TECHNICAL NOTES

Motion Control

Celera MOTION
Encoders | Motors | Mechatronics

Servo Terminology

TN-9002 | REV 160627

INTRODUCTION

Many engineers without formal servo control training are asked to size motors and choose encoders for servo systems. During this process, these engineers interact with motor and encoder vendors, discussing how the hardware they choose will deliver the performance required. This tech notes describes many of the commonly used servo terms, giving the reader some useful insight when designing a motion system.

WHAT IS A "SERVO"

Before we dive into the terminology of servo systems, let's start with the basic word "servo" and what it means. Engineers will often use the term "servo" or "servo loop" to describe any type of closed loop control system, whether it be a DC motor and encoder, or a piezoelectric stack actuator and capacitance probe. In this case, the word is typically used as a noun. An example might be, "The servo is stable and performed the move within its allotted time requirement." The word servo, however, is sometimes used as verb. For example, an engineer might say "the stage servo'd to the final position without any problem." In summary, the word "servo" is commonly meant to represent the entire closed loop system, meaning the actuation device, the sensing device, the drive and control electronics, and the control algorithm.

Figure 1 below shows the basic servo block diagram applicable to most electro-mechanical servo systems. There are five main elements to the servo system, 1) Controller, 2) Drive Electronics, 3) Actuation Device, 4) Plant and 5) Sensor. The controller is the brains behind the servo system. It is regulating the direction and magnitude of control effort that is needed to arrive at the desired goal. The desired goal in all servo systems is to reduce or minimize error. The Drive is the component that converts low voltage electrical signals from the controller into high power current or voltage signals that deliver power to the Actuation Device (Item 3). Common actuation devices for electro-mechanical servos are DC motors (linear and rotary), and piezo-electric actuators. These devices are often built into, or coupled with more complex devices such as valves, gear boxes, etc. The fourth component is the Plant. This component is described in more detail later, so for now we will describe it as the structure that supports the actuation device and the sensor. It also includes the moving load that the actuation device is connected to. The last component is the sensor. This component converts a physical performance metric, such as position, velocity, pressure, or flow, into a signal the controller can use as an input (i.e., voltage or counts).

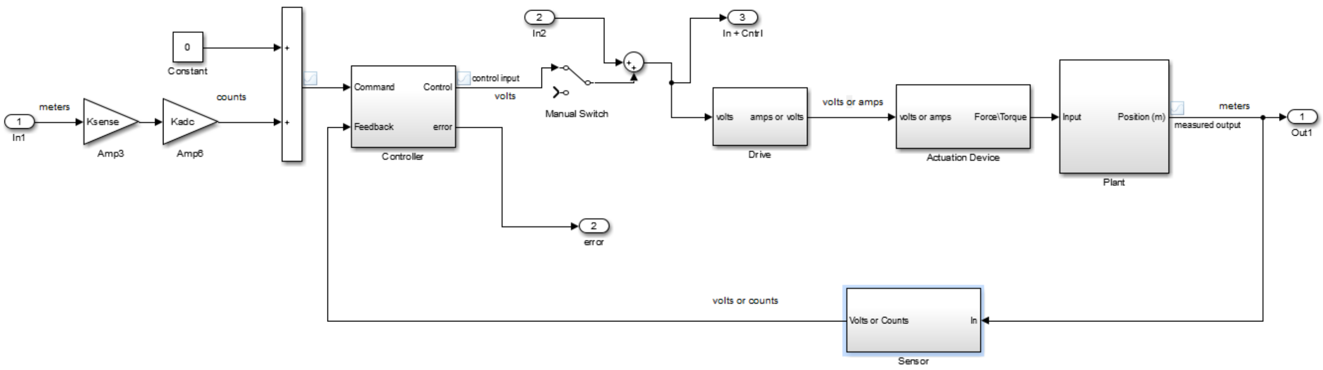

Figure 1. Servo Loop Diagram

TRANSFER FUNCTION

The term transfer function is used to describe a frequency domain relationship between an input and output signal. For example, if a signal goes through a typical inverting op-amp with gain of -10, the transfer function would be 0 at 0.1 Hz, 10 at 10 Hz, and 10 and 1 kHz and the phase would be -180 at each frequency.   The -180 degrees is the equivalent representation of the (-) sign of the gain. In this simple case, the op-amp adjusts the magnitude of the signal the same amount at all frequencies. Now consider a first order low pass filter. If we compare the signal out to the signal in, we see the signals are identical in magnitude before the filter frequency, and a smaller output signal for frequencies above the filter frequency. While the magnitude ratio is changing from unit to a number less than 1, there is a corresponding change in the phase, starting at 0 degrees and ending at -90 degrees.

The magnitude and phase relationship versus frequency is known as "Transfer Functions". This measurement is a key tool for servo engineers to analyze components within the servo loop, and the servo loop as a whole. Transfer Functions across the controller block are often simulated, so the servo engineer can tune the gains accordingly. Transfer Functions across the plant are frequently simulated, but are also measured, so engineers can see exactly where the mechanical resonances are occurring and adjust the control algorithm or mechanical design accordingly. Once the loop is closed, transfer functions are made again to measure the stability of the entire servo loop. In this case, you may wonder how it is possible to close the loop without first knowing if the gains will be stable (i.e. taking the transfer function). In most cases, there are models that provide enough information to close a low bandwidth loop first, then after stability analysis is done based on the actual transfer function data, the loop is re-tuned to a higher bandwidth.

PLANT

The "Plant" is a term used by control engineers to describe the component(s) that are under servo control. The plant could be RL circuit, or it could be spring, mass system, or it could be fluid flowing in a pipe. These physical components can all be represented in the Laplace domain with poles and zeroes, forming the "Plant Transfer Function" or PTF as referred to in this tech note. Ideally the Plant should not include drive dynamics or sensor dynamics, as those components have their own blocks as shown in Figure 1. Having said that, it is very common for the drive and sensor to be lumped into the Plant, so one must be sure to know what is included and what is not when discussing Plant transfer function data.

An example of a PTF is shown in Figure 2 below. In this example there are two moving masses, one with a spring and damper to ground and a second connected to the first through a second spring and damper pair (Figure 3). The peak seen near 10 Hz is the resonant frequency of the two masses moving out of phase with one another. The units on the magnitude plot of Figure 3 are shown generically as dB, but for the plant, the units are meters per Newton.

Shown here is a very simplistic PTF. In the real world, the structure that supports the motor and sensor are not infinitely stiff and will have dynamics affecting the stability of the servo loop. These high frequency dynamics (or poles) will typically require filtering form the control algorithm. It is critical that care be taken in design of the supporting structure of any motion system.

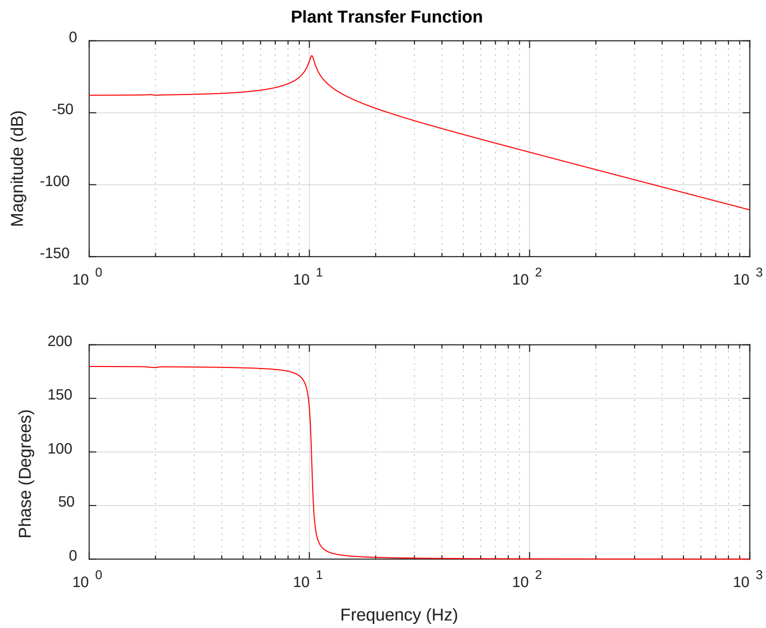

Figure 2. Plant Transfer Function Example

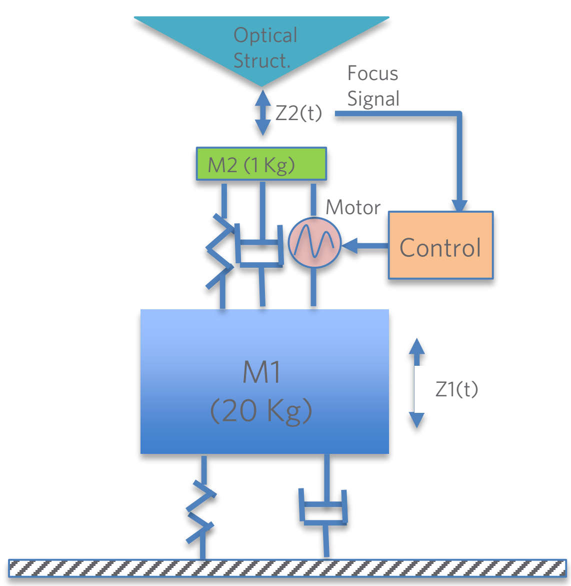

Figure 3. Physical Plant model

GAINS

A servo gain is attributed to any component in the servo loop that alters the control signal's magnitude. There are many types of "Gain" in a servo loop. Some gains are frequency dependent and some gains are constant at all frequencies. The most notable servo gains are Control Gain(s), Drive Gain, Motor Gain, Plant Gain, and Sensor Gain. Every servo gain gets multiplied together to create the overall Loop Gain. Changing any one component will affect the closed loop stability. Component specific gain examples would be; a) for an encoder, the gain would be counts per micron, for a capacitance probe it would be volts per micron; b) for a rotary motor, the motor gain would be Newton-meter per amp; c) for a current drive it would amps/volt; and d) for a PWM drive it would be Volts per %PWM.

One common misconception is that systems with large masses or inertias cannot have high servo bandwidths or cannot move and settle at high speeds. This statement is simply not correct. The mass or inertia of a system is just another gain in the Loop. The more mass or inertia there is in a system, the lower the gain is for the plant. But the Control Gain for example, can be increased to achieve the maximum bandwidth, provided stability rules are met. The limiting factor, however, in large mass servo systems, is the available power to move the load or reject disturbances. This is where the confusion comes in. A large mass system might be stable with a 100 Hz bandwidth, but when the mass is moved, or subjected to the disturbance input, the dissipated power becomes so large that either the motor heats up, or performance is limited due to lack of current or voltage from the drive electronics. A motor with a larger motor constant will heat up less, and more available current or voltage will eliminate saturations. The limiting factor for large mass or inertia systems is not servo bandwidth, rather it is the components selected to drive the mass (i.e. the motor and drive electronics).

PID

PID is probably the most commonly used term of a servo system. The PID is an acronym for the Proportional, Integral and Derivative gain set that make up all, or part of the control algorithm. The "P" gain acts on the error term directly, the "I" gain acts on the integral of the error term and in most cases "D" gain acts on the derivative of the error term (there are some cases where the D gain will act on the velocity of the moving load). For position loops, the P gain acts like a spring in a mechanical system, meaning the P gain generates force proportional to displacement (position error). I gain acts on the integral of the error, or the cumulative history of the load's position. This term is used to bring steady state error to zero but can be very problematic for systems with friction, dead-band, or low sensor resolution. The most common issue for systems with a dead-band and integral control is "limit cycling".

Limit Cycling is when the system is not fully unstable, meaning the output force is not growing without bound, rather it keeps overshooting the target position over and over again in a repeated fashion. There are many solutions for cases such as these and we recommend working with the vendor's technical staff to find the best solution for your needs.

Figure 4 below is the transfer function of a typical PID control block In this example, the sample rate is 1 kHz, the Red PID gains are 110, 10, and 1000 and the blue are 700, 20, and 3500. The red plot depicts the general shape of all PID transfer functions. In general, I gain affects the low frequency part of the curve, P gain the middle region, and D gain the high frequency region. All of the gains combine to affect the frequency of the inflection point and the shape near the inflection point. This is evident when comparing the blue and red curves between 1 and 100 Hz. You can see that the blue curve has a lower inflection point frequency and a softer shaped curve at the inflection point frequency. These subtle differences can play a huge role in servo loop performance. Also, it should be noted that the PID block should always be paired with a low pass filter. The blue trace shows the effect of the filter, where the gain above 100 Hz is not increasing. Not using a low pass filter with a PID block leaves the system very sensitive to any high frequency disturbances. It should be noted that there is no universal set of PID gains, as they depend on all of the other Loop Gains in the system.

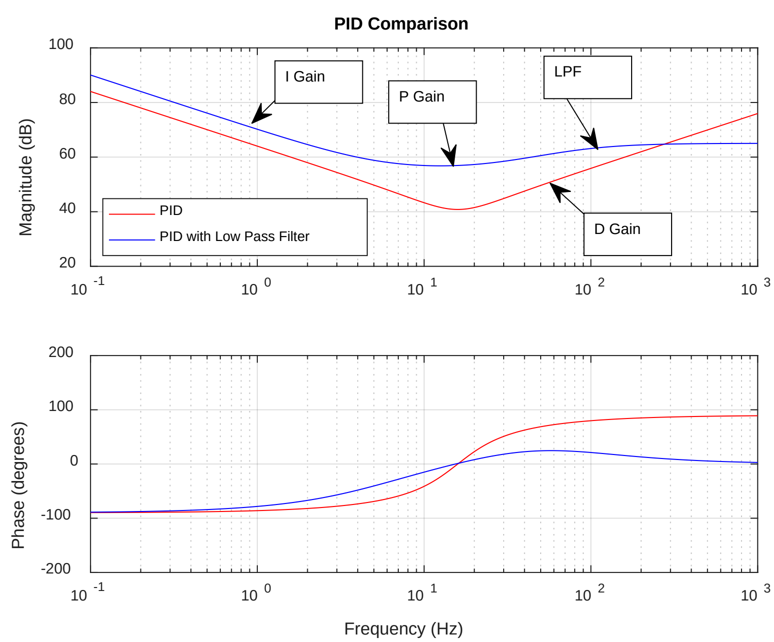

Figure 4. Transfer Function of a PID and PID with a Low Pass Filter

CLOSED LOOP TRANSFER FUNCTION

The Closed Loop Transfer Function (CLTF) is the most common way to describe the control system in the frequency domain. The data is taken while the loop is closed, so all of the closed loop dynamics are captured in this transfer function. The gain of the CLTF is very close to unity at low frequency and rolls off at high frequency, with some amplification in between. The phase of the CLTF is near 0 at low frequency and near -180 degrees at high frequency. This curve is typically used to state the system's "Bandwidth", a term we will describe later. A simulated CLTF is show in Figure 5 below.

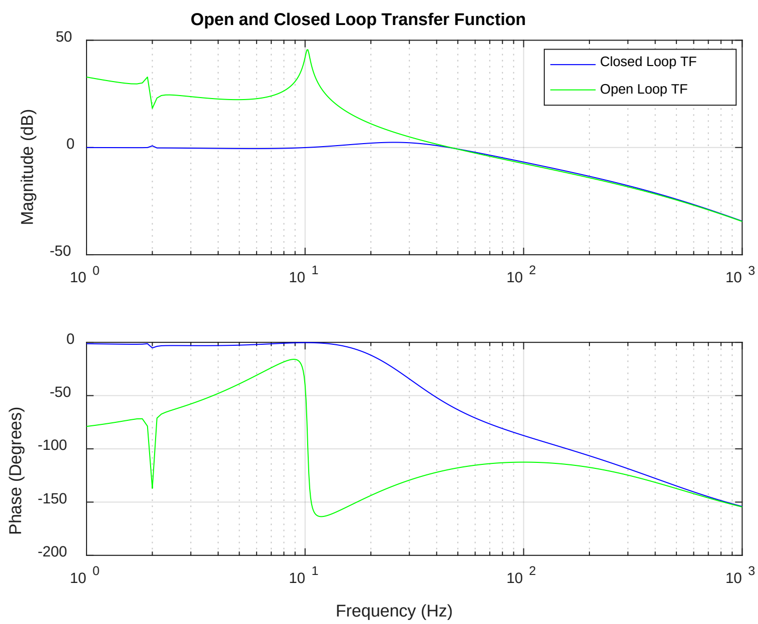

Figure 5. Closed Loop Transfer Function Example

OPEN LOOP (LOOP) TRANSFER FUNCTION

The Open Loop, or "Loop" transfer function (OLTF) is representative of all the frequency dependent blocks that make up of the servo loop, meaning the control, the drive, the plant, and the sensor. The OLTF for some systems can be taken with the servo loop open, but in many cases it must be either derived from the Closed Loop, or taken with the loop closed, and done so using the test point shown in Figure 1. The OLTF can be found from the Closed Loop using the formula CLTF/(1-CLTF), where CLTF is the complex representation of the closed loop transfer function. Referring back to Figure 1, the CLTF can be measured by driving the system with random or swept sine at In2 and measuring Out2/In2. The

OLTF can be found by driving with In2, but measuring Out2/Out3. The open Loop and closed loop transfer function are shown in Figure 5 above.

BANDWIDTH

"Bandwidth" is often proceeded by the word "Servo". "Servo Bandwidth" or "Bandwidth" describes the maximum frequency at which the control system will exert beneficial effort into a control system. Electrical engineers who deal with low pass filter types such as Butterworth, Elliptical, Chebyshev, etc., will often quote the -3 dB magnitude frequency as the filter's bandwidth. The equivalent for servo systems would be the -3 dB point of the Closed Loop Transfer Function. This definition, however, is not universal. Many control engineers will refer to the 0 dB of the Open Loop (or Loop) Transfer Function as the Control System's Bandwidth. This tech note will not dive into the details of each, rather, we want the reader to: a) be clear by what you mean when you use the word. Are you quoting the -3 dB point in the Closed Loop, or 0 dB in the Open Loop?; and b) ask what others mean when they use the word so there is no miscommunication that results in over-designing a control system, or having one not meet its intended performance.

In control textbooks bandwidth is described as a number that characterizes a system's response time. BW = 0.35 * TR is the most commonly quoted formula, but is also the most commonly mis-used formula when describing closed loop mechanical systems. This formula is for 1st order systems, and closed loop motion control systems do not act like first order systems. Using the above formula can absolutely steer one in the wrong direction. There are many variations of the bandwidth formula that take into account the system's damping ratio "zeta". Zeta for a closed loop system is related to the phase margin of the system. The formula relating Phase Margin and zeta is discussed in the Phase Margin section later in this tech note.

It is important to note, however, almost all of the formulas you will find equating bandwidth and response time for second order systems are based on linear systems that are not in saturation. If one compares response times from actual step response data to the calculations made via bandwidth formulas, they will rarely be equivalent and the most likely reason is from voltage or current saturation. Drive systems have a limited amount of voltage or current, and if the step values chosen, in combination with the proportional and integral gain are too large, voltage or current saturation will occur, slowing down the system's response versus the formula's predicted response time.

Bandwidth is also frequently used incorrectly to state the control system's ability to track, reject, or attenuate error. The most common misconception is that the higher the bandwidth the better the performance. For tracking, attenuating, or rejecting disturbances, "Loop Gain" is what matters, not Bandwidth. Loop Gain is frequency dependent. Two systems can have the same bandwidth, but very different low frequency gain. Both systems will have the same mathematically calculated response time, but one will track, or reject low frequency disturbances more so than the other.

Related to this concept is understanding the limitations of disturbance rejection (or tracking) as the frequency nears the bandwidth frequency. If we use the 0 dB cross over in the OLTF as our bandwidth frequency, this means that the Loop Gain just before the bandwidth is above 0 dB, but is also very small, say just a few dB. As stated above, Loop Gain is what drives disturbance rejection, and if Loop Gain is low, so is your disturbance rejection or tracking ability. This means is that one cannot expect the same performance at frequencies just before the bandwidth as compared to frequencies that are more than 2x lower than the bandwidth frequency. This is a key concept to understand when writing and responding to control system's requirements.

GAIN MARGIN\PHASE MARGIN

Gain Margin and Phase Margin are terms used to describe the stability of a servo loop. Gain Margin is the OLTF gain at the frequency when the Phase of the OLTF passes through -180 degrees. Conversely, Phase margin is the phase difference between -180 and the OLTF phase when the gain crosses 0 dB (see Figure 6 below). A good rule of thumb for mechanical servo loops is to have at least 6 dB Gain Margin and 30 degrees of Phase Margin.

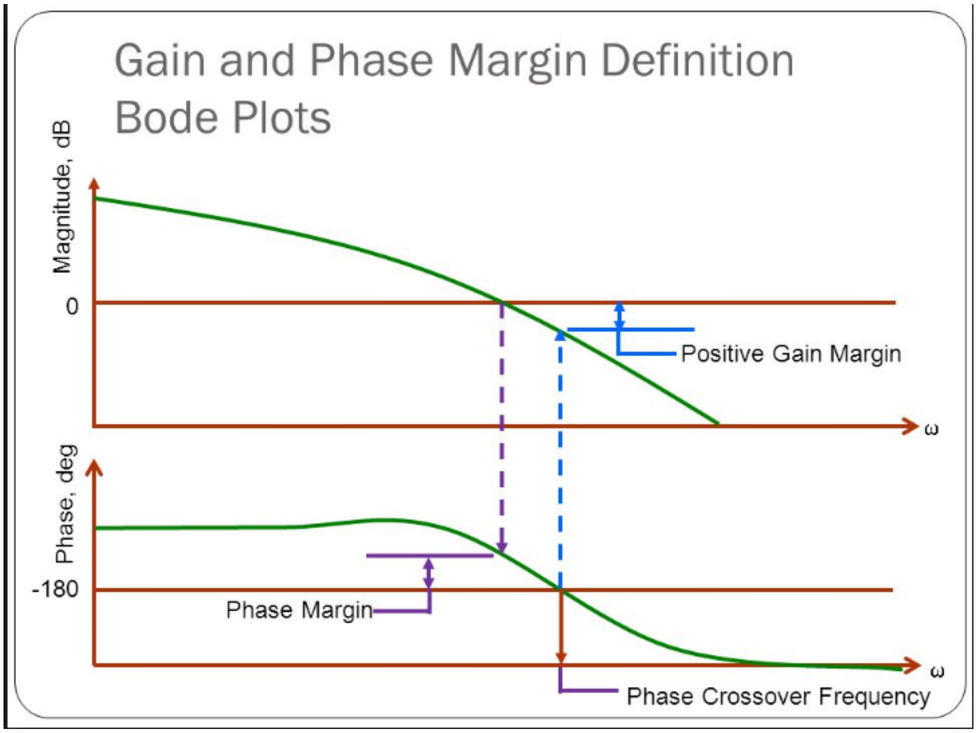

Figure 6. Physical model

Professor Walter W. Olson Department of Mechanical, Industrial and Manufacturing Engineering University of Toledo Stability Margins

A common misconception is that Phase Margin needs to be greater than 45 degrees. In most cases, using 30 degrees will keep your servo loop very stable and well behaved. Another useful strategy for Gain Margin is to limit any resonant peak in the OLTF gain plot, above the phase -180 degree frequency, to have at least 6 dB of Gain Margin. This constraint, although conservative, eliminates the need to determine if large amplitude peaks are phase stable. By assuming the peaks are not phase stable, it forces the control algorithm to reduce the peaks using filters or by adjusting the D gain of the PID. In many applications, having Gain and Phase Margin data at every location of the load is not possible. To counter this limitation, having a conservative approach when dealing with high frequency dynamics, that are often position dependent in terms of magnitude and phase, is a great countermeasure.

Peaking in the Closed Loop Transfer function described earlier is governed by the Phase Margin. Phase Margin and the damping ratio zeta are related through equation below, but can be simplified as PM = 100 * Zeta for damping ratios less than 1.0 (see Figure 7 below). Figure 8 shows how the shape of the CLTF changes for different amounts of Phase Margin. It should be noted that requiring no peaking in the CLTF is a very stringent requirement and one that will likely limit servo limit bandwidth dramatically. The peaking shown below in Blue and in Green are typical of mechanical servo systems. As the Phase Margin decreases towards the Red case, the resonance (or peak) may start to appear as a dominant peak in the performance variable.

$$Phase\ Margini\ (degrees) = \ tan^{-1}\frac{2*\zeta}{\sqrt{-2*\zeta^2+\sqrt{1+4*\zeta^4}}}$$

higheredbcs.wiley.com/legacy/college/nise/0471794759/justask/SNT10060.html

$$where\ \zeta \text{ is the damping ratio and closed loop amplification } Q = \frac{1}{2*\zeta}$$

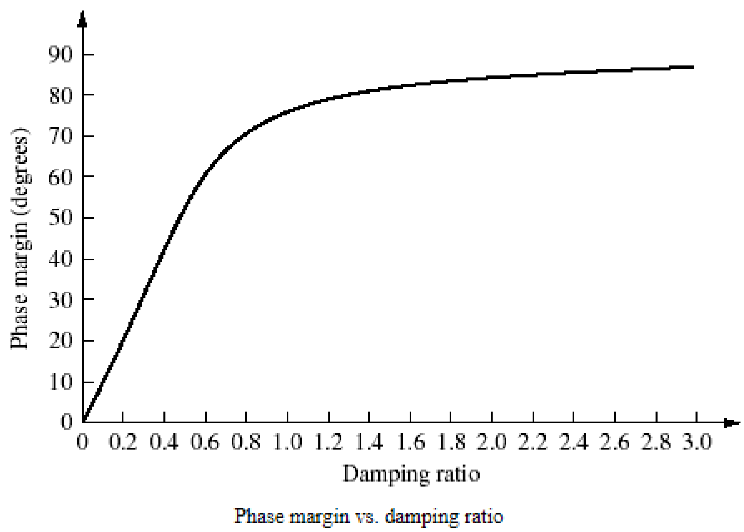

Figure 7. Phase Margin vs. Damping Ratio

higheredbcs.wiley.com/legacy/college/nise/0471794759/justask/SNT10060.html

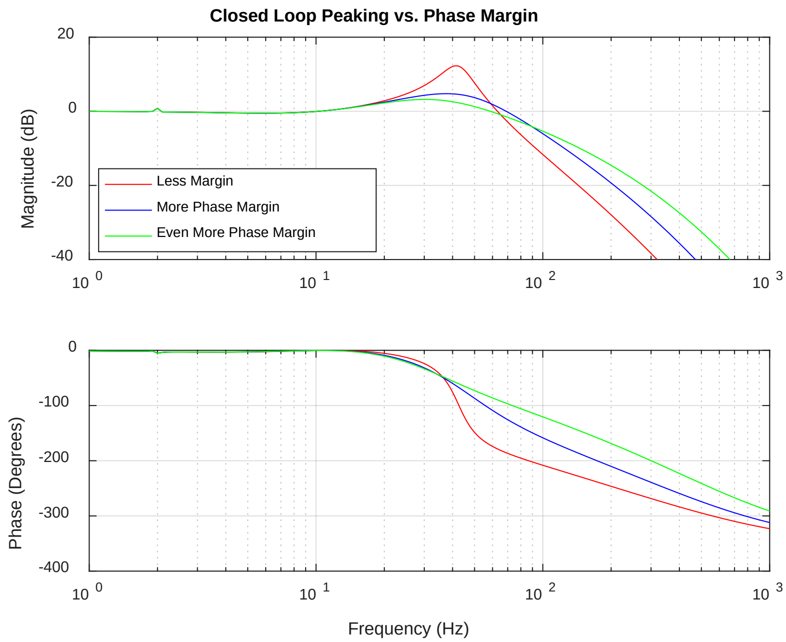

Figure 7. Closed Loop Peaking vs. Phase Margin

INCORRECTLY USED or MISUNDERSTOOD TERMINOLOGY

When servo terms are incorrectly bundled together, confusion can quickly occur. Here are some good examples of servo terms used incorrectly.

- "The servo rate is 2 kHz". In this example the user is mixing up two key elements of the servo loop, the servo "sample rate" and the "servo bandwidth". These are two very different terms. The servo sample rate (Fs), sometime called the "update rate" is the rate at which the controller samples the feedback sensor and outputs a control action. Typical sample rate will vary between 1 and 10 kHz. As described previously, the servo bandwidth is the maximum frequency at which the control system will exert beneficial effort into a control system

- "The system will oscillate at the motor frequency". In this example, the term "motor frequency" is really meant to be "servo bandwidth" frequency. Motors do not have a frequency, servo loops have a frequency.

- "The motor resonance is 100 Hz". Motors typically do not have "resonant" frequency per say*, but the servo loop does have a servo bandwidth frequency and also is likely to have plant resonant frequencies. In this example it is not clear if 100 Hz is the bandwidth frequency or a plant resonant frequency. A more clear statement would be, "The servo bandwidth is 100 Hz" or "The Plant has its first resonance at 100 Hz". (*) The housing supporting the motor can have a resonant frequency. The back iron and magnet track for a linear motor can have a resonant frequency, and the coil for a linear motor can have resonant frequency. Typically these resonances (or modes of vibration) are perpendicular to direct which force is created and are not strong contributors in the Plant Transfer Function.

- "My servo loop bandwidth is 100 Hz, but I am not getting good attenuation of the disturbance frequency at 90 Hz". In this case the person has used the term "servo bandwidth" correctly, but is incorrect in the expectation of how much disturbance rejection is achievable for a given servo bandwidth. Disturbance rejection (or tracking) is 100% dependent on Loop Gain, and Loop Gain near the bandwidth is typically low, so it is expected that limited performance will be achieved. The take away here is that "Bandwidth" is not a good indicator of a servo loop's ability to track or reject disturbances. Loop Gain is a much more direct value to specify rejection or tracking performance. Bandwidth does tell how fast a servo can respond, but in most cases the speed of the response is limited by saturation of current or voltage.

- "Will the bandwidth double if I doubled the encoder resolution, going from 1 count = 0.5 um to 1 count = 0.25 um"? This is a commonly asked question, and the answer is "No". The reason is that the max Loop Gain is always conserved and is limited by stability. This means that components within the loop can change their gains up or down, but stability rules govern the max loop gain. In this example, if the encoder gain doubles, another component within the loop must be cut in half to conserve the overall loop gain value. Also, in this question, the user is mixing bandwidth and loop gain. Bandwidth goes up and down with a change in the overall Loop Gain, but the ratio is not necessarily 1 to 1. Loop Gain is 1 to 1 with each of its components, so if the encoder gain were doubled, and no other changes made, the Loop Gain would also double.

- "If higher resolution encoders do not provide higher bandwidth, what is the benefit"? The answer is repeatability (not to be confused with accuracy). The servo will move the load into position with +/- 1 count of error (assuming some amount of integral gain). The higher resolution encoder, the more precise the final position will be. In this case, accuracy is not improved with higher resolution because it is governed by the runout of the encoder scale markings and other factors, and not based on how the makings are interpolated.

CONCLUSION

Servo systems have a vocabulary all of their own. Terms such as Bandwidth, Transfer Function, Plant, or Gain Margin may not have ever been described to a mechanical or an electrical engineer tasked with delivering an electro-mechanical servo system. This tech note has described some of the basic servo terminology, and given examples of common misuses and common misconceptions of servo terms. Understanding the content of this tech note should help engineers navigate their way through the servo component selection process.