This manuscript was the basis for the article
"A Refresher Course in Control Theory"
printed in Machine Design, September 9, 1999.

Use Control Theory to Improve Servo Performance

George Ellis

Introduction

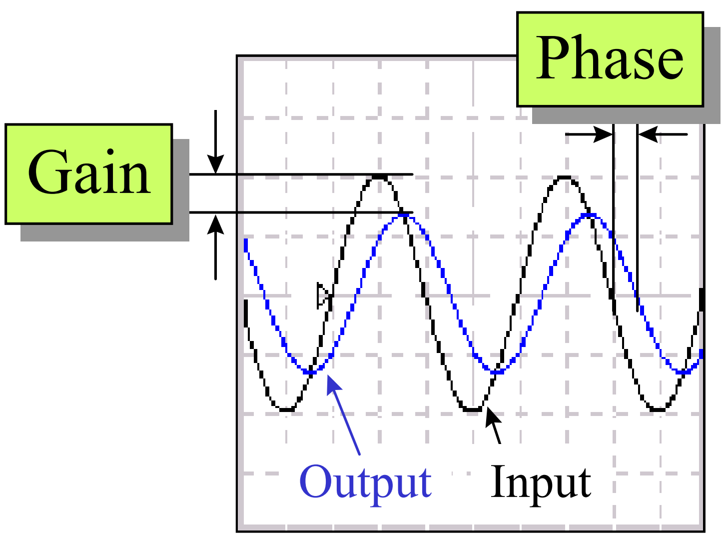
For the people who install and commission motion systems, control theory can seem mysterious. It is true that some controls problems require great expertise, but that's not the normal case. Most engineers can address the majority of controls problems by knowing how to apply just a few principles. Of course, the engineer who's only goal is getting machinery operational, does not need to understand control theory at all. However, if the goal is to produce the best performing machines in your market, an understanding of applied control theory is necessary.

Many engineers learn the basics of controls in college, but they don't learn how to apply them at work and the unused information is quickly forgotten. In this article we will try to revive that knowledge as we discuss a few basic principles. We will review Bode plots and the *open loop method*, a reliable and simple method that characterizes system stability.

Bode Plots

Bode plots provide abundant information about the stability and responsiveness in one picture. Many times, the step response displayed on an oscilloscope is used to measure the performance of control systems. That's understandable because a step is measured with an oscilloscope (usually available). Bode plots provide much more information than a step-response.

Bode plots show the phase and gain of a control system over a range of frequencies. Phase and gain describe the response of a control system to a sine wave as shown in Figure 1. For linear control systems, sine waves are unique in that sine wave inputs generate sine wave outputs. Phase describes the time shift between input and output. Gain describes amplitude variation.

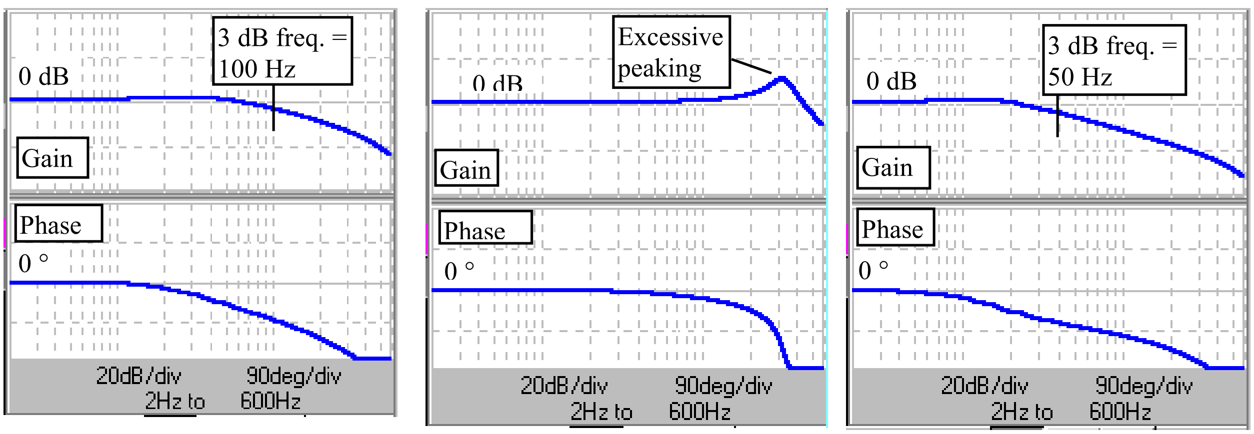


Phase and Gain
Figure 1

Phase is usually measured in degrees where at any frequency, one cycle is defined as 360°. For example, for a 100 Hz sine wave, 360° is equivalent to a 10 ms. If t_{DELAY} is the time shift between input and output, phase ≡ t_{DELAY}*Freqency*360. Gain is measured on a logarithmic scale called decibels or dB. Decibels are defined as 20 times the 10-based logarithm of the ratio of output to input: gain ≡ 20*$\log_{10}$(Output/Input). Bode plots show phase and gain against frequency plotted on a log scale. Example Bode plots are shown in Figure 2.

Measuring Performance

Most information of the Bode plot is in the gain. Stability is seen in the smoothness of the plot. Desirable plots such as Figure 2a have flat gain which, for low frequencies is 0 dB (or unity), and which falls as frequency increases. Peaking, the undesirable characteristic where gain increases above 0dB (see Figure 2b), usually indicates marginal stability. Such a system yields excessive overshoot and ringing. Bode plots also indicate responsiveness. The further to the right the gain curve falls off, the more responsive the system. The frequency where the gain falls to -3dB is often called the bandwidth. To compare to the time domain, step response settling time is roughly 0.5/bandwidth. The Bode plot in Figure 2c shows a stable but sluggish system—the bandwidth is half that of Figure 2a.



2a) Well tuned: little peaking, 100Hz Bandwidth

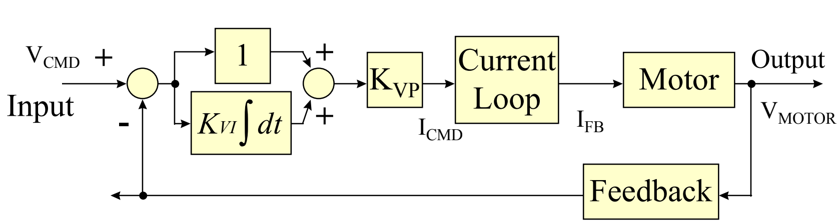
2b) Peaking indicating marginal stability

2c) Stable, but sluggish (50 Hz BW)

Figure 2 Bode plots for various systems

Transfer Functions

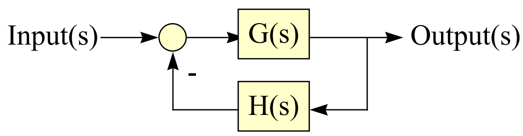
Reasonably linear systems, such as most commercial motion control systems, can be diagrammed as a series of transfer functions. Transfer functions show the relationship between the input and output of a component such as a motor, feedback device, a PID controller, or a low-pass filter. Block diagrams show how these components connect to each other. Figure 3 below shows a block diagram of a simple proportional-integral (PI) control system.



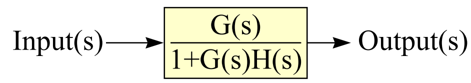
Block diagram for a PI Velocity Controller
Figure 3

G/(1+GH) Rule

An important conceptual block diagram is the simple feedback loop shown in Figure 4a. This block diagram depicts functions of "s", the Laplace operator. Control systems rely on s-based transfer functions. If your Laplace-Transform skills are rusty, don't be concerned—control engineers rarely calculate Laplace transforms. Normally we use the s-domain because s-based functions can be combined easily. For example, with a little algebra, the transfer function of this system of Figure 4a is seen to be G/(1+GH) as shown in Figure 4b.



a) Simple feedback loop



b) Single-block equivalent to simple feedback loop

Block diagrams
Figure 4

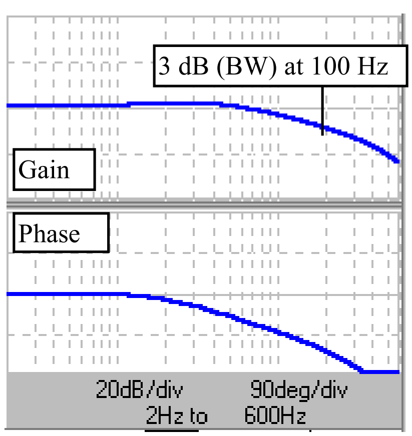
As simple as the loop of Figure 4a appears, it provides insight into control system instability. Instability occurs when the denominator in Figure 4b becomes zero. In other words, when G(s)H(s) = -1. Recalling that 0 dB is equivalent to unity, -1 can be written as 0dB∠-180°. G(s)H(s) is called the *open loop transfer function*.

The Open Loop Method

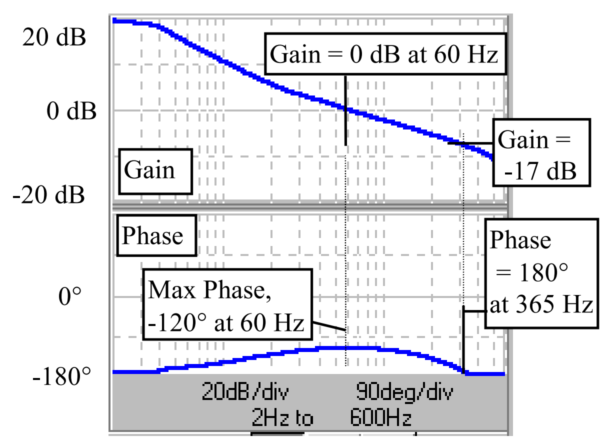
The *open-loop method* measures system stability by characterizing how close the open loop transfer function, G(s)*H(s), gets to 0dB∠180° (or -1). Figure 5 shows closed and open loop Bode plots side-by-side. As we discussed, the closed loop gain (Figure 5a, top) starts at 0 dB and rolls off near the bandwidth. Notice how the open loop gain (Figure 5b, top) is very high at low frequencies and falls as frequency increases. This is typical—motor inertia is easy to move at low frequencies but difficult to move at high frequencies.

Phase of the Closed and Open Loops

The phase of the closed loop (Figure 5a, bottom) at low frequencies is near zero, as you would expect. This means the closed loop follows the command with little time delay at low frequencies. At higher frequencies, the closed-loop phase rolls off. Again, this is expected—motors follow high frequency commands with more phase delay. The open loop phase (Figure 5b, bottom) is different. It starts at -180, rises a little, and then falls again. This is typical for PI control systems.



a) Closed-Loop Bode Plot



b) Open-Loop Bode Plot

Figure 5

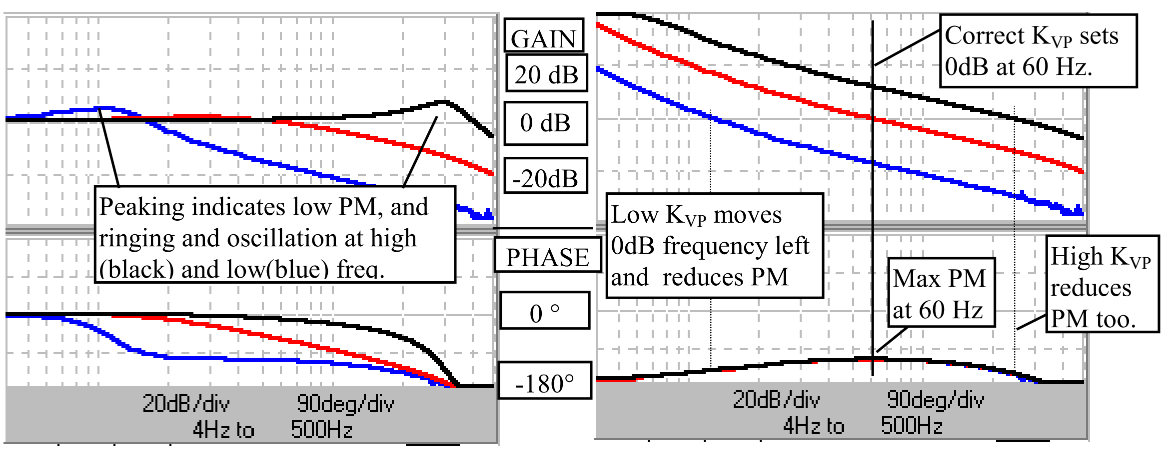
Phase and Gain Margin

The open-loop method works by estimating two measures of stability or *margins*. The first is *phase margin* or *PM*. Find PM by locating the frequency where the loop gain is 0 dB. For Figure 5b, that is about 60 Hz. Knowing that the system will be unstable if the phase is -180°, PM is the difference between the actual phase and -180°. For Figure 5b, the open loop phase is about -120°so the PM is 180°-120° or 60°. The PM can be seen graphically in Figure 5b as the distance from the open loop phase to -180°, the bottom of the graph. *Gain margin* or *GM* is similar. Find the frequency where the open-loop phase is -180° and then measure the gain. Since the system will be unstable at 0dB∠-180°, the GM is 0dB less the gain. For Figure 5b, the open-loop phase is -180° at about 365Hz and the gain at that frequency is about -17dB; thus, the GM is 17 dB. Typical values for gain and phase margin vary by application, just as overshoot and settling time requirements vary. Most machines require at 10-20dB GM and 50° to 75° PM.

Example: A PI Velocity Controller

Consider the PI controller of Figure 3 to demonstrate the open loop method. This form of PI control has two gains: K_{VI} for the integral of velocity error and K_{VP}, the loop gain constant. The Bode plots from Figure 5 show the open and closed loop gains. As you can see from Figure 5b, the gains have been chosen to maximize the PM. Notice that the open loop gain crosses through 0 dB at 60 Hz, the frequency where the open-loop phase is at its largest.

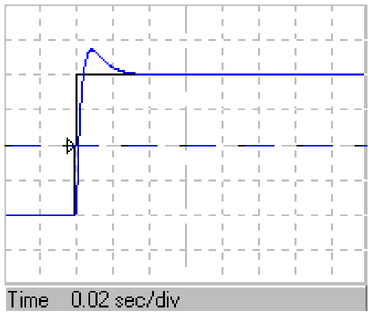
PM can be used to explain a curious characteristic of control systems. When the loop gain (K_{VP}) is too high, the system rings excessively. The same thing happens when the loop gain is too low. As discussed above, PM is maximized at 60 Hz, the frequency where the gain is 0dB in the correctly tuned (red) system of Figure 6b. This occurs when $K_{VP} = 0.72$. Raising K_{VP} to 3.0 (Figure 6b, black) moves the entire gain plot up (by virtue of raising the loop gain) and moves the frequency of 0 dB gain too far right, reducing PM. Lowering K_{VP} to 0.1 moves the point of 0 dB too far left, which also reduces PM. Figure 6a shows the closed loop gain of three systems. Low phase margin (Figure 6b) correlates closed loop peaking (Figure 6a).



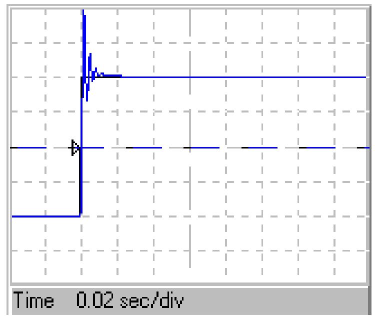
b) Closed Loop Gain (top) and Phase *b) Open-Loop Gain(top) and Phase*

Bode plots for PI controller with three gains
Figure 6

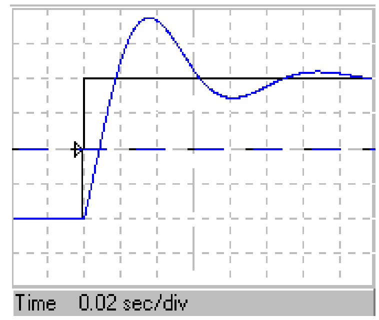
Figure 7 shows the oscilloscope graphs of the Bode plots from Figure 6. Note how peaking in the closed-loop gain correlates to excessive overshoot in the scope plots. Figure 7a ($K_{VP} = 0.72$) has little overshoot, but the high K_{VP} (Figure 7b) and the low K_{VP} (Figure 7c) have excessive ringing albeit at widely differing frequencies. This confirms the principle: low phase margin implies peaking in the closed loop which implies excessive overshoot and ringing in the scope plots.



Correct K_{VP} (0.72), corresponds to Figure 6, Red



High K_{VP} (3.0), corresponds to Figure 6, Black



Low K_{VP} (0.1), corresponds to Figure 6, Blue

Scope Plots of the Systems Shown in Figure6
Figure 7

In general, raising K_{VP} increases the loop gain and increases the frequency where the gain is 0 dB. Increasing K_{VI} increases the gain at low frequencies and simultaneously reduces the phase margin. Reducing the PM is usually the dominant effect, which explains why large integral gains are so often associated with overshoot and ringing.

Conclusion

So, how does understanding these concepts help you build a better servo system? First, the understanding of PM and GM help explain how tuning gains work and allow engineers to develop consistent methods of tuning. This in turn leads to more uniform operation between machines. Second, understanding that lower phase margin implies less stability shows that you should reduce delay in the control loop where possible. This includes eliminating unnecessary low-pass filters and increasing the sample rate of digital controllers. Finally, the advanced user can measure these margins on actual machines for even more consistent performance and more certain identification of mechanical problems (see sidebar, "Measuring the Machine".)

Although this article discusses the open-loop method for a simple system, these concepts can extended to other velocity controllers and to position controllers. Unlike most methods taught in school, the open-loop method extends easily from analog to digital systems because the effect of sampling delay, the predominant effect in digital controls, is simply another contributor to phase lag. The simplicity and flexibility of the open-loop method make it an ideal way to measure and improve system performance for the practicing controls engineer.

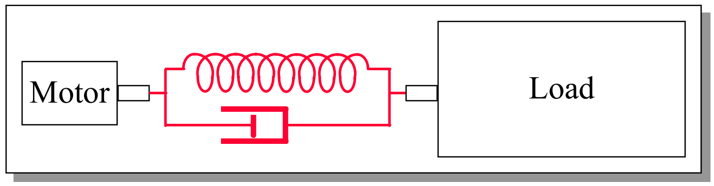
ModelQ

Would you like to try the open loop method yourself? All the examples in this article are produced with a simulation environment called *ModelQ*. *ModelQ* is a stand-alone simulation package designed for teaching principles of control theory. The models are fixed, but you can easily adjust constants to observe the effects of control system parameters on performance. *ModelQ* is available free of charge at *www.???.com\??*. Just download, install, click "Run." On-line help is included.

Resonance

Resonance is a problem for a lot of motion systems. A compliant coupling load and motor cause resonance. In effect, the shaft and transmission become a spring imperfectly linking two inertias: that of the motor and load. Those two inertias resonate with one another. Resonance is a primary limitation in many high-performance applications. Typically, the controls engineer tunes a system by increasing gains until the system begins to show signs of instability. This instability is often the result of mechanical resonance.

What can you do to improve resonance? After the machine is built, the most common steps are using low-pass and notch filters, which are provided by most drive manufacturers to deal with resonance. Low pass filters generate phase lag, which reduces PM. If the filter frequency is too low, system performance is compromised. Notch filters provide filtering over a narrow band of frequencies configured to attenuate the resonance. Notch filters have little effect on PM, but they don't work well if the resonant frequency varies over machine operation.



Resonance is caused by compliant coupling between motor and load
Figure A

While you can use the control system to patch over resonance at the end of the design, the best place to correct resonance is at the source. Frequently the machine needs to be stiffened. For example, you can use stronger shaft couplings, shorter shafts, idlers on long belt spans, and stronger frames. This is the reason why higher quality servo motors have large diameter shafts. Making the machine stiffer has the effect of increasing the system spring constant, raising the resonant frequency. In the best case, stiffening the machine corrects the problem entirely. At the very least, it provides a better starting point for using filters. Treating resonance as a mechanical problem rather than correcting it with filters usually provides superior results.

Measuring the Machine

Bode plots and the open loop method help engineers understand how control systems work. This understanding brings benefits such as helping engineers design consistent tuning procedures and guiding engineers by helping to understand the need to remove unnecessary phase lag. For those that want to go further, modeling programs like MatLab {web-site}, VisSim {web-site}, and P-Spice {web-site} can be used to build simple models of machines. Here, machine and drive parameters are entered as closely as possible and system performance is approximated. Modeling, verified by comparing the step responses of the model and machine, frequently provides satisfying results.

The most demanding motion applications require that the actual machine Bode plots be measured. This gives the most complete identification of system operation. The measured Bode plot tells how the machine is operating across a broad frequency range and provides much more information than a step response. It also identifies mechanical problems, especially resonances. Measurements help the initial machine integration by identifying problems that may need to be corrected. Measurements also help with problem machines both at the factory and in the field. For example, if a machine in production could not be tuned well, a Bode could identify an unusual resonance that was caused by, say, a damaged coupling. Measuring PM and GM on every machine provides assurance of consistent performance that cannot be achieved relying on step responses.

Instruments

When you have decided to measure a Bode plot, you will probably need to acquire an instrument called a "Dynamic Signal Analyzer." These products inject a waveform into a control system and measure the output. Often, that waveform is a series of sine waves, but sometimes complex waveforms such as pseudo-random noise or a *chirp* signal are used. The output and input are compared and used to generate information for a Bode plot. These instruments are produced by several companies including DSP Technologies (www.DSPT.com), HP, and Schlumberger.

Kollmorgen ServoStar System (left) being measured by DSP Technologies SigLab, a Dynamic Signal Analyzer which uses a laptop computer
Figure B

If you want to consider purchasing a dynamic signal analyzer, you need to shop around. The devices are expensive ($5,000-$15,000) and vary widely. Some are tightly integrated to a PC environment, while others are stand-alone. If you plan to take the instrument into the field, bear in mind that these instruments vary from the size of a laptop computer to that of a large suitcase.

Be aware that measuring a motion system can require cooperation with your drive vendor. These analyzers are generally analog and the drive must produce the necessary analog signals (velocity or torque). Also, depending on the types of position or velocity loops in the system, you may need to do some work to convert the measured data into open-loop plots (for GM and PM) or closed-loop plots (for bandwidth.) Measuring a system's Bode plot can be a big job, but the payoff is often bigger.