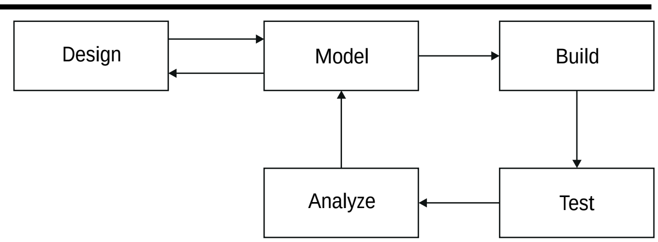
Test: the collection of stimulus and/or response data.

Analyze: determining the value of parameters, either physical or theoretical, used to characterize the action or function of a device. Also establishing the relationships, if any, between those parameters.

When grouped into a process flowchart, these five processes can be used to model the development of a control system. A generalized example of a "development process" model/ flowchart is shown in Figure 2-1.

To emphasize the DSA's ability to contribute throughout the development of a control system, the following chapters examine the tasks associated with each development process (with the exception of build) and present the tools provided by DSAs for accomplishing those tasks. To provide a structured introduction, the chapters are presented in the following order: Test, Analyze, Model and Design.

Figure 2-1:



Chapter 2: Test

Test: the collection of stimulus and/or response data.

There are many tests which conform with the above definition; however, the most common control system tests are the measurement of a system's response to a step change in the input (i.e., the step response) and the frequency response of the system and/or any of its components.

Instruments which have commonly been used to perform these tests include frequency response analyzers, network analyzers, waveform recorders, strip-chart recorders, and storage oscilloscopes. Typical control system tests often required at least two of these instruments: one instrument to record time domain data (e.g., the impulse response or step response) and another to record frequency domain data (e.g., the open-loop or closed-loop frequency response).

The high performance DSA, however, is a single instrument capable of providing all the measurement capability needed in the dc to 100 kHz frequency range. Technological advances allow DSA to assume 1 to 3 basic configurations: a waveform recorder for direct measurement of time domain data, a frequency response analyzer (i.e. Swept Fourier Analyzer) for providing frequency domain data, or a Fast Fourier Transform (FFT analyzer which also provides frequency domain information.

In addition to providing three analyzers within one test instrument, the DSA also provides several signal monitoring functions. These functions allow the DSA to automatically optimize measurement conditions during a test, reducing the need for operator interaction.

The remainder of this chapter presents the DSA's basic capabilities for measuring both time domain and frequency domain data.

2-1: Time Domain Measurements

Time domain measurements require the test instrument to record the reaction of a device in response to some controlled change in the system's input. A measurement is generally considered successful if it records the entire response and allows the operator to examine both the long term trend of the response and the details of any short term events.

DSAs provide this measurement capability by sampling the signals applied to their inputs and recording the samples as blocks of contiguous data called time records. How the time records are stored and how the data within them can be accessed depends on which of two measurement modes, time capture or time throughput, is used to collect the data.

2-1.1: Time Capture

Responses which decay to a steady state value within a few time records can easily be recorded using the DSA's time capture mode. The time capture mode stores a limited number of contiguous time records within the DSA's internal memory. Once collected, all the data can be compressed onto a single trace on the DSA's display. Segments of the compressed data can then be expanded and closely examined on the second trace of the display, as shown in Figure 2-2.

2-1.2: Time Throughput

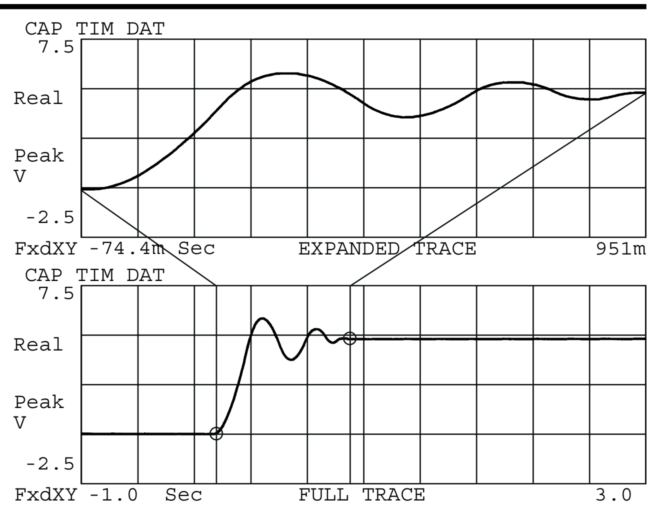
Occasionally, a device with a very long settling time will require very large amounts of data to be recorded. In these situations, the DSA's time throughput mode can be used to store contiguous time records[1] directly to a mass storage disc without the need for an instrument controller. To study a recorded event, time records are recalled from the disc and presented on the DSA's display.

To ensure that an entire response can be recorded, both time capture and time throughput provide pre- and post-trigger data recording functions. The pre-trigger function allows a specified amount of data obtained before a trigger occurs to be recorded. The post-trigger function allows a specified amount of data to be ignored when obtained after a trigger occurs.

For systems with very fast response times, the pre-trigger function can be used to record the steady-state operation of a system just before a step change is introduced. Alternatively, the post-trigger function can be used to ignore the large amounts of dead time in systems with very slow response times.

In addition to recording and displaying time domain data, DSAs are also capable of recalling recorded data and processing it through a Fast Fourier Transform algorithm. This allows the DSA to provide both time and frequency domain information from one set of recorded data. This capability can be especially valuable for extracting the maximum amount of information from tests which can be performed only once, such as destructive tests.

Figure 2-2: A display of a step response recorded with time capture showing 10 time records compressed on a single trace (lower trace) and a portion of the response expanded to reveal detail (upper trace).



[1] If the DSA collects data much faster that the connected disc can record data, or the DSA collects data faster than it can process the data through its own I/O section, then the time records will no be contiguous. The rate at which time records can be transferred in a contiguous fashion is referred to as the "real-time bandwidth" of the throughput function. More information on real-time bandwidths is available in Hewlett-Packard Application Note 243, *The Fundamentals of Signal Analysis.*

2-2: Frequency Domain Measurements

Virtually all closed-loop control system development requires the frequency response of the system and/or some of its components to be evaluated by experiment. Unlike most conventional test instruments, advanced DSAs provide two independent techniques for measuring the frequency response of a device; Swept Fourier Analysis and Fast Fourier Transform analysis.

2-2.1: Swept Fourier Analysis

Swept Fourier Analysis (SFA) is a very common measurement technique involving a swept sine wave source and an integration process which emulates a tracking bandpass filter, as shown in Figure 2-3. The primary objective of this measurement technique is to measure the gain and phase shift of a device by measuring only the fundamental component of the stimulus signal and only the fundamental component of the device's response signal (the frequencies of the fundamentals are assumed to be the same). A series of measurements are made at different frequencies to provide a frequency response based on the fundamental of the stimulus and response signals (i.e. ignoring any other spectral components including those generated by nonlinearities such as harmonic distortion).

By using very narrow bandwidths, the effects of nonlinearities such as harmonic distortion, dc offset and random noise can be minimized. This measurement technique also allows those types of nonlinearities which are not affected by narrow filter bandwidths (such as level saturation and frequency shifting of resonances) to be characterized by either making several measurements at different stimulus levels or by sweeping in both directions.

To achieve the narrow filter bandwidths required to measure low frequency systems, DSAs utilize a Discrete Fourier Transform to evaluate the energy within a narrow frequency span. The transform is evaluated at several points during a sweep with the center frequency of the analysis corresponding to the frequency of the swept sine source (thus the term Swept Fourier Analysis). This technique emulates a tracking bandpass filter with very narrow bandwidths, very good harmonic rejection and excellent dc rejection.

An added advantage of using a DSA to make SFA measurements is the availability of automated measurement functions. By constantly monitoring the signals applied to its inputs and referencing past measurements, the DSA can automatically:

- adjust its input sensitivity
- reject measurements in which input overloads occurred
- adjust the frequency resolution of the measurement relative to the rate of change in gain and phase
- repeat a measurement at a given frequency and average the results until an acceptable variance in the measurement is obtained
- adjust the source level to maintain a constant stimulus or response level
- allow the operator to simultaneously monitor the signals applied to the analyzer (in either the time or frequency domains) and view the current measurement.

2-2.2: Fast Fourier Transform Analysis

Compared to SFA, FFT analysis represents more of a parallel approach to measuring a device's frequency response. Rather than sweeping a single bandpass filter as the SFA technique does, the FFT process uses a different form of Fourier integration to create many adjacent bandpass filters (up to 800 in advanced DSAs), as shown in Figure 2-4. These filters selectively and simultaneously measure the energy distributed over an entire frequency span.

A useful analogy is to think of each filter as the bandpass filter of an SFA analyzer. However, rather than collect new data for each measurement point sweep, the FFT process uses time records to collect time domain data and then processes the data through 800 filters simultaneously. This form of parallel processing provides exceptional measurement speeds. It is worth noting, however, that unlike an SFA measurement, an FFT measurement does not filter out energy converted to other frequencies by nonlinearities in the system. Instead, these frequency components (if they are not coherent with the stimulus) are removed by averaging several measurements.

One of the most powerful attributes of the FFT measurement technique is that it allows virtually any type of signal to be used as a stimulus. Common stimulus signals used with FFT measurements include: actual operating signals, sine wave chirps, fixed sine waves, random noise, burst random noise, step functions and impulse functions.

Figure 2-3: Synchronized sweep of sinewave source and bandpass filter created by Fourier integration.

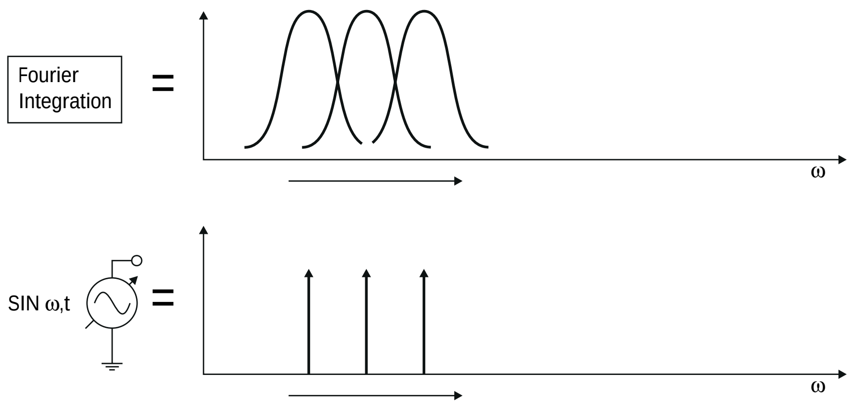
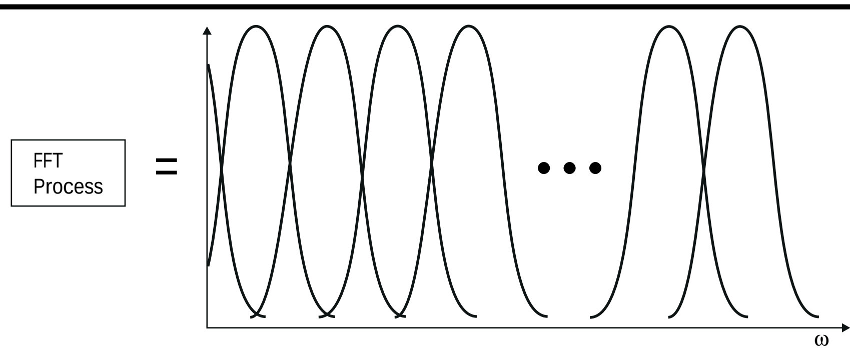


Figure 2-4: Multiple bandpass filters produced by fast Fourier transform (FFT) process.



This broad range of stimulus signals increases the resources available for characterizing the operation of a system. Often, selecting the right stimulus signal can provide a better understanding of nonlinearities present in the system and, in some cases, even reduce the overall testing time. The following paragraphs cite some of the benefits offered by certain source types.

An important class of stimulus signals are those stimuli which produce energy at all of the frequencies being analyzed by the FFT algorithm and do so within one time record. Stimuli which meet this criteria (such as the sine chirp, random noise, burst chirp and burst random noise stimuli provided by advanced DSAs) allow the FFT algorithm to provide frequency response information over the entire frequency span being analyzed with just one measurement. If any of these stimuli (with the exception of random noise) are used to test a system which is relatively noise-free and linear, a single time record is often sufficient data to produce an accurate frequency response.

When testing a nonlinear system, selecting a stimulus signal which approximates the signals present during normal operation can provide results which more accurately predict the system's operation. The ability to use a random noise stimulus can be very useful in this respect. For example, random noise superimposed on a dc level often resembles the signals present in a servo system much more than a sine wave superimposed on a dc level.

Signals with random amplitude distribution, such as true random and burst random, can be used to provide an approximation of the frequency response of a system with amplitude nonlinearities. Because random noise is characterized by a random level distribution at a given frequency, a random noise measurement produces a frequency response which represents an average of responses taken at several stimulus levels. When attempting to measure the frequency response of a device with an amplitude nonlinearity such as gain compression, a random noise measurement may provide a better approximation of the device's actual operation than a single swept sine measurement.

A random stimulus signal can also reduce the effects of nonlinearities influenced by the direction of a sine sweep. Such nonlinearities often show up as a change in resonance frequencies corresponding to a change in sweep direction (not to be confused with skewed responses caused by excessive sweep speed). Since random noise continuously produces energy over an entire frequency spectrum, the measurement is not affected by transferring energy from one frequency to another.

Some forms of nonlinearities preclude the use of certain stimulus types. For example, when testing systems with a significant amount of dead zone or hysteresis, such as large gear trains, signals such as random noise can be inappropriate. The waveform of a random signal is typically characterized by many changes in slope and a greater concentration of lower level voltages than high level voltages. This would create a lot of noise in a gear train while producing little output. Instead, a sine wave stimulus which spends more time at higher voltage levels and makes fewer slope transitions may be a much better overall stimulus choice.

The decision of which stimulus/analysis combination should be used is driven in part by the known attributes of the device being tested and the kind of information being sought. For example, several swept sine measurements made at different stimulus levels can be used to characterize the operation of a device with an amplitude nonlinearity. Alternatively, an FFT measurement using random noise and averaging can be used to provide a single frequency response which approximates the device's operation over a range of stimulus levels.

Chapter 3: Analyze

If the device being tested is essentially linear (at least within the range of amplitudes and frequencies being tested), the selection of a stimulus/analysis combination is simply a matter of measurement speed. Any stimulus/analysis combination would be able to produce accurate results.

It is important to note, however, that before any assumption can be made about a system's linearity, at least two measurements (with variances in the stimuli between them) must be compared. If the system is found to be nonlinear, it may take several more measurements to characterize the nonlinearity so that its effect on the operation of the system can be understood.

It is in response to these measurement needs that advanced DSAs have incorporated the ability to make time domain measurements, traditional swept sine frequency response measurements and non-traditional frequency response measurements utilizing virtually any type of stimulus signal and FFT analysis. With these measurement capabilities, the DSA provides a total measurement solution for fully characterizing the operation of control systems.

Analyze: determining the value of parameters, either physical or theoretical, used to characterize the action or function of a device. Also, establishing the relationships, if any, between those parameters.

This definition of analysis, when applied to classical control theory, generally implies the evaluation of parameters such as gain margin, phase margin and settling time.

Typically, these parameters are not evaluated by the test instrument. More often than not, they must be derived from the measured data and, in some cases, derived from several sets of data. With respect to extracting useful information from measured data, the Dynamic Signal Analyzer represents one of the most powerful measurement and analysis tools available to the control systems engineer.

The DSA's major contributions toward analyzing data center around three major functions: waveform math, curve fitting and coherence. The following sections briefly describe each function and present typical applications.

3-1: Waveform Math

Waveform math provides the ability to use standard math operators such as +, -, × and ÷ between two displayed data traces, or perform any of the other math functions shown in Table 2-1 on individual traces. Waveform math therefore allows many of the control system calculations which have historically been done graphically, with plotted data, to be performed within the analyzer. This not only reduces calculation times, but also preserves the full resolution and accuracy of the original data. The following examples present only a few of the many possible applications for the waveform math function.

A very straightforward application of waveform math is the extraction of the normalized value of maximum overshoot from a step response measurement. The left half of Figure 2-5 shows a measured step response with a Y-axis marker positioned on the steady-state value. Using waveform math, the display can be normalized by simply specifying the ÷ operator and entering the response's steady-state value. The normalized value of maximum overshoot can then be read directly from the X-axis marker as shown in the right-half of Figure 2-5.

Table 2-1: Waveform math functions available with advanced dynamic signal analyzers.

+	√	Multiply by jω	T/(1 - T)
–	Reciprocal	Multiply by jω - 1	Real Part
×	Negate	FFT	Complex Conjugate
÷	Differentiate	FFT -1	Log of Data

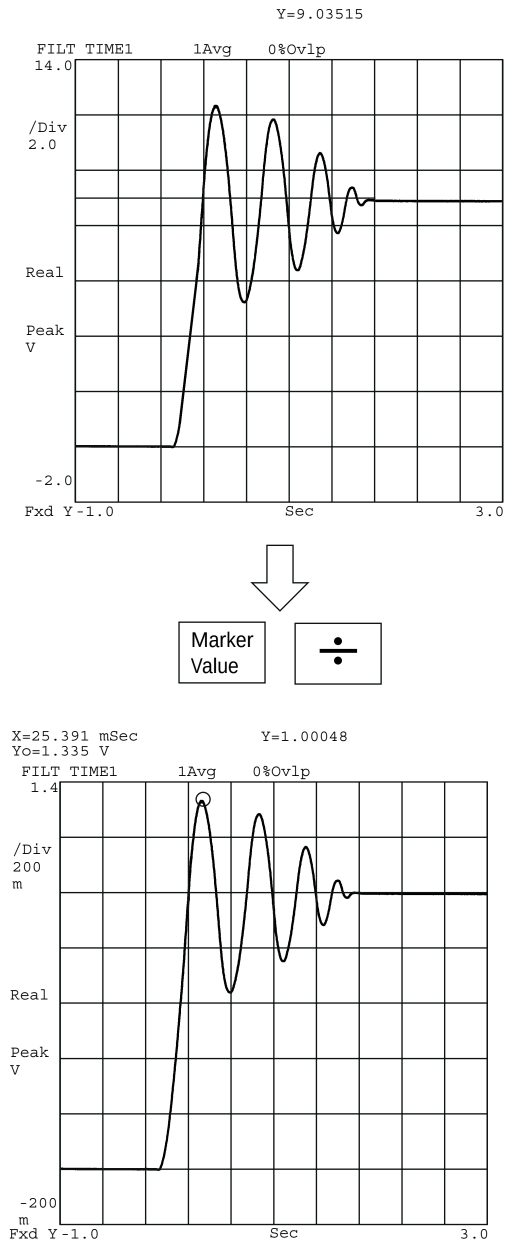
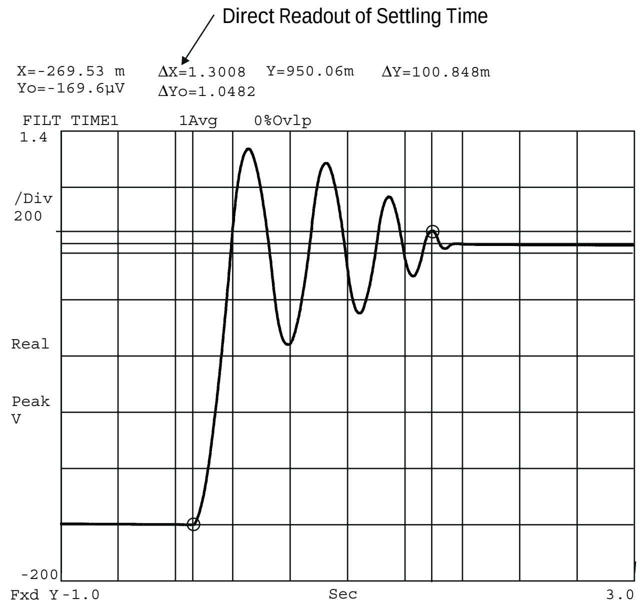


Figure 2-5: Using waveform math to normalize a step response. Marker on normalized display reads out normalized peak overshoot directly.

Using the normalized display, the settling time can also be quickly evaluated. The upper and lower boundaries relative to the steady-state value can be clearly marked by simply programming the Y-axis markers to those values (i.e., for a restriction of ± 5% of final value, the markers can be set to 1.05 and 0.95). The X-axis marker can then be used to display the settling time, as shown in Figure 2-6. The information shown on the display of the DSA, including trace, display grid and annotation, can then be sent directly to a digital plotter to provide hardcopy documentation.

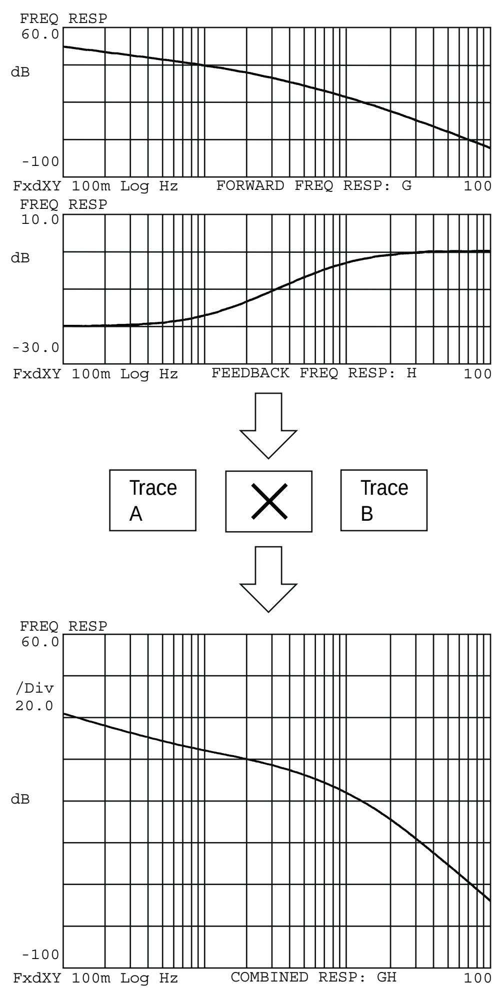
Figure 2-6: Using X- and Y-axis markers to read out settling time of a step response from a normalized display.



The DSA's waveform math function can also be used with frequency domain data to execute much more complex calculations. For example, two sets of frequency response data representing the forward gain path and feedback path of a system could be quickly combined to predict the system's open-loop frequency response.

Combining frequency responses can be accomplished by simply displaying one set of frequency response data in one display trace and a second set of frequency response data in the other display trace. The operator then selects an active trace, the multiply operator and the second operand (in this case the nonactive display trace)[1]. The result of the calculation is then displayed in the active trace, as shown in Figure 2-7.

Figure 2-7: Using waveform math to combine frequency response data.

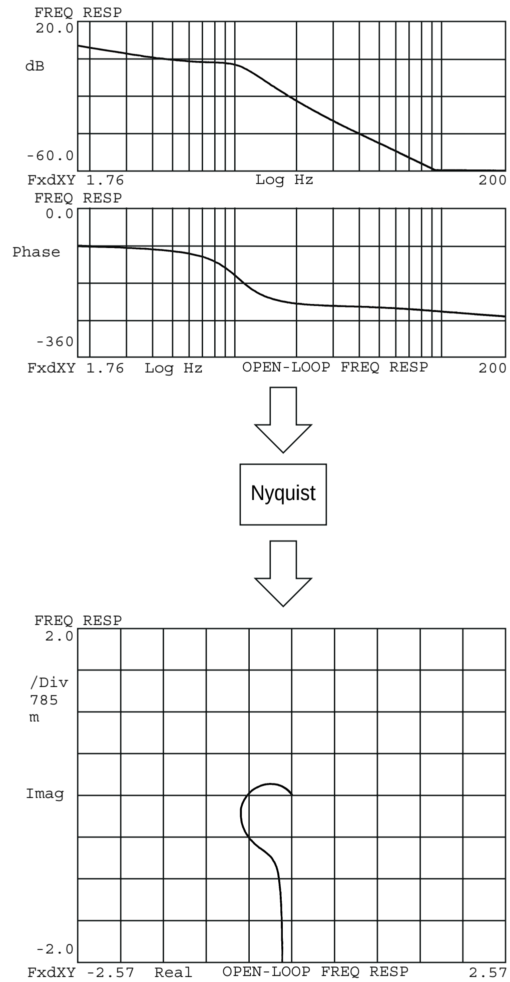


[1] The order in which the waveform math operations are executed may differ between DSAs.

The resultant frequency response can then be presented in virtually any desired scale and in one of many display formats. For example, the derived frequency response can be displayed in a Bode plot, as shown in Figure 2-8A, to allow the gain margin, phase margin and open-loop bandwidth to be quickly read from the X-axis markers. The frequency response can then be displayed on a Nyquist plot, as shown in Figure 2-8B, to provide a quick check of the system's absolute stability.

Since either display trace may contain either current measurement data, calculated data, or data recalled from a mass storage device (such as a magnetic disc or tape drive), waveform math can be used to combine many frequency response data sets. This capability could be used to predict the frequency response of a system from a library of previously stored component frequency response data.

Figure 2-8: CRT Display of data allows quick change of display formats.



A. Bode plot

B. Nyquist plot

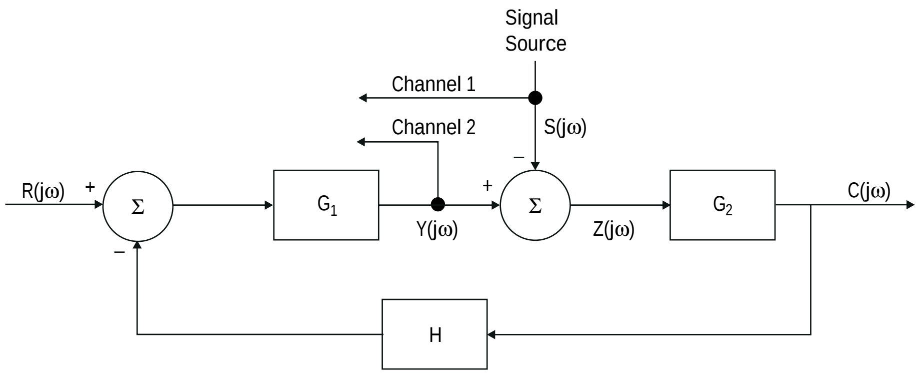


Figure 2-9: Typical measurement setup for determining the open-loop frequency response of a closed-loop system using FFT.

$$\frac{Y(j\omega)}{S(j\omega)} = \frac{G_1G_2H(j\omega)}{1 + G_1G_2H(j\omega)} = T(j\omega) \qquad \frac{T(j\omega)}{1 - T(j\omega)} = G_1G_2H(j\omega)$$

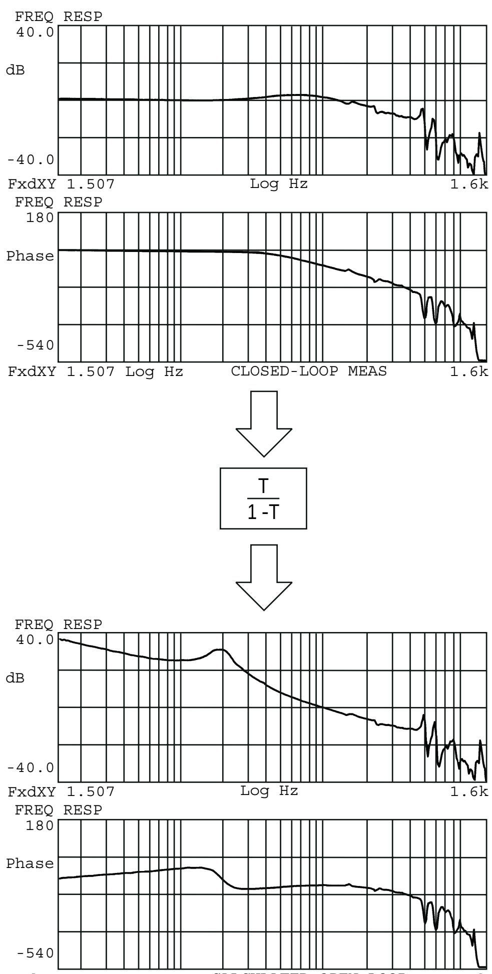


Figure 2-10: Using the "*T/(1 - T)*" waveform math function to calculate the open-loop frequency response from a measurement of *Y(jω)/S(jω)*.

Waveform math also makes it possible to easily calculate the open-loop frequency response of a system from a closed-loop measurement. Typically, a stimulus signal is injected into the loop and, when using FFT analysis[1], the frequency response between the stimulus signal S and the response to the stimulus signal at the point Y is measured as shown in Figure 2-9. The open-loop frequency response of the system can then be calculated by evaluating the equation:

$$\textit{open-loop frequency response} = \frac{T(j\omega)}{1 - T(j\omega)}$$

where $T(j\omega)$ is the measured frequency response $Y(j\omega)/S(j\omega)$.

This equation can be easily evaluated using either a series of waveform math calculations or by using the single waveform math operator $T/(1 - T)$ as shown in Figure 2-10.

3-2: Curve Fitting

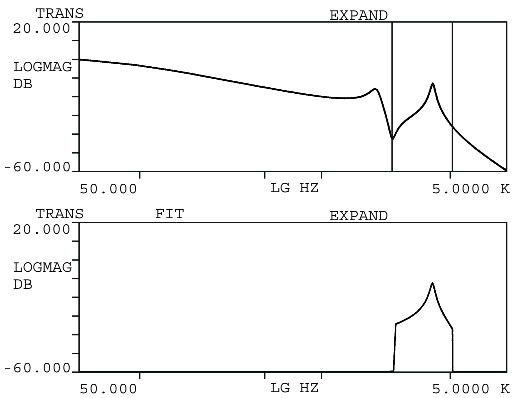
Curve fitting is a function which estimates an equation whose solution, when plotted, will be identical to the measured frequency response. Depending on the curve fitter available with a given DSA, the derived equation may be expressed to the operator in one of three formats: a table of poles and zeros, a table of poles and residues (i.e., partial fraction expansion form), or a ratio of polynomials.

Advanced DSAs are usually equipped with one of two curve fitters, either a basic single-degree-of-freedom (SDOF) curve fitter or a multiple-degree-of-freedom (MDOF) curve fitter. SDOF curve fitters provide pole/residue information for each resonance identified by the operator, as shown in Figure 2-11A. MDOF curve fitters represent a more versatile generation of curve fitters which can automatically process an entire spectrum; using up to 40 poles and 40 zeros in the estimation process (see Figure 2-11B). The latter curve fitters are typically accompanied by a synthesis capability which allows the pole/zero information to be quickly converted to a pole/residue format or a polynomial format as shown in Figure 2-12.

For extracting information from measured data, the curve fitting function is an exceptionally powerful analysis tool. Its applications, however, lie mostly in the area of modeling and design and are discussed in chapters 4 and 5, in Part Two, respectively.

[1] When using an FFT analyzer to derive the open-loop frequency response of a closed-loop system, the ratio $Y(j\omega)/S(j\omega)$ or $Z(j\omega)/S(j\omega)$ is measured rather than $Y(j\omega)/Z(j\omega)$ (the ratio commonly measured with frequency response analyzers) to prevent a bias error from degrading the calculation. The bias error can be avoided and is typically not a significant factor when using SFA analysis.

Figure 2-11: Using curve fitters to reduce frequency response data to analytical equations.



Data Being Analyzed

Curve Produced by Curve Fitter Estimation

FREQUENCY AND DAMPING

MODE NO.	FREQUENCY HZ	R/S	DAMPING %	HZ	R/S
1	2.047 K	12.863 K	5.834	119.638	751.708
2	4.002 K	25.143 K	2.598	103.990	653.386
					0

LARGEST MODE USED: 2

And

MODAL RESIDUES

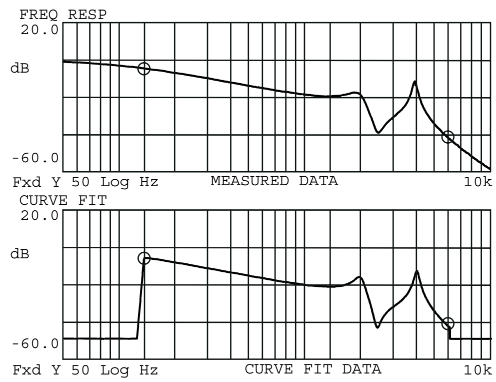
MODE: 2 | FREQ(HZ): 4.002 K | DAMP(%): 2.598

MEASMT	CHAN#1 PT	CHAN#1 DIR	CHAN #2 PT	CHAN #2 DIR	RESIDUE
1	11	3	1	3	321.651
2	0	0	0	0	0.000
3	0	0	0	0	0.000
4	0	0	0	0	0.000
5	0	0	0	0	0.000
6	0	0	0	0	0.000
7	0	0	0	0	0.000
8	0	0	0	0	0.000
9	0	0	0	0	0.000
10	0	0	0	0	0.000

LARGEST MEAS USED: 1

$$\frac{R_1}{s + P_1} + \frac{R_1^*}{s + P_1^*} + \frac{R_2}{s + P_2} + \ldots = \text{Transfer Function}$$

A. Using an SDOF curve fitter to evaluate the frequency, damping and residue of each resonance in the displayed frequency response. Curve fit data is accumulated in the Frequency and Damping and Modal Residues tables.



Measured Data

Curve Produced by Curve Fitter Estimation

Curve Fit

Poles And Zeros

POLES 5	ZEROS 2
1 -74.6323	-79.9162 ±j 2.43186k
2 -119.071 ±j 1.99714k	
3 -103.437 ±j 4.00214k	

Time delay= 0.0 s Gain= 1.02E+9 Scale= 1.0

$$\frac{(s + Z_1)\,(s + Z_2)\, s + Z_3) \ldots}{(s + P_1)\,(s + P_2)\, s + P_3) \ldots} = \text{Transfer Function}$$

B. Using an MDOF curve fitter to evaluate the poles and zeros associated with the displayed frequency response, up to 40 poles can be evaluated in one analysis.

Figure 2-12: Converting pole/zero data from the curve fitter to pole/residue and polynomial formats via the frequency response synthesis table conversion function.

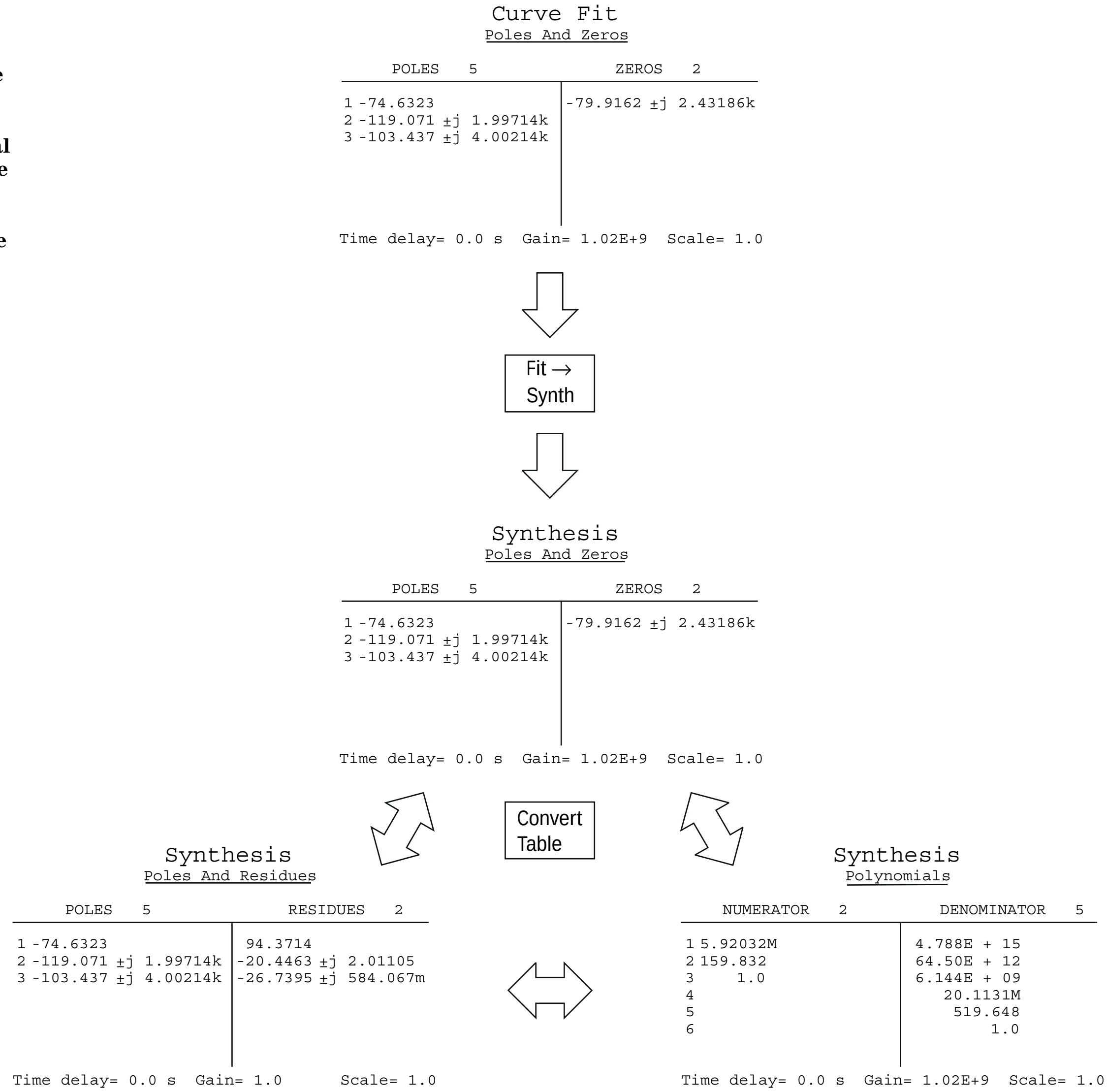
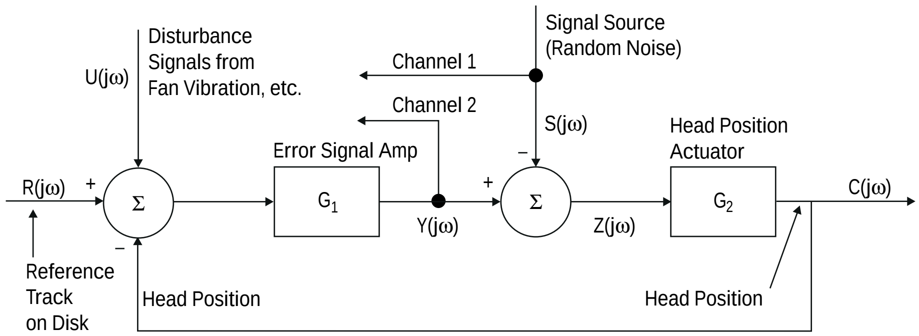
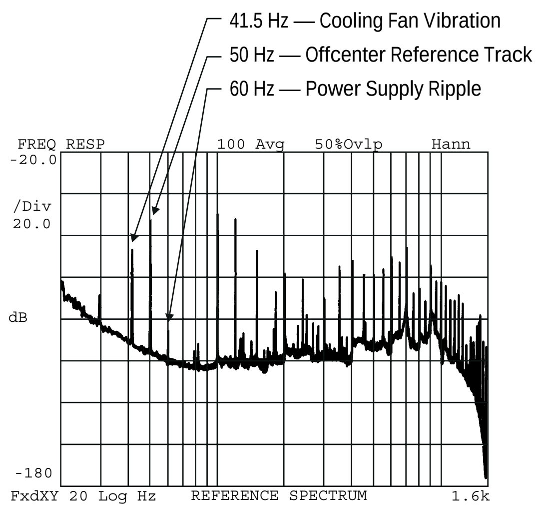


Figure 2-13: Using the noncoherent power spectrum of the response to monitor known periodic disturbances in the control loop.



A. Measurement setup used to measure the frequency response *Y(jω)/S(jω)* of a disk drive read/write head positioning servo.



B. Noncoherent power spectrum derived from *Y(jω)/S(jω)* measurement and the associated coherence data. Note that the error signal caused by an offcenter reference track appears as a noncoherent signal at 50 Hz.

3-3: Coherence

The coherence function is a statistical quantity whose dimensionless values represent the fraction of system output power directly related to the input. Values of coherence are used in two primary applications: 1) as a measure of the quality of a frequency response measurement and 2) to discriminate between those response signals which are directly related to (coherent with) the stimulus signal and those response signals which are not directly related to (not coherent with) the stimulus signal.

When more than one average is taken per measurement point, the coherence function produces a value from 0.0 to 1.0 for each point. (For example, when using SFA, a value of coherence will be produced for each step in the sweep if the analyzer is programmed to average two or more measurements per step.) A coherence value of 1 indicates that all of the output power (response) is coherent with the input power (stimulus) but not necessarily a result of the input power. A coherence value of 0 indicates that virtually none of the output power is coherent with the input power.

Since a low value of coherence indicates that only a small percentage of the response is directly related to the stimulus, it is reasonable to assume that the corresponding measurement data may not accurately reflect the transfer of energy through the tested device. In this respect, the coherence function acts as a qualitative tool which can be used to verify the general quality or credibility of a measurement. Typical causes of low coherence include very poor signal-to-noise ratios, the presence of noncoherent signals generated within the tested device or, when using FFT analysis, leakage due to improper window selection or insufficient time record length[1].

Coherence can also be used to separate the output power spectrum into two power spectra: the coherent power spectrum which represents the output power directly related to the input and the noncoherent power spectrum which represents the output power not related to the input.

Both the coherent and noncoherent power spectra have been used in several interesting applications. One example is the use of the noncoherent power spectrum by a disc drive manufacturer to monitor the disturbance signals within the read/write head positioning servo. By using a random stimulus signal, the periodic signals within the control loop (such as those caused by cooling fan vibration, power supply ripple bleeding into the control loop or an off-centered reference track on the disc) appear in the response as noncoherent signals. By correlating the known characteristic frequencies of these signals with the spectral components of the noncoherent power spectrum, the amplitudes of these noncoherent signals were effectively monitored, providing more information about the overall health of the positioning system. A simplified drawing of the measurement setup and an actual plot of the noncoherent power spectrum are shown in Figure 2-13.

[1] Complete definitions of leakage, window functions and time records are available in Hewlett-Packard Application Note 243, *The Fundamentals of Signal Analysis*.

Chapter 4: Model

Since all of the data needed to calculate the noncoherent power spectrum is provided with each frequency response measurement, it can be provided without increasing measurement time. The ability to increase the information obtained from each measurement can be especially valuable in situations where testing time is considered a valuable commodity, such as production line testing. A copy of a production test report dumped directly to a digital plotter by a DSA is shown in Figure 2-14.

The coherent and noncoherent power spectra mentioned above can easily be obtained by using waveform math to calculate the following formulas:

coherent power spectrum =
(output power spectrum) × (coherence spectrum)
noncoherent power spectrum =
(output power spectrum) × (1 – coherence spectrum)

where: *coherence spectrum* refers to the collective set of coherence values which exist when more than one average is taken and *(1 – coherence spectrum)* implies the subtraction of each value of coherence in the coherence spectrum from 1.

The output power spectrum, like the coherence function, is a normal by-product of a DSA's frequency response calculations and can be viewed at any time.

More applications for the coherence function (as well as a detailed definition) are provided in Hewlett-Packard Application Note 245-2, *Measuring the Coherence Function with the HP 3582A Spectrum Analyzer.*

Model: the process of transforming the observed characteristics of some device or process into theoretical representations consistent with the analysis/design technique being used.

This definition, when applied to classical control theory, generally implies the creation of equations which accurately predict the action or function of some device in the frequency or time domains. Since most design work is done in the frequency domain, the modeling process can further be generalized as the development of frequency domain equations, typically in a pole/zero format, which accurately predict a device's frequency response.

4-1: Curve Fitting Applied to the Modelling Process

As an aid in accomplishing this task, the MDOF curve fitter offered with high performance DSAs represents one of the most powerful tools ever offered by a test instrument.

By simply displaying a measured frequency response and activating the MDOF curve fitter, the DSA automatically provides an estimate of the s-plane poles and zeros and the gain required to produce the displayed response, as shown in Figure 2-15.

The use of a curve fitter to extract pole/zero information from a measured frequency response represents a significant advancement over the graphic techniques commonly used to derive pole/zero information. The curve fitter has the advantage of utilizing the full frequency and amplitude resolution of the measured data and, in many cases, provides the pole/zero information in the time normally required to obtain and prepare hardcopy plots for graphic interpretation.

Figure 2-14: A production test report showing the data provided from a measurement made on an operating closed-loop servo system.

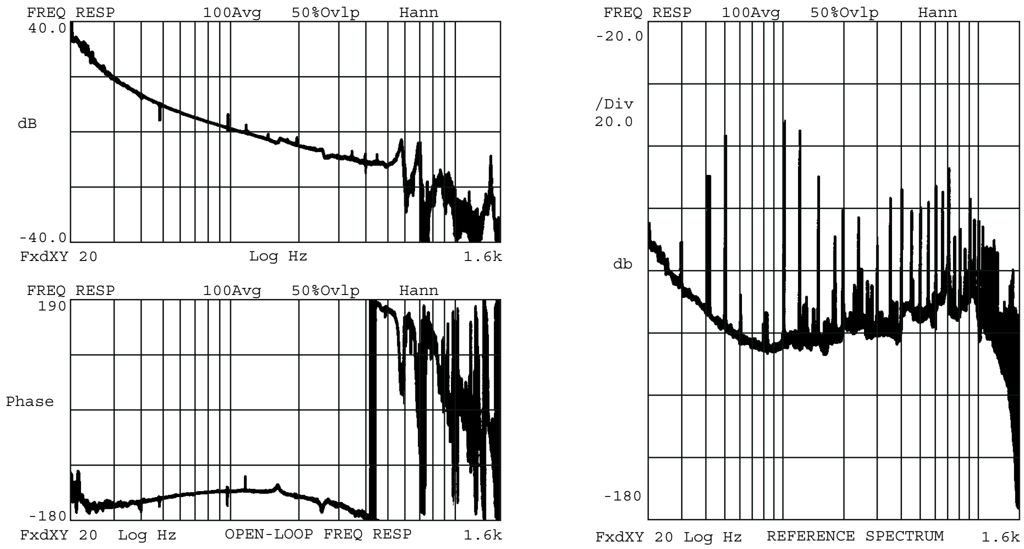
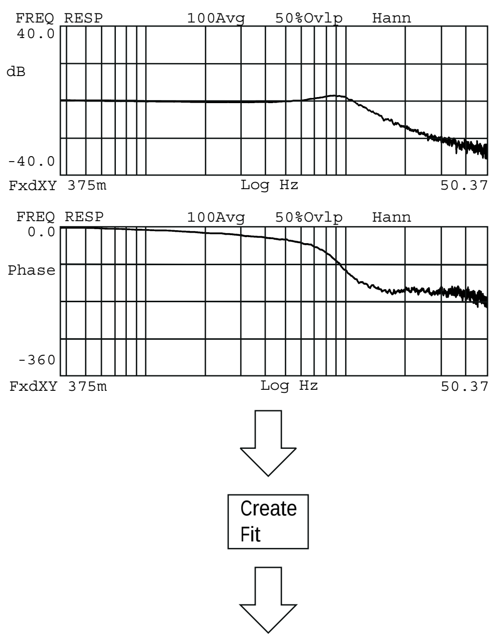


Figure 2-15: Using an MDOF curve fitter to estimate the pole/zero locations and gain from frequency response data.



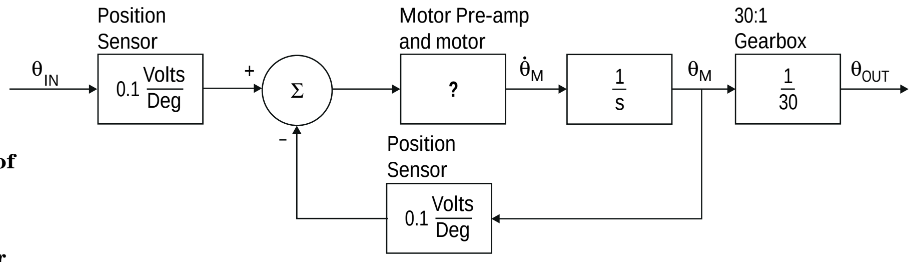
Curve Fit

Poles And Zeros

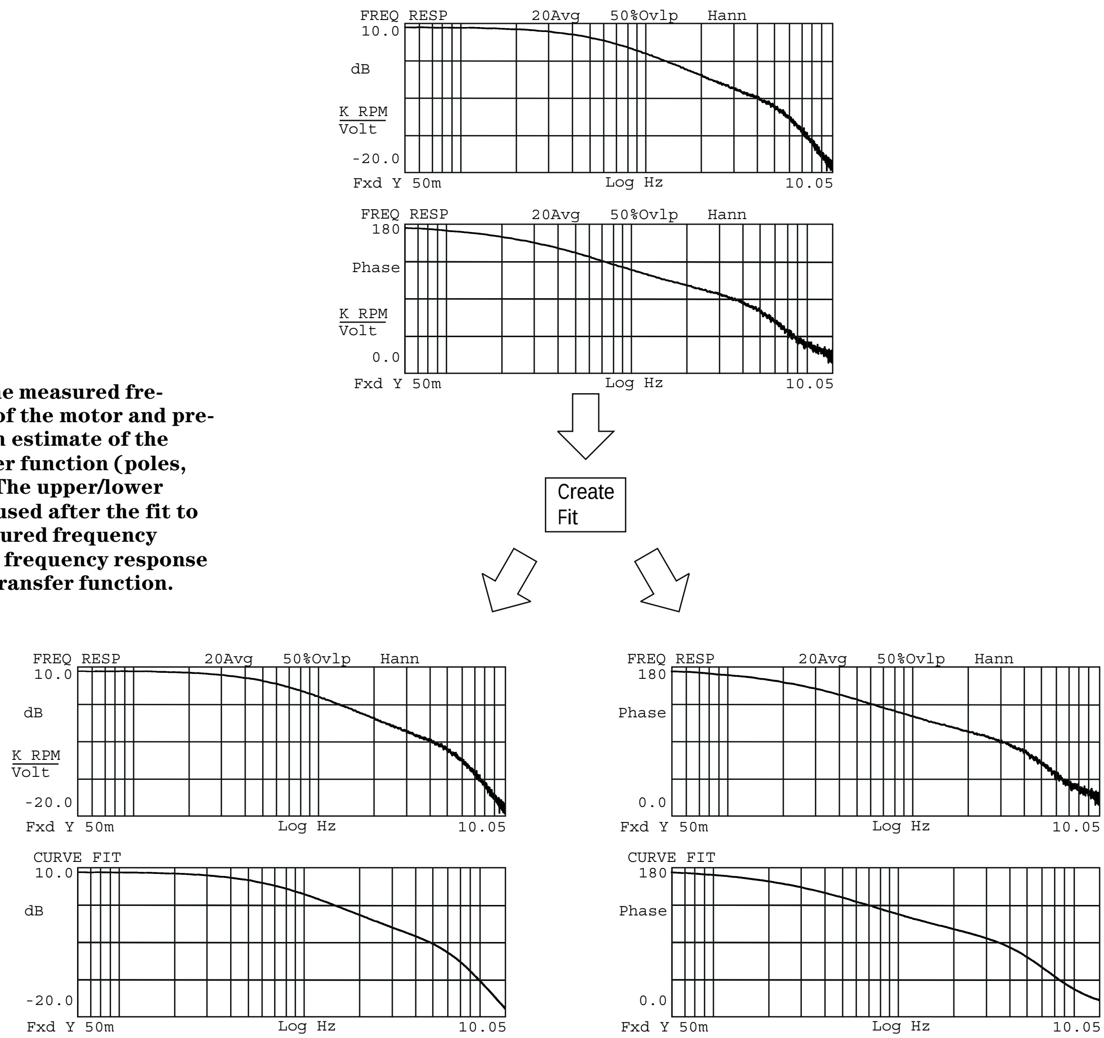
	POLES 5	ZEROS 2
1	-98.6842	-30.8428
2	-8.03376 ±j 6.78611	-5.36744 ±j 10.7513
3	-1.98173 ±j 9.71016	95.0489 ±j 97.0916

Time delay= 0.0 s Gain= 11.79m Scale= 1.0

Figure 2-16:



A. Block diagram of position control system with unknown transfer function for motor and pre-amp.



B. Curve fitting the measured frequency response of the motor and pre-amp to produce an estimate of the associated transfer function (poles, zeros and gain). The upper/lower display format is used after the fit to compare the measured frequency response with the frequency response of the estimated transfer function.

Figure 2-16: (cont.)

Curve Fit

Poles And Zeros

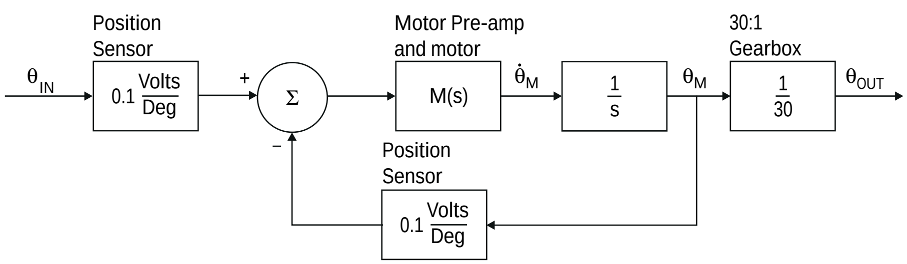
POLES 3	ZEROS 1
1 -599.83m	-4.56721
2 -3.27977 ±j 4.19772	

Time delay= 0.0 s Gain= 10.83 Scale= 1.0

C. Producing the transfer function of the motor and the pre-amp from pole/zero table generated by DSA's MDOF curve fitter.

$$\frac{-10.83\,(s + 4.57)}{(s + 0.6)(s + 3.28 + j4.2)(s + 3.28 - j4.2)} = M(s)$$

D. Completed block diagram.



Amongst other applications, the pole/zero data obtained from frequency response measurements can be used to either verify the poles and zeros used in an existing analytical model or create an initial model of a device with unknown characteristics. An example of the latter application is illustrated in Figure 2-16. In this example, a transfer function is generated for a combination armature controlled motor and preamplifier (of a position control system) whose specifications, such as motor inertia and forward gain, are unknown.

To obtain the motor/preamps transfer function, the frequency response of the motor/preamp is first measured using a DSA equipped with a MDOF curve fitter. The curve fitter is then activated resulting in a table of poles and zeros. The pole/zero information is automatically synthesized to provide a frequency response which can be compared with the measured frequency response, as shown in Figure 2-16B. The pole/zero data is then used to generate a transfer function of the motor/preamp as illustrated in Figure 2-16C. The derived transfer function can now be added to the system block diagram to complete the system model, as shown in Figure 2-16D.

4-2: Frequency Response Synthesis Applied to the Modeling Process

Another useful modeling tool provided with advanced DSAs is the frequency response synthesis function (commonly referred to as the synthesis function). DSAs equipped with this function allow analytical equations (e.g., transfer functions) to be entered directly into the analyzer. The DSA then calculates and displays the frequency response associated with the transfer function, as shown in Figure 2-17.

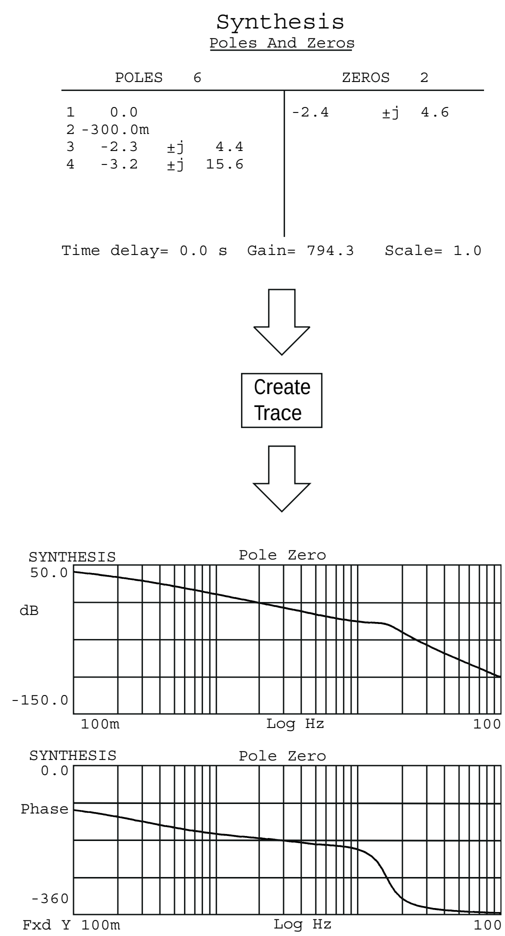
Equations may be entered in one of three formats: pole/zero, pole/residue (i.e., partial fraction expansion), or ratio of polynomials in s. In addition to providing a conversion function for transferring data from one format to another, high performance DSAs also provide direct transfer of data between the synthesis and curve fitting functions.

In the modeling process, the synthesis function is commonly used in conjunction with the curve fitter. For example, if the curve fitter produces more detailed information than required for a given application, the pole/zero data can be transferred to the synthesis function where insignificant poles and zeros can be deleted. The frequency response of the remaining poles and zeros can then be synthesized and compared to the measured frequency response. This allows the engineer to verify that the remaining poles and zeros sufficiently model the measured frequency response.

Another use of the synthesis function utilizes modeling information to optimize the initial testing of systems. By synthesizing the frequency response of a system which has never been tested (i.e. the model has been developed from data sheet information or initial design parameters), an initial estimate of the system's frequency response can be obtained. This information can then be used to estimate the transducers and stimulus levels required to properly test the system, reducing test time and, in many cases, preventing damage to the system or device being tested.

These examples illustrate only a few of the applications in which the DSA's precision measurement hardware and computational power contribute to the modeling process. By providing analysis tools such as frequency response synthesis and curve fitting, the DSA provides a new level of support for meeting the complex as well as the routine challenges of modeling today's control systems.

Figure 2-17: Synthesizing the frequency response of a control system using the DSA's frequency response synthesis function.



Chapter 5: Design

Design: determining the combination of physical or theoretical components or parameters that will produce a desired action or result.

The design process, as defined above, occurs throughout the development of control systems. It begins with the initial conception of a system and becomes one of an unpredictable sequence of development processes which ultimately result in a refined, fully operational control system. Typically, the purpose of most design work (after conceiving the initial system) is to generate modifications to the initial system which will allow it to comply with the original design goals or specifications. Modifications can range from simple changes in component values to the design and addition of complex compensation networks.

5-1: Applying Frequency Response Synthesis, Waveform Math and Curve Fitting to the Design Process

As a design tool, DSAs offer several data processing functions which can aid the engineer in choosing combinations of components which will accomplish a desired task. For example, the frequency response synthesis function[1] can be used to predict the frequency response of compensation networks before they are actually built. The waveform math function[2] can then be used to predict the effects of a synthesized compensation network on a system's open-loop frequency response or predict the system's new closed-loop frequency response. It can even be used to estimate the step or impulse response of the modified system before the compensation network is built.

To illustrate the use of the DSA's data processing functions in the design process, the following case study examines the development of a simple compensation network for a motor speed controller.

Initial measurements on the motor speed control were taken with the control loop closed and the system's open-loop gain set approximately 8 dB below the desired operating level. The closed-loop measurement indicated a sharp resonance at approximately 87.5 Hz, as shown in Figure 2-18A. The open-loop frequency response was then calculated from the measurement of $Y(j\omega)/S(j\omega)$ using the $T/(1 - T)$ calculation, as shown in Figure 2-18B.

The magnitude of the resonance at 87.5 Hz indicated that an 8 dB increase in the gain would cause the open-loop gain at 90 Hz to exceed 0 dB with the phase less than -180 degrees, creating an unstable operating condition, as shown in Figure 2-19. Therefore, to achieve the desired increase in the system's open-loop gain, a compensation network was added to the system to reduce the level of the 87.5 Hz resonance.

The compensation network, in this case a two-pole low-pass filter, was developed by entering an initial estimate of the pole locations, gain and delay into the pole/zero table of the DSA's frequency response synthesis function. The synthesized frequency response of the low-pass filter was then displayed on the CRT of the DSA, as shown in Figure 2-20.

[1] See section 4-2 for a brief description of the frequency response synthesis function.

[2] See section 3-1 for a brief description of the waveform math function.

Figure 2-18: Closed-loop measurement and calculated open-loop frequency response showing resonance at 87.5 Hz.

A. Closed-loop measurement (*Y(jω)/S(jω)*)

B. Calculated open-loop frequency response (*GH(jω)*)

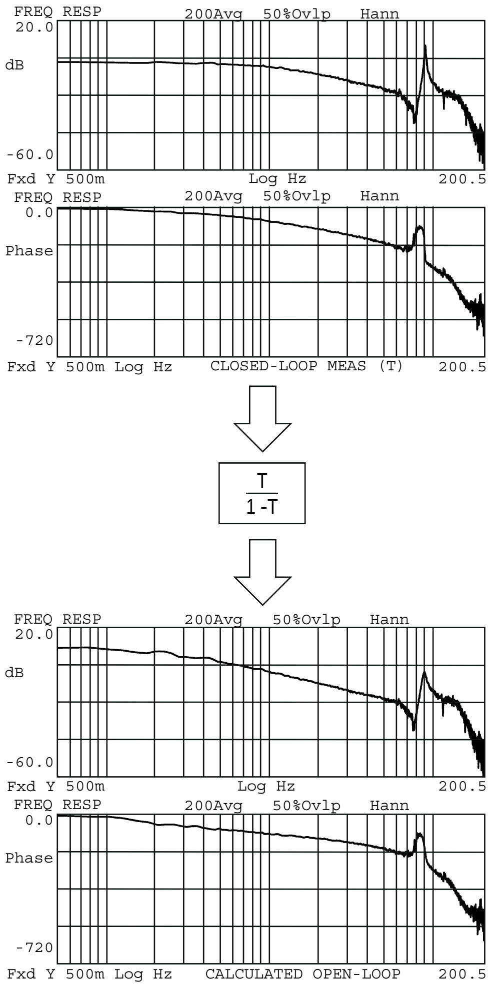


Figure 2-19: Using markers to predict the effect of increasing the open-loop gain by 8 dB.

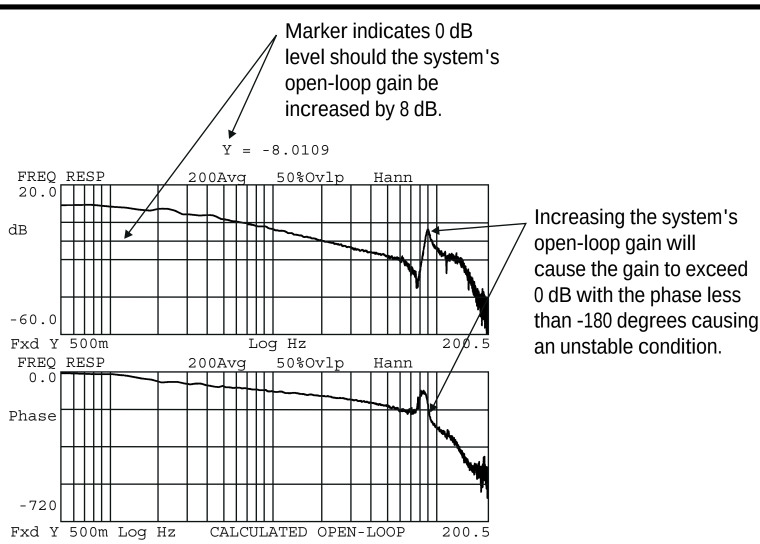
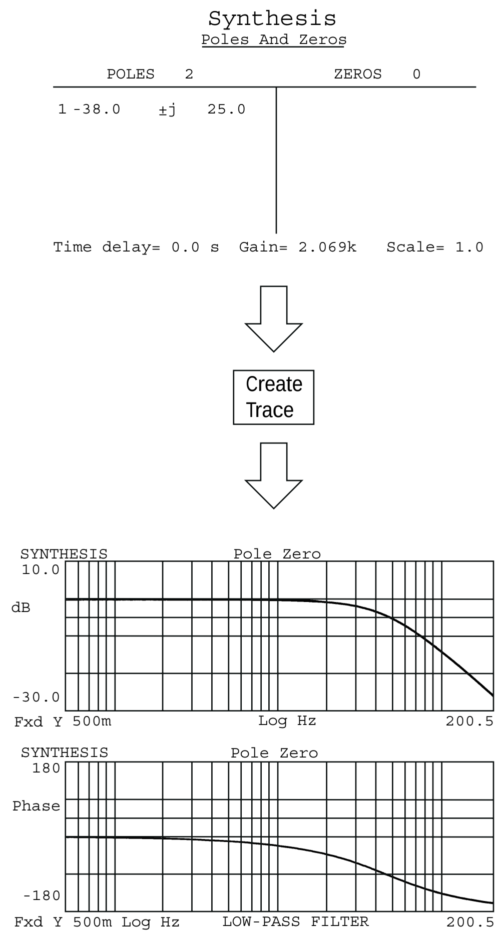
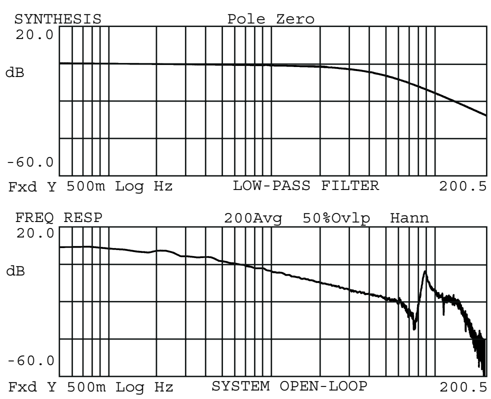


Figure 2-20: Using the synthesis function to calculate the frequency response of a low pass filter.

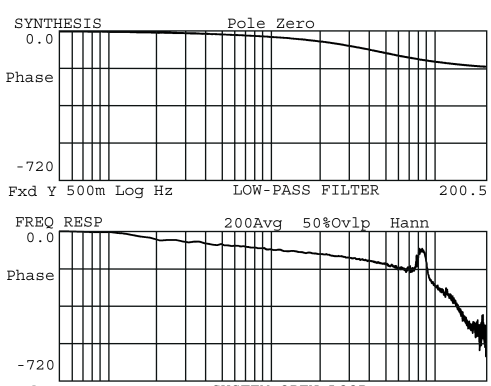


The frequency response of the speed control system and the synthesized frequency response of the low-pass filter were then displayed adjacently, as shown in Figure 2-21. By displaying both frequency responses in this fashion, the low-pass filter pole locations which provided the best trade-off between level rejection and phase shift could quickly be determined.

Figure 2-21: Synthesized frequency response of low pass filter (upper trace) and calculated open-loop frequency response of motor speed controller (lower trace).



A. Gain



B. Phase

Figure 2-22:
Using waveform math to calculate the effect of the low-pass filter on the open-loop frequency response.

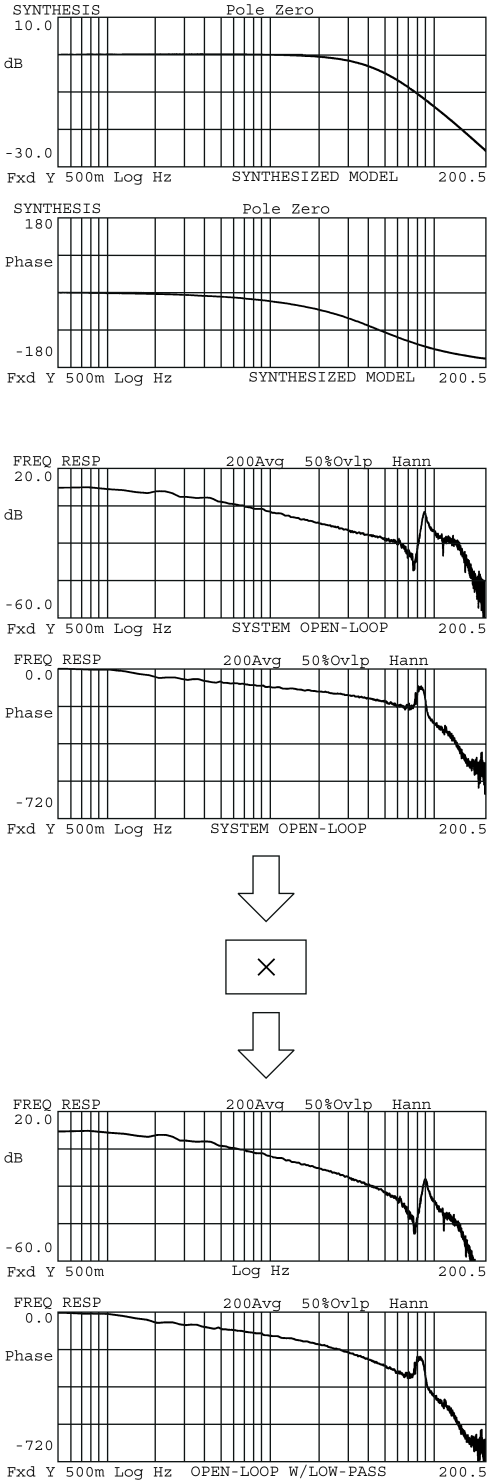
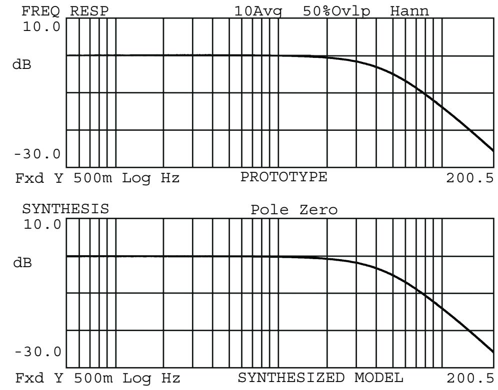
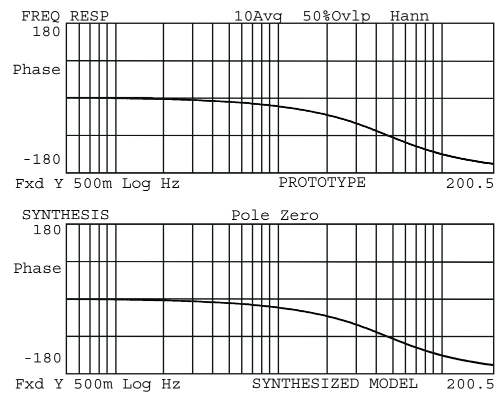


Figure 2-23:
Measured frequency response of prototype (upper trace) compared to synthesized frequency response (lower trace).



A. Gain



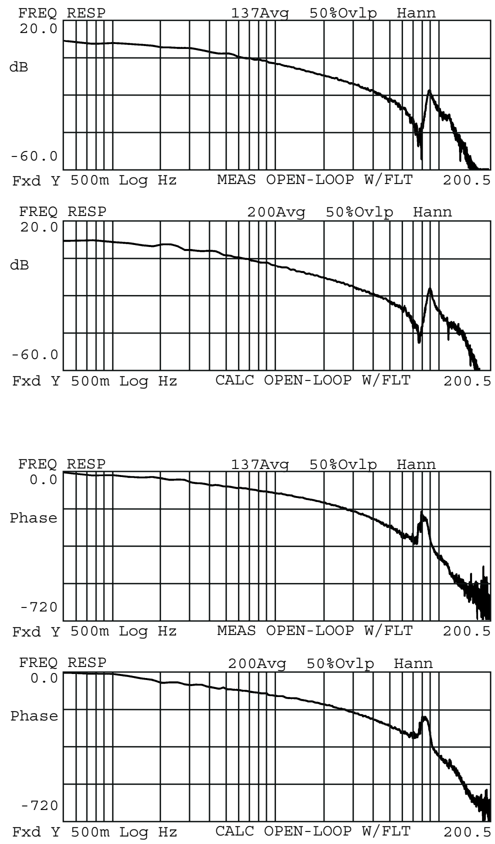
B. Phase

To verify the visual approximation, the synthesized frequency response of the low-pass filter was combined with the open-loop frequency response of the speed control system using waveform math, as shown in Figure 2-22.

Using the information provided by the pole/zero table and a passive filter design guide, the component values for the low-pass filter were determined and a prototype filter constructed. The frequency response of the prototype was then measured and compared to the synthesized frequency response, as shown in Figure 2-23.

With the low-pass filter installed in the forward signal path of the motor speed control, the open-loop frequency response was again measured and compared to the predicted response, as shown in Figure 2-24. Finally, the gain of the speed control was raised by 8 dB to provide the desired performance while maintaining reasonable gain margin and phase margin, as shown in Figure 2-25.

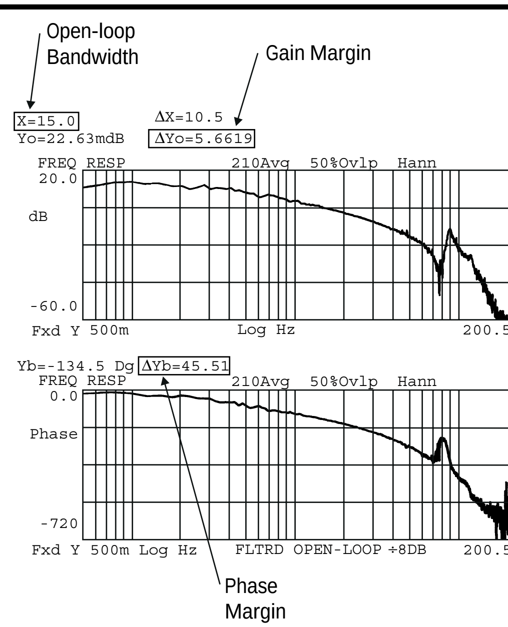
Figure 2-24: Comparison of the measured (upper trace) and the predicted (lower trace) open-loop frequency response with low-pass filter installed.



A. Gain

B. Phase

Figure 2-25: Measured open-loop frequency response with low-pass filter installed and gain increased by 8 dB.

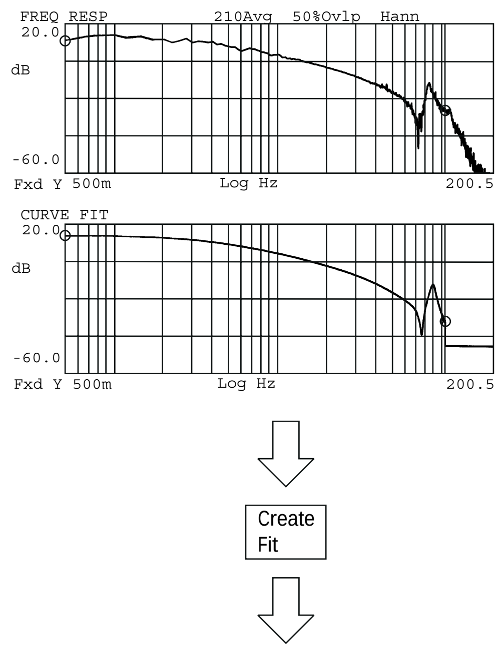


In this example, the low-pass filter provided enough compensation to achieve the desired system performance. However, for more demanding applications, a lag-lead network could be added to the system to further improve the system's performance.

When building compensation networks such as the lag-lead network mentioned above, the DSA's curve fitter can be used to locate the dominant poles and zeros of a system's open-loop frequency response, as shown in Figure 2-26. This information can then be used with design tools such as a root locus plot to select the most advantageous position for the poles and zeros of the compensation network.

The DSA's curve fitter function can also be used to suggest the location of a compensation network's poles and zeros. For example, the pole/zero model of a "perfect" compensation network can be derived using a combination of the frequency response synthesis, waveform math and curve fitting functions. First, the frequency response synthesis function is used to synthesize the "ideal" frequency response for a system. Waveform math is then used to divide the synthesized response by the system's measured frequency response. The result is the frequency response of the cascade compensation network needed to achieve the "ideal" frequency response for the system. By curve fitting this resultant frequency response, the DSA supplies the designer with a table of poles and zeros which will produce that response.

Figure 2-26: Curve fitting over frequency range of interest to determine the dominant poles and zeros of the system.



Curve Fit
Poles And Zeros

POLES 9		ZEROS 2
1	-3.9	-963.235m ±j 74.3426
2	-24.549 ±j 21.4567	
3	-24.3135 ±j 63.7758	
4	-2.72413 ±j 85.6206	
5	-15.2349 ±j 108.683	

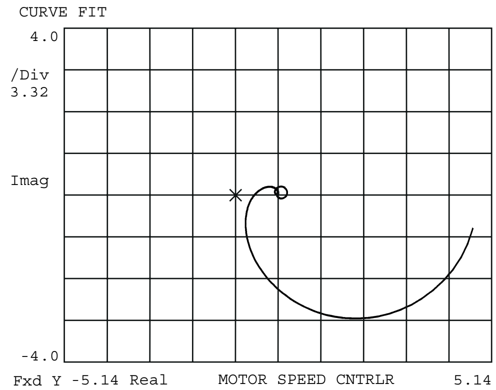
Time delay= 0.0 s Gain= 1.1E+12 Scale= 1.0

5-2: Using Display Formats Other Than the Bode Plot

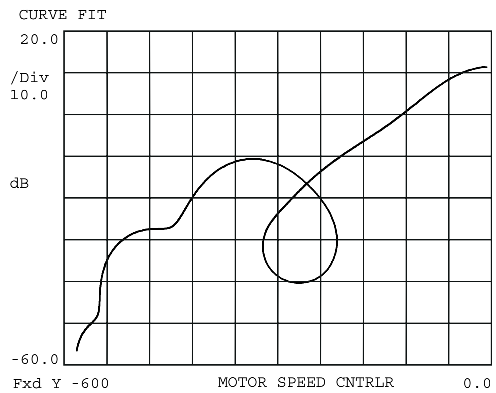
By providing a wide choice of coordinate formats, advanced DSAs allow the operator to observe frequency response data in the display format which best conforms with the design technique being used. For example, the open-loop frequency response of the motor speed controller can be displayed in either the Nichols or Nyquist formats as shown in Figure 2-27.

This rapid exchange of data between display formats not only allows the engineer to capitalize on the advantages of each display format, it also serves as a convenient way to bridge communication gaps between engineers accustomed to different display formats.

Figure 2-27: Open-loop frequency response of motor speed controller redisplayed in the Nichols and Nyquist formats.



A. Nuquist plot of open-loop frequency response.



B. Nichols plot of open-loop frequency response.

Chapter 6:
Summary

By combining the computational power of the microprocessor with the accuracy of precision measurement hardware, the Dynamic Signal Analyzer has expanded its functional scope to include contributions in virtually all aspects of control system development.

In the area of testing, the DSA has provided the facilities for making both time domain and frequency domain measurements. Using either the time capture or time throughput measurement modes, the DSA can store large quantities of time domain data. The data can then be either displayed in the time domain or routed to the FFT processor and transformed into frequency domain data.

For making frequency domain measurements, the DSA provides both FFT analysis and Swept Fourier Analysis. This combination of measurement capabilities allows the DSA to analyze a control system's response to a wide range of stimulus signals. This capability can often be used to gain greater insight into the operation of a control system as well as minimize measurement times.

In addition to providing multiple measurement capabilities, the DSA utilizes the power of the microprocessor to provide a host of automated measurement aids capable of optimizing measurement conditions and rejecting undesirable data.

In the area of analysis, the DSA provides functions such as coherence, waveform math, curve fitting and advanced display formatting as tools for reducing raw data to valuable information.

In the areas of modeling and design, the DSA's frequency response synthesis and advanced analysis functions can be utilized in the development of accurate system models and effective system designs.

Perhaps the DSA's most significant contribution is that it has brought both advanced measurement capabilities and powerful analysis tools together in a single instrument. This consolidation of development tools allows the DSA to provide a great deal of valuable information — not just data.

Appendicies
Appendix A: Glossary

Bandwidth. The interval separating two frequencies between which both the gain and the phase difference (of sinusoidal output referred to sinusoidal input) remain within specified limits.

Bode diagram. A plot of log-gain and phase-angle values on a log-frequency base, for an element transfer function $G(j\omega)$, a loop transfer function $GH(j\omega)$. The generalized Bode diagram comprises similar plots of functions of the complex variable $s = \sigma + j\omega$.

Characteristic equation. Of a feedback control system, the relation formed by equating to zero the denominator of a rationalized transfer function of a closed loop.

Closed loop (feedback loop). A signal path which includes a forward path, a feedback path and a summing point, and forms a closed circuit.

Compensation. A modifying or supplementary action (also, the effect of such action) intended to improve performance with respect to some specified characteristic.

Control system. A system in which deliberate guidance or manipulation is used to achieve a prescribed value of a variable.

NOTE: It may be subdivided into a controlling system and a controlled system.

Control system, automatic. A control system which operates without human intervention.

Control system, feedback. A control system which operates to achieve prescribed relationships between selected system variables by comparing functions of these variables and using the difference to effect control.

Control system, open-loop. One which does not utilize feedback of measured variables.

Critically damped. Describing a linear second-order system which is damped just enough to prevent any overshoot of the output following an abrupt stimulus. See also damping.

Critical point. (1) In a Nyquist diagram for a control system, the bound of stability for the locus of the loop transfer function $GH(j\omega)$, the $(-1, j\omega)$point. (2) In a Nichols chart, the bound of stability for the $GH(j\omega)$ plot; the intersection of $GH = 1$ with $GH = -180$ degrees.

Damping. (1) (noun) The progressive reduction or suppression of the oscillation of a system. (2) (adj.) Pertaining to or productive of damping.

Decibel. In control usage, a logarithmic scale unit relating a variable x (e.g., angular; displacement) to a specified reference level x, $dB = 20 \log x/x_0$.

NOTE: The relation is strictly applicable only where the ratio x/x_0 is the square root of the power ratio P/P_0, as is true for voltage or current ratios. The value $dB = 10 \log P/P_0$ originated in telephone engineering, and is approximately equivalent to the old "transmission unit".

Dither. A useful oscillation of small amplitude introduced to overcome the effects of friction, hysteresis or clogging.

Error constant. In a feedback control system, the real number K by which the nth derivative of the reference input signal is divided to give the resulting nth component of the actuating signal.

Frequency, damped. The apparent frequency of a damped oscillatory time response of a system resulting from a non-oscillatory stimulus.

Frequency, gain crossover. On a Bode diagram of the loop transfer function of a system, the frequency at which the gain becomes unity (and its decibel value zero)

Frequency, phase crossover. Of a loop transfer function the frequency at which the phase angle reaches ± 180 degrees.

Frequency response. In a linear system, the frequency-dependent relation in both gain and phase difference, between steady-state sinusoidal inputs and the resulting steady-state sinusoidal outputs.

Function describing. Of a nonlinear element under periodic input, a transfer function based solely on the fundamental, ignoring other frequencies.

Function, loop transfer. For a closed loop, the transfer function obtained by taking the ratio of the Laplace transform of the return signal to the Laplace transform of its corresponding error signal.

Function, output transfer. For a closed loop, the transfer function obtained by taking the ratio of the Laplace transform of the output signal to the Laplace transform of the input signal.

Function, return transfer. For a closed loop, the transfer function obtained by taking the ratio of the Laplace transform of the return signal to the Laplace transform of its corresponding input signal.

Function, system transfer. The transfer function obtained by taking the ratio of the Laplace transform of the signal corresponding to the ultimately controlled variable to the Laplace transform of the signal corresponding to the command.

Function, transfer. A mathematical, graphical, or tabular statement of the influence which a system or element has on a signal or action compared at input and at output terminals.

Gain (magnitude ratio). For a linear system or element, the ratio of the magnitude (amplitude) of a steady-state sinusoidal output relative to the causal input; the length of a phasor from the origin to a point of the transfer locus in a complex plane.

NOTE: The quantity may be separated into two factors: (1) a proportional amplification often denoted as K which is frequency-independent, and associated with a dimensioned scale factor relating the units of input and output; (2) a dimensionless factor often denoted as $G(j\omega)$ which is frequency-dependent. Frequency, conditions of operation, and conditions of measurement must be specified A loop gain characteristic is a plot of log gain vs. log frequency. In nonlinear systems, gains are often amplitude-dependent; see also transfer function.

Gain characteristic, loop. Of a closed loop, the magnitude of the loop transfer function for real frequencies.

Gain, closed-loop. The gain of a closed-loop system, expressed as the ratio of output to input.

Gain, loop. The absolute magnitude of the loop gain characteristic at a specified frequency.

Gain margin. Of the loop transfer function for a stable feedback system, the reciprocal of the gain at the frequency at which the phase angle reaches minus 180 degrees.

NOTE: Gain margin, sometimes expressed in decibels is a convenient way of estimating relative stability by Nyquist, Bode, or Nichols diagrams, for systems with similar gain and phase characteristics. In a conditionally stable feedback system, gain margin is understood to refer to the highest frequency at which the phase angle is minus 180 degrees.

M-peak. Of a closed loop, the maximum value of the magnitude of the return transfer function for real frequencies, the value at zero frequency being normalized to unity.

Nichols chart (Nichols diagram). A plot showing magnitude contours and phase contours of the return transfer function referred to ordinates of logarithmic loop gain and to abscises of loop phase angle.

Nyquist diagram. A polar plot of the loop transfer function.

NOTE: The "inverse Nyquist diagram" is a polar plot of the reciprocal function. The generalized Nyquist diagram comprises plots of the loop transfer function of the complex variables, where $s = \sigma + j\omega$ and σ and ω are arbitrary constants, including zero.

Overdamped. Damped sufficiently to prevent any oscillation of the output following a step or impulse input.

NOTE: For a linear second-order system the roots of the characteristic equation are real and unequal.

Phase angle, loop. Of a closed loop, the value of the loop phase characteristic at a specified frequency.

Phase characteristic, loop. Of a closed loop, the phase angle of the loop transfer function for real frequencies.

Phase margin. Of the loop transfer function for a stable feedback control system, 180 deg. minus the absolute value of the loop phase angle at a frequency where the loop gain is unity.

NOTE: Phase margin is a convenient way of expressing relative stability of a linear system under parameter changes in Nyquist, Bode or Nichols diagrams. In a conditionally stable feedback control system where the loop gain becomes unity at several frequencies, the term is understood to apply to the value of phase margin at the highest of these frequencies.

Pole. (1) Of a transfer function in the complex variable *s*, a value of *s* which makes the function infinite. (2) The corresponding point in the s-plane.

NOTE: If the same value is repeated n times, it is called a pole of nth order; if it occurs only once, a simple pole.

Resonance. Of a system or element, a condition evidenced by large oscillatory amplitude which results when a small amplitude of a periodic input has a frequency approaching one of the natural frequencies of the driven system.

NOTE: In a feedback control system, this occurs near the stability limit.

Response, steady-state. Of a stable system or element, that part of the time response remaining after transients have expired.

NOTE: The term steady-state may also be applied to any of the forced response terms: for example, "steady-state sine-forced response".

Root locus. For a closed loop whose characteristic equation is $KG(s)H(s)+1 = 0$, a plot in the s-plane of all those values of *s* which make $G(s)H(s)$ a negative real number; those points which make the loop transfer function $KG(s)H(s) = -1$ are roots.

NOTE: The locus is conveniently sketched from the factored form of $KG(s)H(s)$; each branch starts at a pole of that function, with $K = 0$. With increasing K, the locus proceeds along its several branches toward a zero of that function and, often asymptotic to one of several equiangular radial lines, toward infinity. Roots lie at points on the locus for which (1) the sum of the phase angles of component $G(s)H(s)$ vectors totals 180 deg., and for which (2) $1/K = |G(s)H(s)|$. Critical damping of the closed loop occurs when the locus breaks away from the real axis; instability when it crosses the imaginary axis.

Servomechanism. An automatic feedback control system in which the controlled variable is mechanical position or any of its time derivatives.

Servomechanism type number. In control systems in which the loop transfer function is:

$$\frac{K(1 + a_1 s + a_2 s^2 + \ldots + a_i s^i)}{Sn(1 + b_1 s + b_2 s^2 + \ldots + b_k s^k)}$$

where K, a, b etc. are constant coefficients, the value of the integer n.

Stability. For a control system, the property that sufficiently bounded input or initial state perturbation result in bounded state or output perturbations. Time, rise. The time required for the output of a system (other than first-order) to make the change from a small specified percentage (often 5 or 10) of the steady-state increment to a large specified percentage (often 90 or 95), either before overshoot or in the absence of overshoot.

NOTE: If the term is unqualified, response o a unit-step stimulus is understood, otherwise the pattern and magnitude of the stimulus should be specified.

Time, settling (correction time). The time required following the initiation of a specified stimulus to a linear system for the output to enter and remain within a specified narrow band centered on its steady-state value.

NOTE: The stimulus may be a step, impulse, ramp, parabola, or sinusoid. For a step or impulse, the band is often specified as ±2%. For nonlinear behavior, both magnitude and pattern of the stimulus should be specified.

Underdamped. Damped insufficiently to prevent oscillation of the output following an abrupt stimulus.

Zero. (1) of a transfer function in the complex variable *s*, a value of *s* which makes the function zero. (2) The corresponding point in the s-plane.

NOTE: If the same value is repeated n times, it is called a zero of nth order; if it occurs only once, a simple zero.

Appendix B: Bibliography

American National Standard Association, 1981. *Terminology for Automatic Control*, ANSI MC85.1M-1981. American Society of Mechanical Engineers, New York, NY.

Brogan, W. L., 1974. *Modern Control Theory.* Guantum, New York, NY.

D'Azzo, J. J.. and C. H. Houpis, 1981. *Feedback Control System Analysis and Synthesis.* McGraw-Hill, New York, NY.

DiStefano, J. J., A. R. Stubberud, and I. J. Williams, 1967. Schaum's *Outline of Theory and Problems of Feedback and Control Systems.* Schaum, New York, NY.

Dorf, R. C., 1980. *Modern Control Systems.* Addison-Wesley, Reading, MA.

Gabel, R. A., and R. A. Roberts. *Signals and Linear Systems.* John Wiley, New York, NY.

Hewlett-Packard Co. Digital S*ignal Analysis*, *Feedback Control System Measurements.* Application note 240-1.

Hewlett-Packard Co. *Measuring the Coherence Function with the HP 3582A Spectrum Analyzer.* Application note 245-2.

Hewlett-Packard Co., 1982. *The Fundamentals of Signal Analysis.* Application note 243.

Institute of Electrical and Electronics Engineers. Standards Committee, 1972. I*EEE Standard Dictionary of Electrical and Electronics Terms.* IEEE, New York, NY, IEEE Standard 100-1972.

Kuo, B. C., 4th Ed., 1982. *Automatic Control Systems.* Prentice-Hall, Englewood Cliffs, NJ.

MacFarlane, A. G. J., ed. *Frequency-Response Methods in Control Systems.* IEEE Press, New York, NY.

Nagrath, I. J., and M. Gopal, 1982. *Control Systems Engineering.* Wiley Eastern, New Delhi, India.

Thaler, G J., ed., 1974. *Automatic Control: Classical Linear Theory.* Dowden, Hutchinson, & Ross, Stroudsburg, PA.

Thaler, G. J. and R. G. Brown, 1960. *Analysis and Design of Feedback Control Systems.* McGraw-Hill, New York, NY.

Chen, C.T., 1984. *Linear System Theory and Design.* Holt, Rinehart, and Winston.

Appendix C: Acknowledgments

Many of the definitions within this application note have been either influenced by, or directly reprinted from American National Standard ANSI MC85.1M-1981 with permission from the publisher.

For more information about Hewlett-Packard test & measurement products, applications, services, and for a current sales office listing, visit our web site, http://www.hp.com/go/tmdir. You can also contact one of the following centers and ask for a test and measurement sales representative.

United States:
Hewlett-Packard Company
Test and Measurement Call Center
P.O. Box 4026
Englewood, CO 80155-4026
1 800 452 4844

Canada:
Hewlett-Packard Canada Ltd.
5150 Spectrum Way
Mississauga, Ontario
L4W 5G1
(905) 206 4725

Europe:
Hewlett-Packard
European Marketing Centre
P.O. Box 999
1180 AZ Amstelveen
The Netherlands
(31 20) 547 9900

Japan:
Hewlett-Packard Japan Ltd.
Measurement Assistance Center
9-1, Takakura-Cho, Hachioji-Shi,
Tokyo 192, Japan
Tel: (81) 426 56 7832
Fax: (81) 426 56 7840

Latin America:
Hewlett-Packard
Latin American Region Headquarters
5200 Blue Lagoon Drive
9th Floor
Miami, Florida 33126
U.S.A.
Tel: (305) 267-4245
(305) 267-4220
Fax: (305) 267-4288

Australia/New Zealand:
Hewlett-Packard Australia Ltd.
31-41 Joseph Street
Blackburn, Victoria 3130
Australia
Tel: 1 800 629 485 (Australia)
0800 738 378 (New Zealand)
Fax: (61 3) 9210 5489

Asia Pacific:
Hewlett-Packard Asia Pacific Ltd.
17-21/F Shell Tower, Times Square,
1 Matheson Street, Causeway Bay,
Hong Kong
Tel: (852) 2599 7777
Fax: (852) 2506 9285

Printed in U.S.A. 3/98
5953-5136E