Parasitic Effects

- Parasitic effects are present in all real-world systems and are troublesome to account for when the systems are designed. They are rarely disabling, but are debilitating if not dealt with effectively.
- These effects include:
  - Coulomb Friction
  - Time Delay
  - Unmodeled Resonances
  - Saturation
  - Backlash

- Questions:
  - Are they significant?
  - What to do about them?
- Approaches:
  - Ignore them and hope for the best! Murphy's Law says ignore them at your own peril.
  - Include the parasitic effects that you think may be troublesome in the truth model of the plant and run simulations to determine if they are negligible.
  - If they are not negligible and can adversely affect your system, you need to do something – but what?

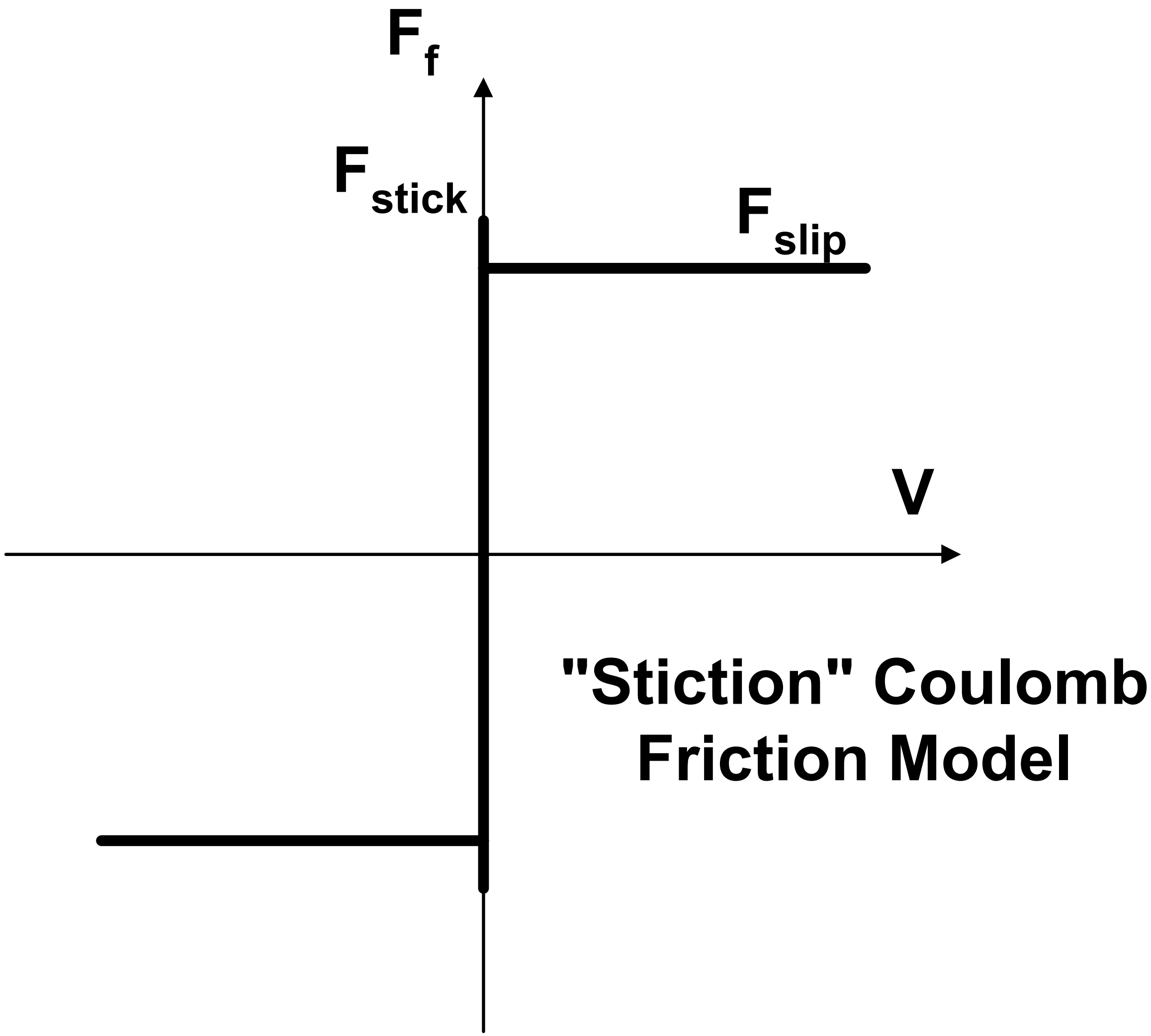
- General remedies include:
  - Alter the design to reduce the effective loop gain of the controller, especially at high frequencies where the effects of parasitics are often predominant. This generally entails sacrifice in performance.
  - Techniques specifically intended to enhance robustness of the design are also likely to be effective, but may entail use of a more complicated control algorithm.

Coulomb Friction: Modeling and Simulation

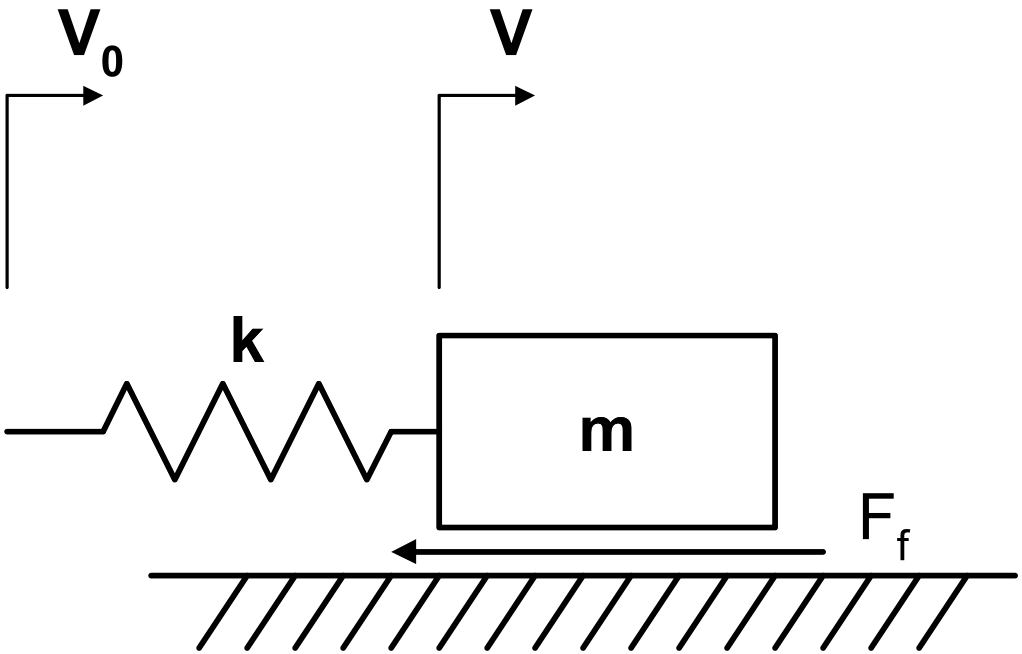
- In most control systems, Coulomb friction is a nuisance.
- Coulomb friction is difficult to model and troublesome to deal with in control system design.
- It is a nonlinear phenomenon in which a force is produced that tends to oppose the motion of bodies in contact in a mechanical system.
- Undesirable effects: "hangoff" and limit cycling
- *Hangoff* (or d-c limit cycle) prevents the steady-state error from becoming zero with a step command input.
- *Limit Cycling* is behavior in which the steady-state error oscillates or hunts about zero.

- What Should the Control Engineer Do?
  - Minimize friction as much as possible in the design
  - Appraise the effect of friction in a proposed control system design by simulation
  - If simulation predicts that the effect of friction is unacceptable, you must do something about it!
  - Remedies can include simply modifying the design parameters (gains), using integral control action, or using more complex measures such as estimating the friction and canceling its effect.
  - Modeling and simulation of friction should contribute significantly to improving the performance of motion control systems.

Modeling Coulomb Friction



Case Study to Evaluate Friction Models



$m = 0.1$ kg

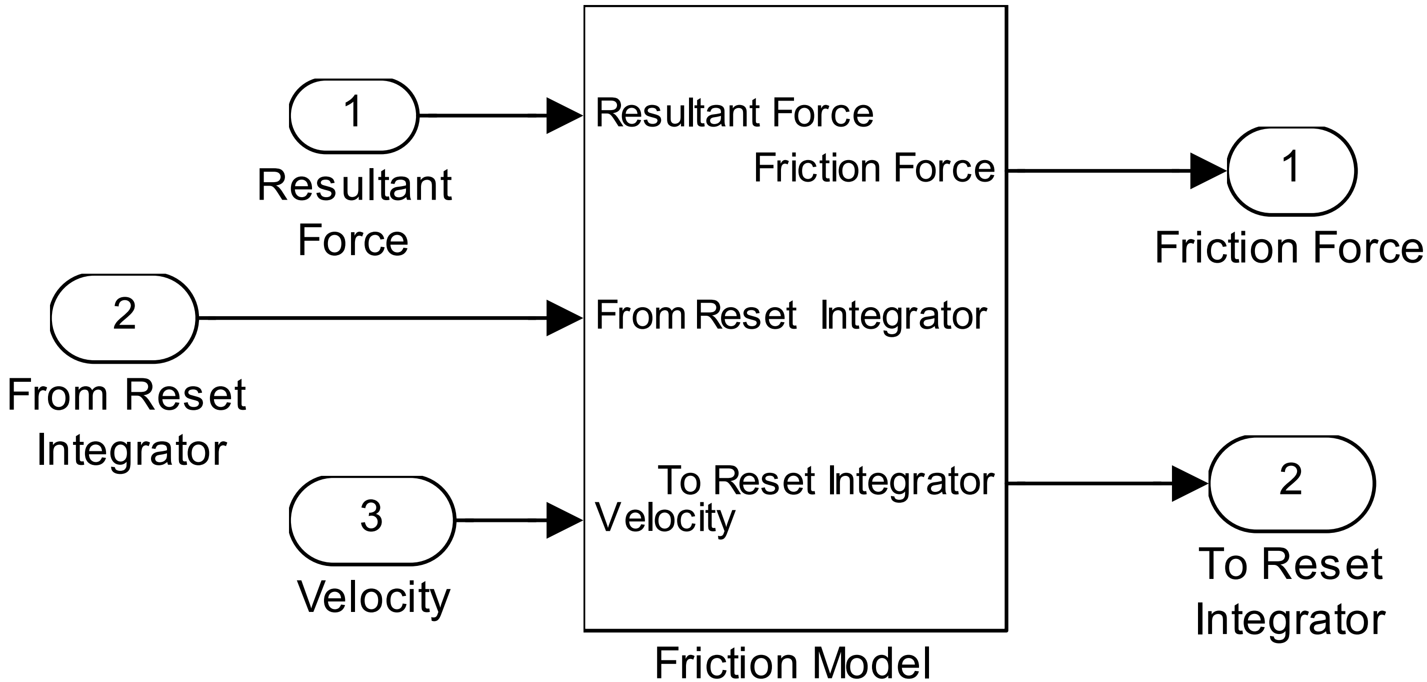
$k = 100$ N/m

$F_{stick} = 0.25$ N

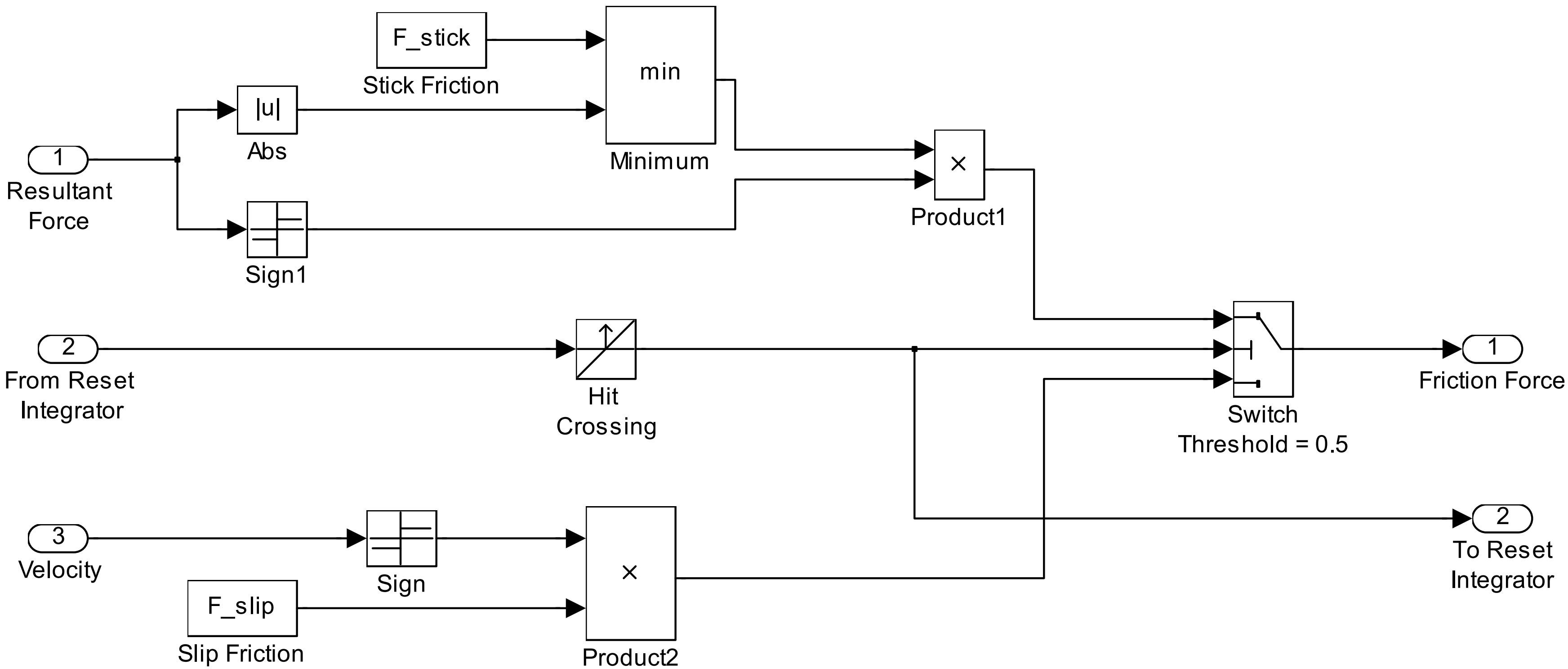
$F_{slip} = 0.20$ N (assumed independent of velocity)

V_0 = step of 0.002 m/sec at $t = 0$ sec

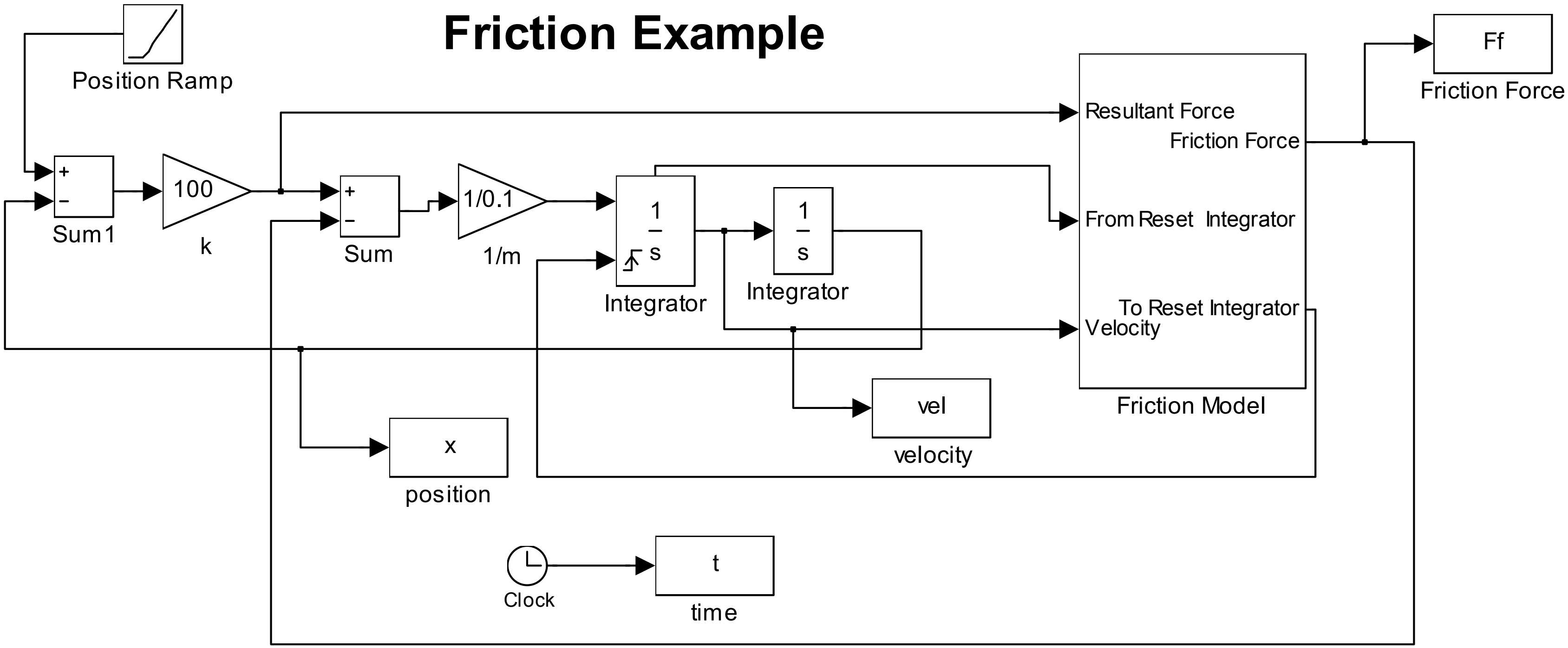
Friction Model in Simulink



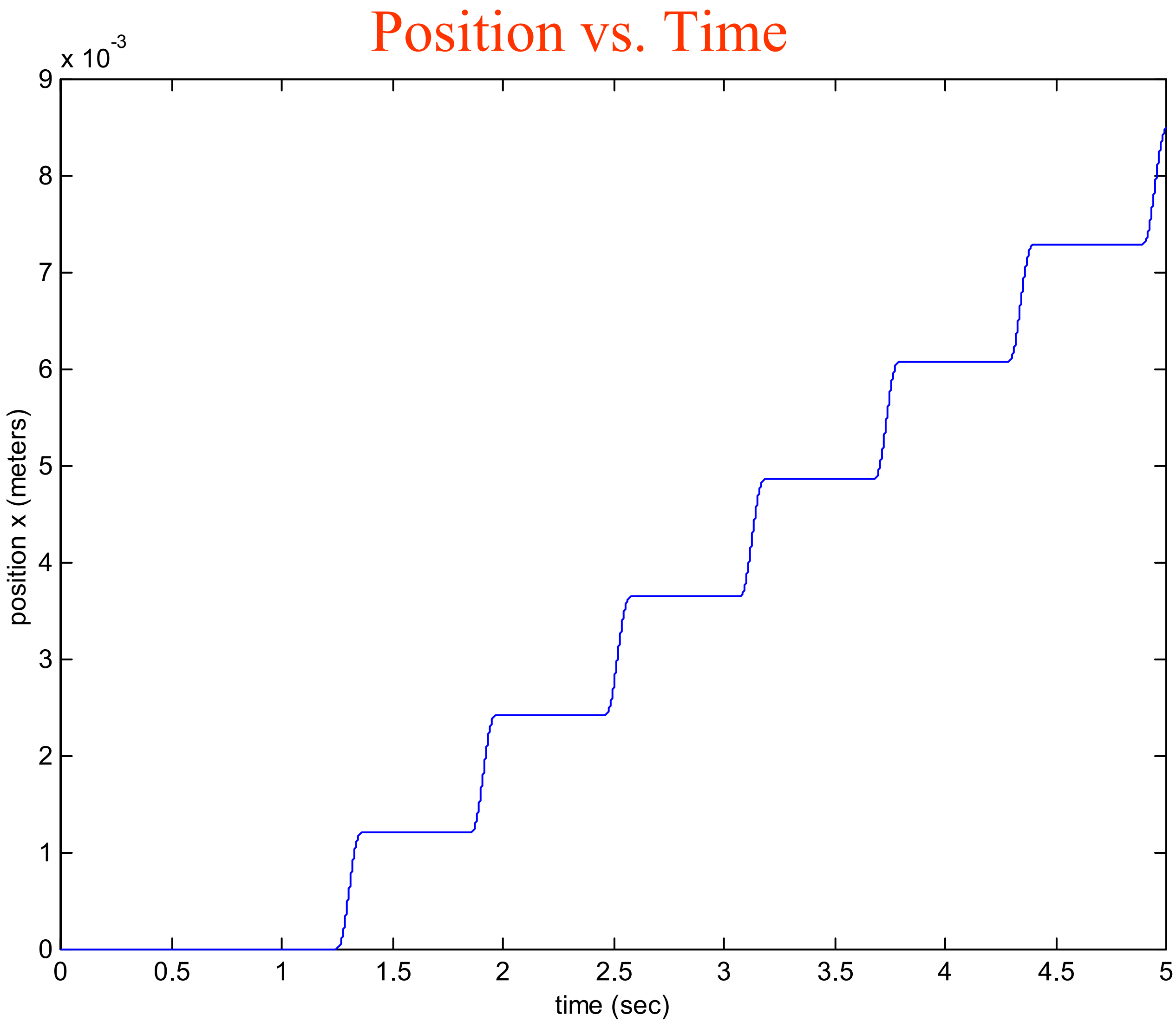
Simulink Block Diagram



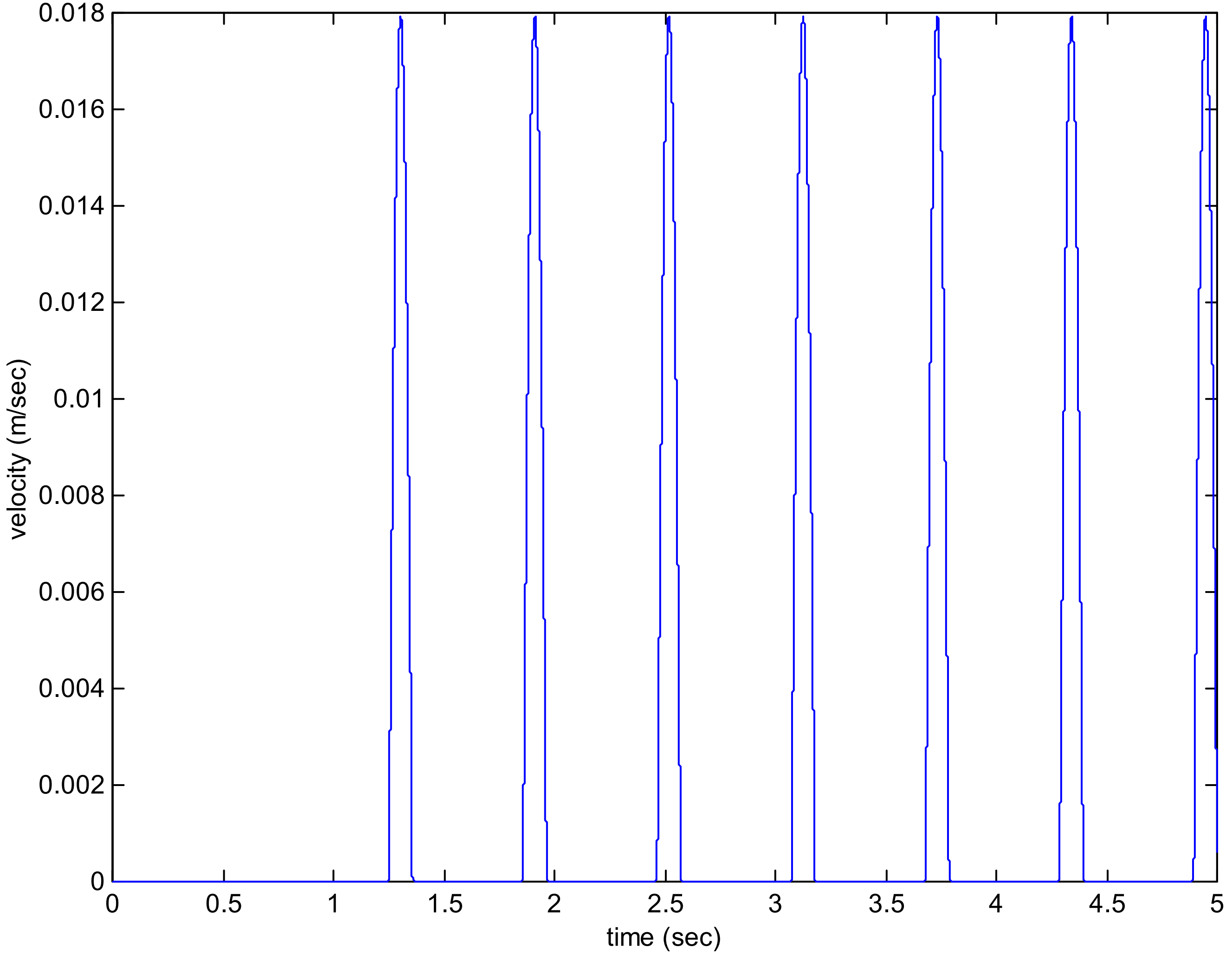
Example with Friction Model



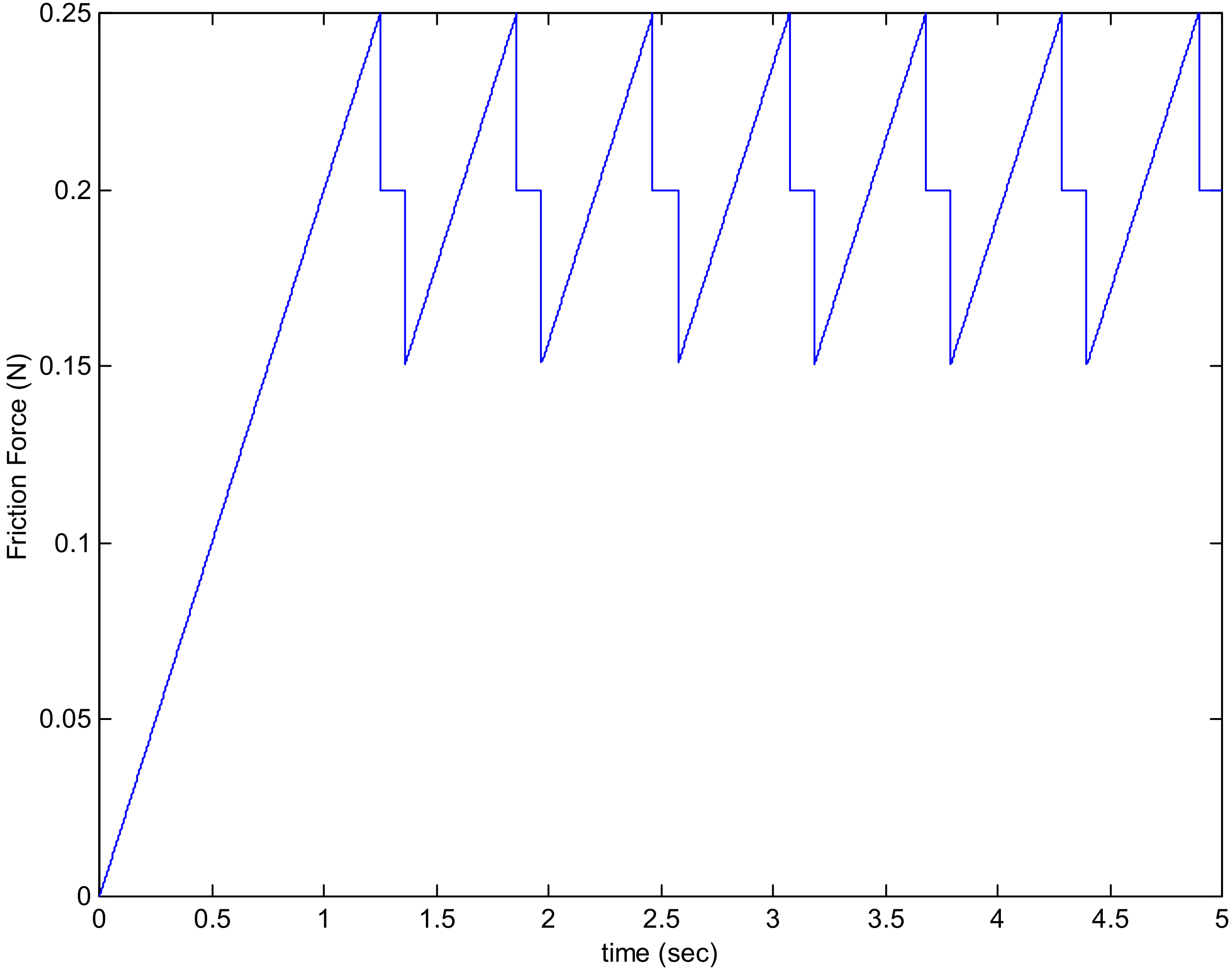
Model Simulation Results



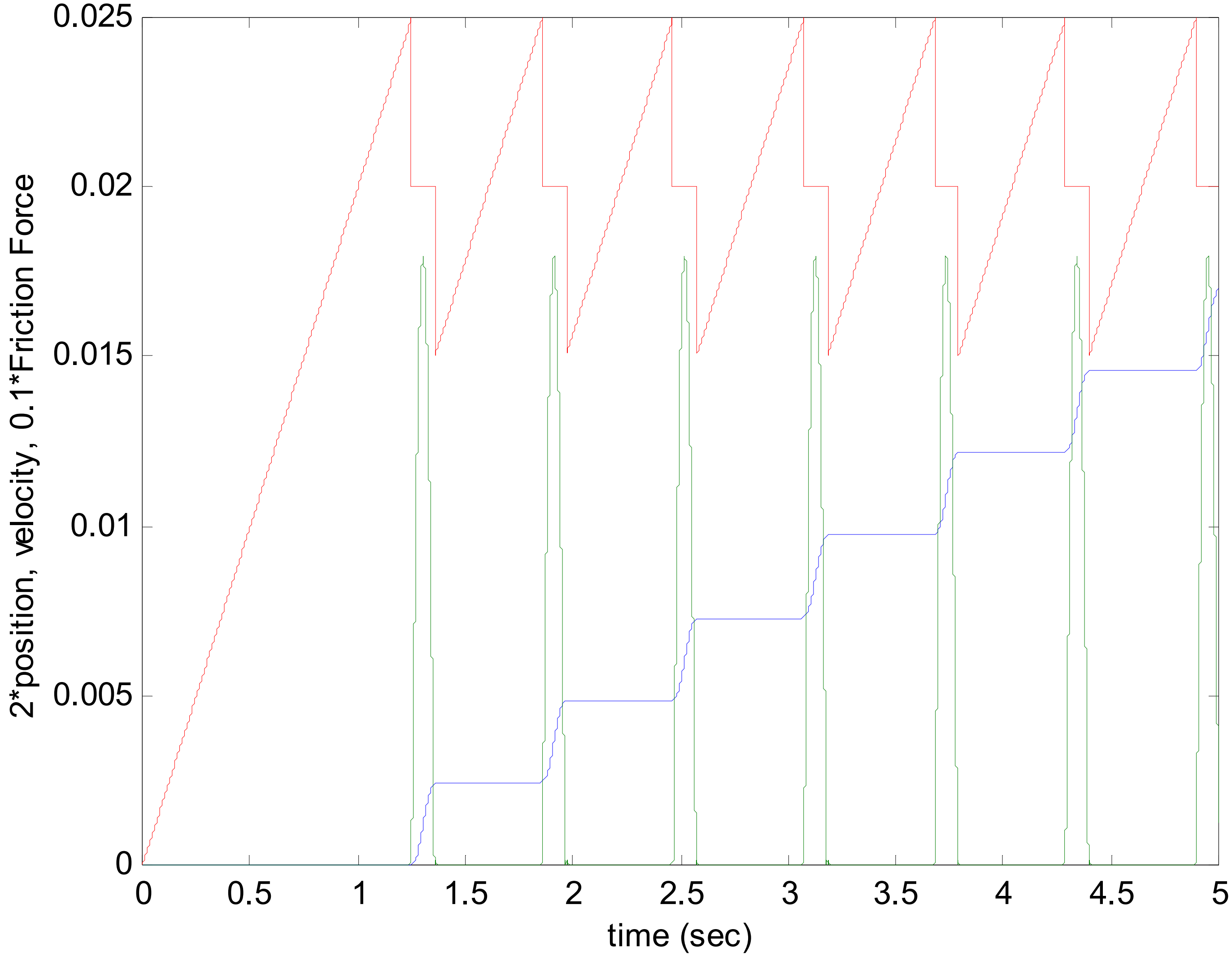
Velocity vs. Time



Friction Force vs. Time

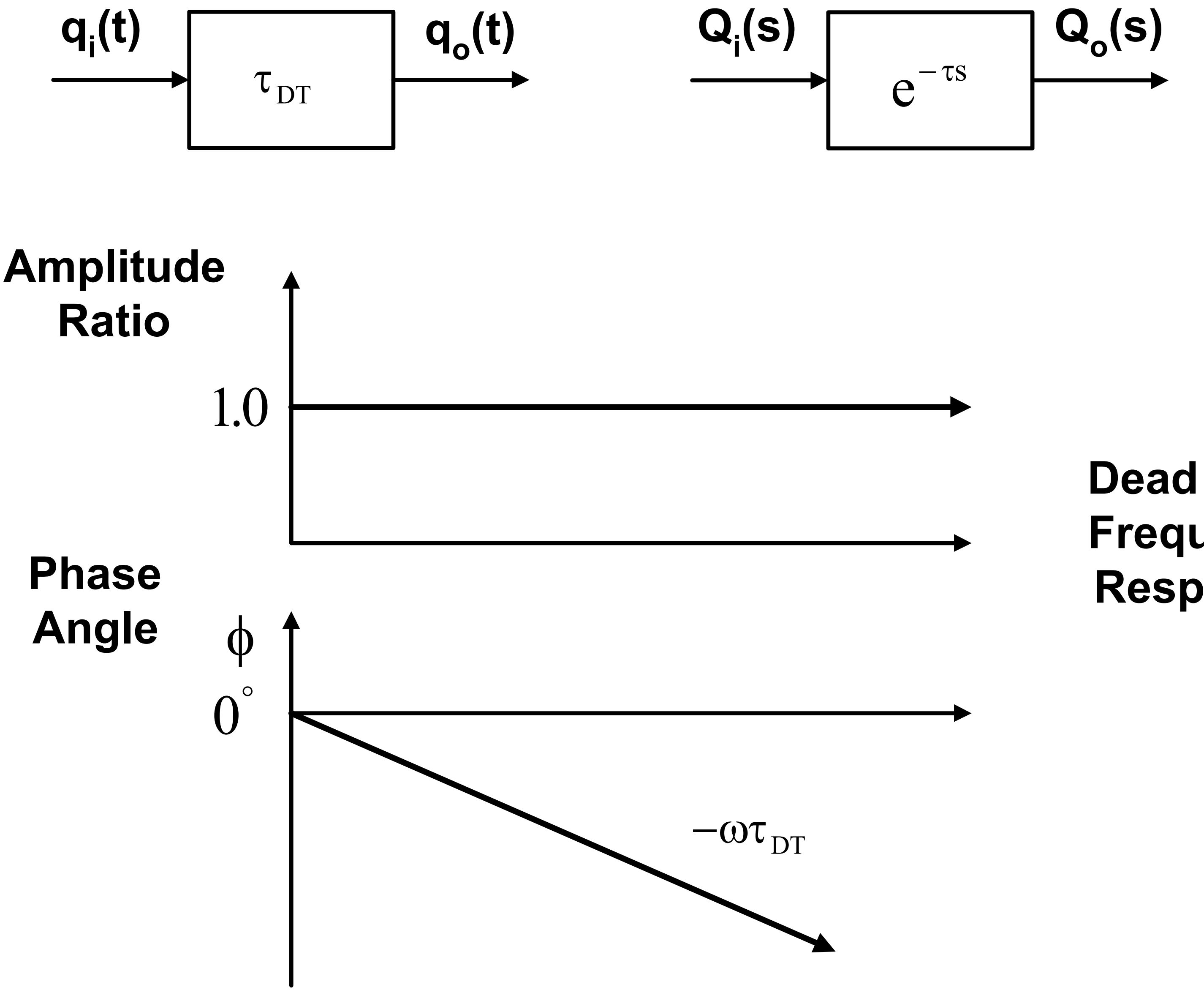


Position, Velocity, Friction Force vs. Time



Time Delay: Modeling, Simulation, and Compensation

- Time delays or dead-times (DT's) between inputs and outputs are very common in industrial processes, engineering systems, economical, and biological systems.
- Transportation and measurement lags, analysis times, computation and communication lags all introduce DT's into control loops.
- DT's are also used to compensate for model reduction where high-order systems are represented by low-order models with delays.
- Two major consequences:
  - Complicates the analysis and design of feedback control systems
  - Makes satisfactory control more difficult to achieve



Dead Time Frequency Response

- Systems with Time Delay
  - Any delay in measuring, in controller action, in actuator operation, in computer computation, and the like, is called *transport delay* or *dead time*, and it always reduces the stability of a system and limits the achievable response time of the system.

- Dead-Time Approximations
  - The simplest dead-time approximation can be obtained by taking the first two terms of the Taylor series expansion of the Laplace transfer function of a dead-time element, τ_{dt}.

$$q_i(t) = \text{ input to dead-time element}$$

$$q_o(t) = \text{ output of dead-time element}$$

$$= q_i\left(t-\tau_{dt}\right)u\left(t-\tau_{dt}\right)$$

$$u\left(t-\tau_{dt}\right)=1 \quad \text{for } t \geq \tau_{dt}$$

$$u\left(t-\tau_{dt}\right)=0 \quad \text{for } t < \tau_{dt}$$

$$L\left[f\left(t-a\right)u\left(t-a\right)\right]=e^{-as}F\left(s\right)$$

$$\frac{Q_o}{Q_i}\left(s\right)=e^{-\tau_{dt}s} \approx 1-\tau_{dt}s$$

$$q_o\left(t\right) \approx q_i\left(t\right)-\tau_{dt}\frac{dq_i}{dt}$$

– The accuracy of this approximation depends on the dead time being sufficiently small relative to the rate of change of the slope of $q_i(t)$. If $q_i(t)$ were a ramp (constant slope), the approximation would be perfect for any value of τ_{dt}. When the slope of $q_i(t)$ varies rapidly, only small τ_{dt}'s will give a good approximation.

– A frequency-response viewpoint gives a more general accuracy criterion; if the amplitude ratio and the phase of the approximation are sufficiently close to the exact frequency response curves of $e^{-\tau_{dt}s}$ for the range of frequencies present in $q_i(t)$, then the approximation is valid.

- The Pade approximants provide a family of approximations of increasing accuracy (and complexity):

$$e^{-\tau s} = \frac{e^{\frac{-\tau s}{2}}}{e^{\frac{\tau s}{2}}} \approx \frac{1 - \frac{\tau s}{2} + \frac{\tau^2 s^2}{8} + \cdots + \frac{\left(-\frac{\tau s}{2}\right)^k}{k!}}{1 + \frac{\tau s}{2} + \frac{\tau^2 s^2}{8} + \cdots + \frac{\left(\frac{\tau s}{2}\right)^k}{k!}}$$

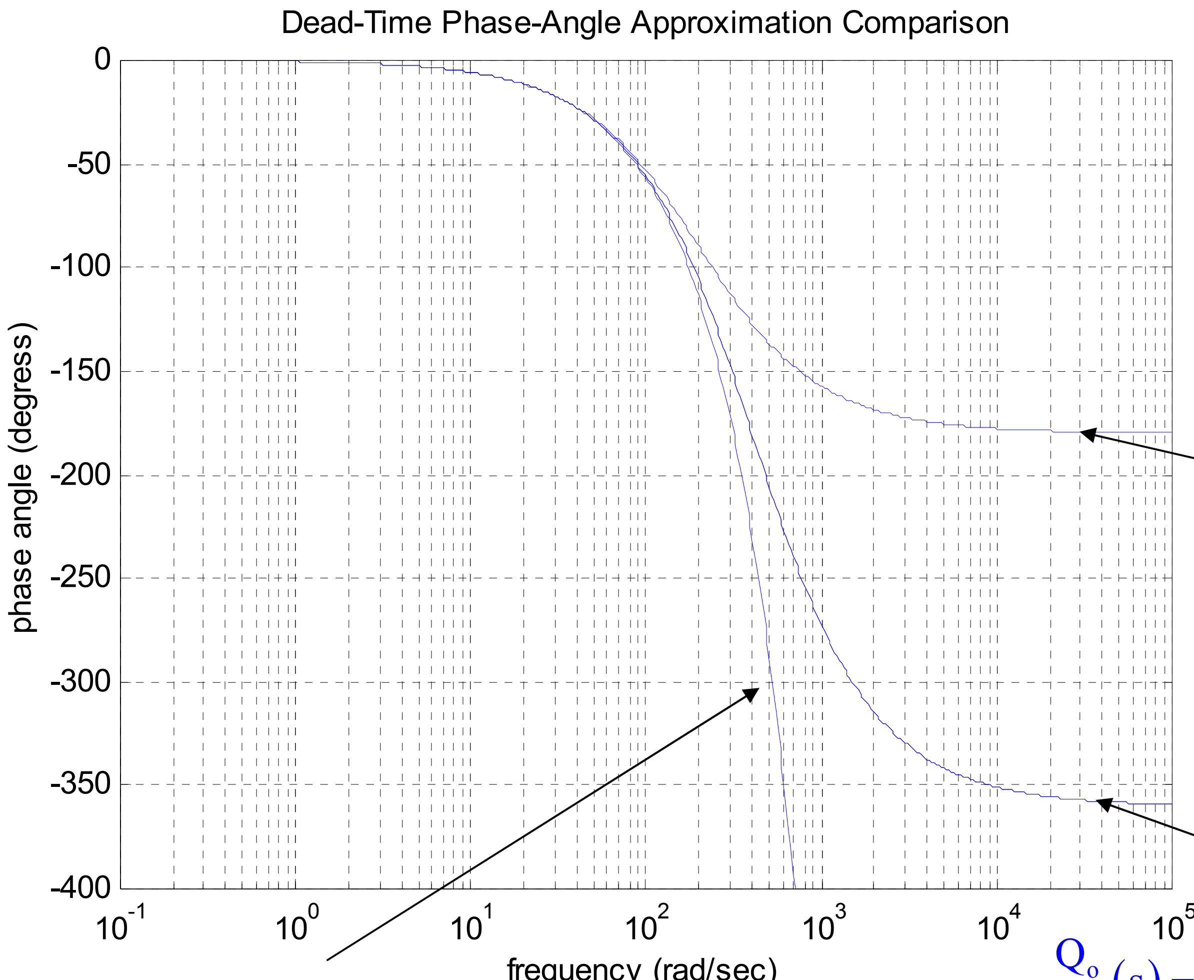
- In some cases, a very crude approximation given by a first-order lag is acceptable:

$$\frac{Q_o}{Q_i}(s) = e^{-\tau_{dt} s} \approx \frac{1}{\tau_{dt} s + 1}$$

- Pade Approximation:
  - Transfer function is all pass, i.e., the magnitude of the transfer function is 1 for all frequencies.
  - Transfer function is non-minimum phase, i.e., it has zeros in the right-half plane.
  - As the order of the approximation is increased, it approximates the low-frequency phase characteristic with increasing accuracy.
- Another approximation with the same properties:

$$e^{-\tau s} = \frac{e^{\frac{-\tau s}{2}}}{e^{\frac{\tau s}{2}}} \approx \frac{\left(1 - \frac{\tau s}{2k}\right)^k}{\left(1 + \frac{\tau s}{2k}\right)^k}$$

– Dead-time approximation comparison:



$$\tau_{dt} = 0.01$$

$$\frac{Q_o}{Q_i}(s) = \frac{2 - \tau_{dt}s}{2 + \tau_{dt}s}$$

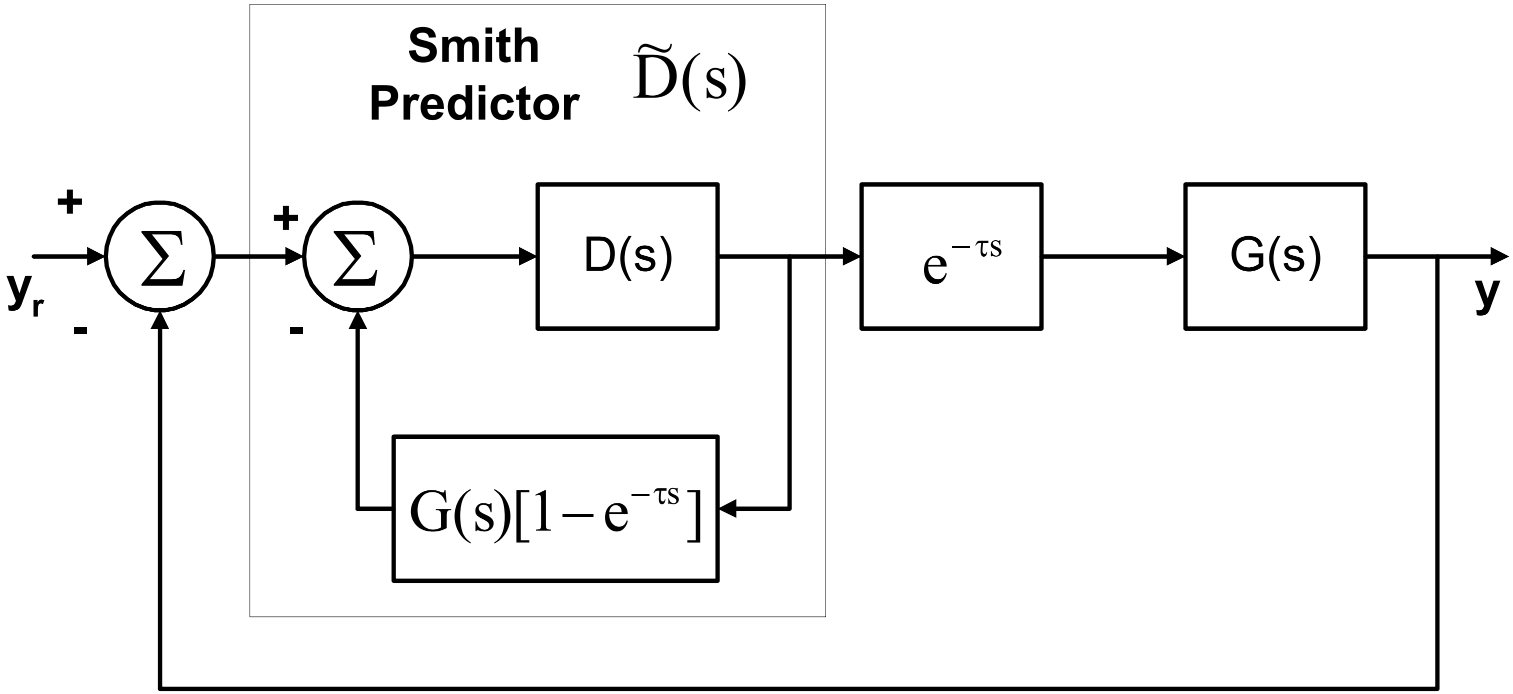
$$e^{-\tau_{dt}s} = 1\angle -\omega\tau_{dt}$$

$$\frac{Q_o}{Q_i}(s) = \frac{2 - \tau_{dt}s + \frac{(\tau_{dt}s)^2}{8}}{2 + \tau_{dt}s + \frac{(\tau_{dt}s)^2}{8}}$$

- Observations:
  - Instability in feedback control systems results from an imbalance between system dynamic lags and the strength of the corrective action.
  - When DT's are present in the control loop, controller gains have to be reduced to maintain stability.
  - The larger the DT is relative to the time scale of the dynamics of the process, the larger the reduction required.
  - The result is poor performance and sluggish responses.
  - Unbounded negative phase angle aggravates stability problems in feedback systems with DT's.

- The time delay increases the phase shift proportional to frequency, with the proportionality constant being equal to the time delay.
- The amplitude characteristic of the Bode plot is unaffected by a time delay.
- Time delay always decreases the phase margin of a system.
- Gain crossover frequency is unaffected by a time delay.
- Frequency-response methods treat dead times exactly.
- Differential equation methods require an approximation for the dead time.
- To avoid compromising performance of the closed-loop system, one must account for the time delay explicitly, e.g., Smith Predictor.

Smith Predictor



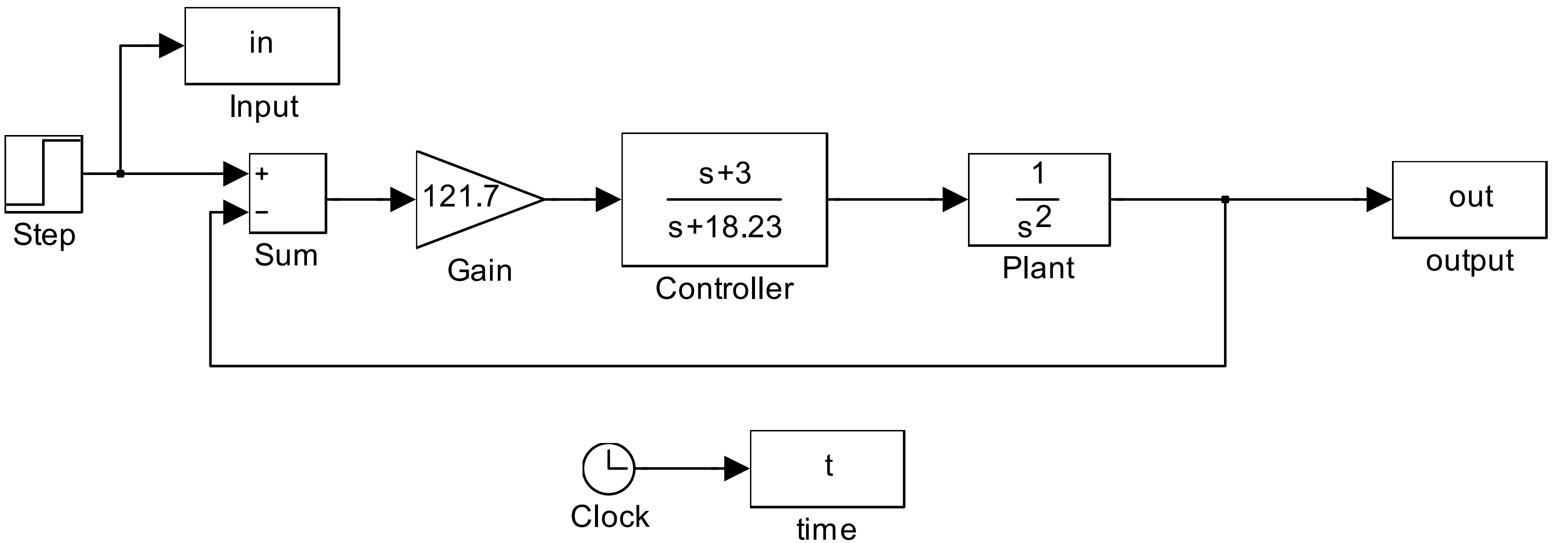
$$\tilde{D}(s) = \frac{D(s)}{1+(1-e^{-\tau s})D(s)G(s)} \qquad \frac{y}{y_r} = \frac{\tilde{D}(s)G(s)e^{-\tau s}}{1+\tilde{D}(s)G(s)e^{-\tau s}} = \frac{D(s)G(s)}{1+D(s)G(s)}e^{-\tau s}$$

- D(s) is a suitable compensator for a plant whose transfer function, in the absence of time delay, is G(s).
- With the compensator that uses the Smith Predictor, the closed-loop transfer function, except for the factor $e^{-\tau s}$, is the same as the transfer function of the closed-loop system for the plant without the time delay and with the compensator D(s).
- The time response of the closed-loop system with a compensator that uses a Smith Predictor will thus have the same shape as the response of the closed-loop system without the time delay compensated by D(s); the only difference is that the output will be delayed by τ seconds.

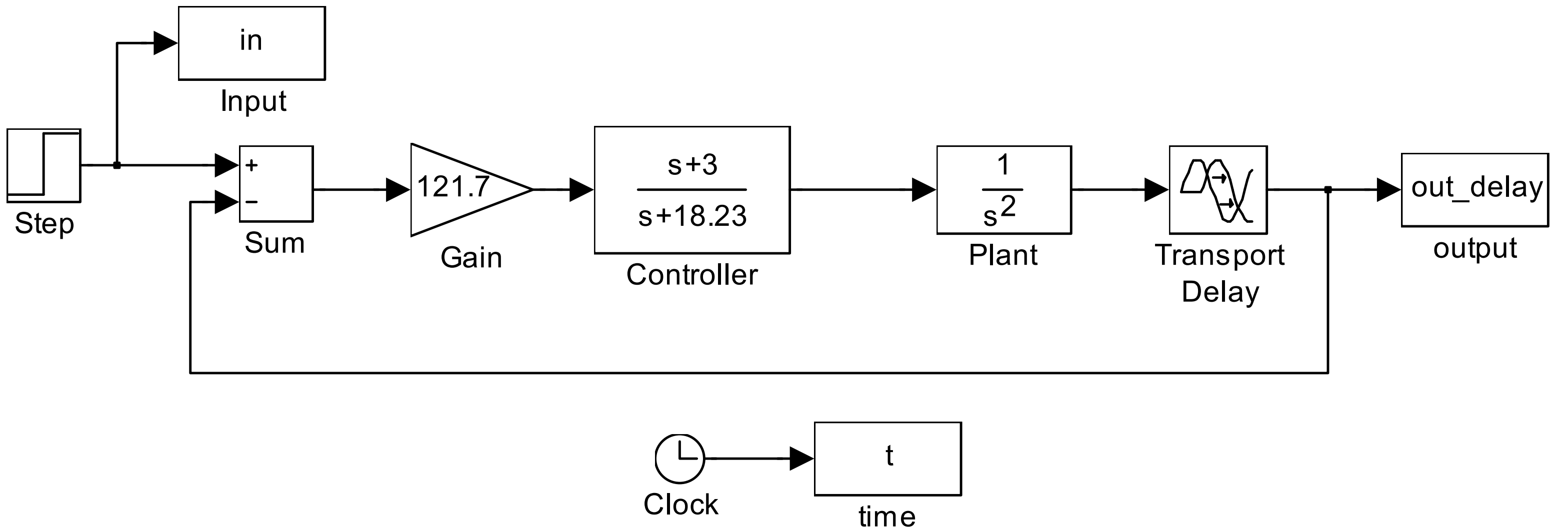
- Implementation Issues
  - You must know the plant transfer function and the time delay with reasonable accuracy.
  - You need a method of realizing the pure time delay that appears in the feedback loop, e.g., Pade approximation:

$$e^{-\tau s} = \frac{e^{\frac{-\tau s}{2}}}{e^{\frac{\tau s}{2}}} \approx \frac{1 - \frac{\tau s}{2} + \frac{\tau^2 s^2}{8} + \cdots + \frac{\left(-\frac{\tau s}{2}\right)^k}{k!}}{1 + \frac{\tau s}{2} + \frac{\tau^2 s^2}{8} + \cdots + \frac{\left(\frac{\tau s}{2}\right)^k}{k!}}$$

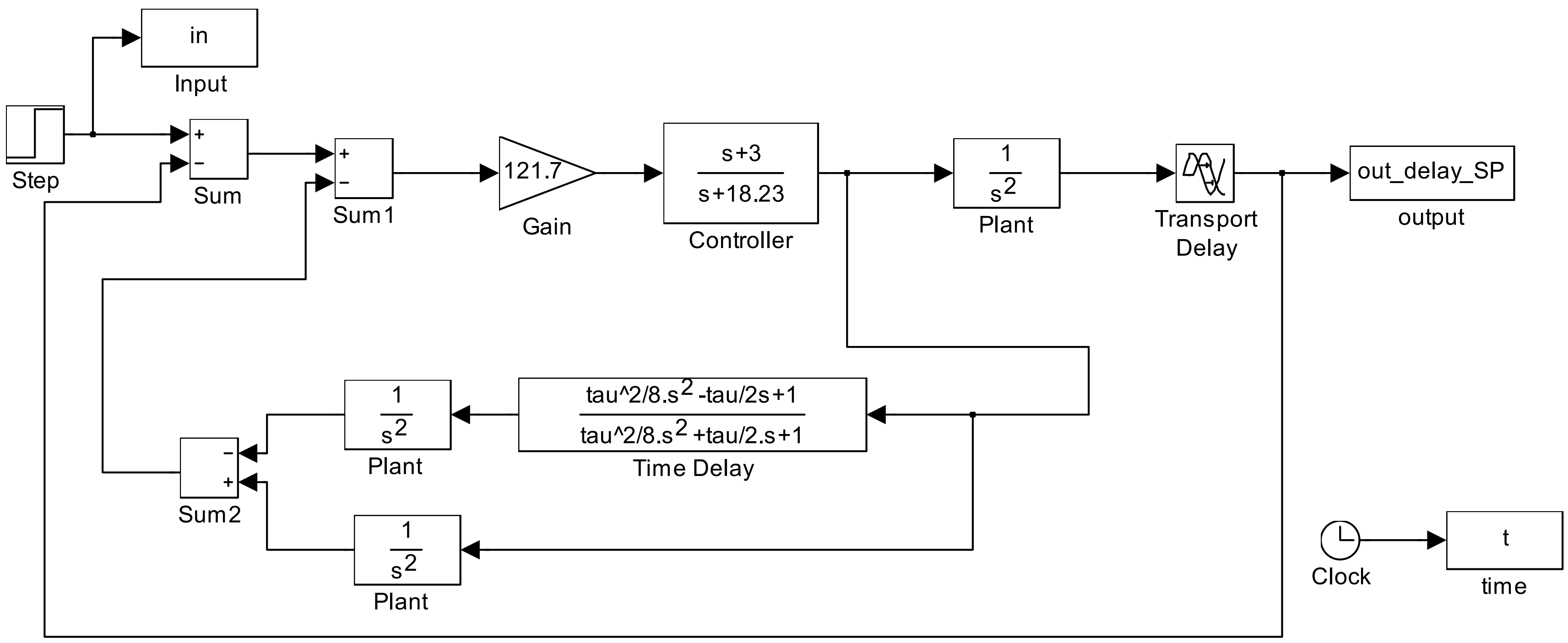
Example Problem



Basic Feedback Control System with Lead Compensator

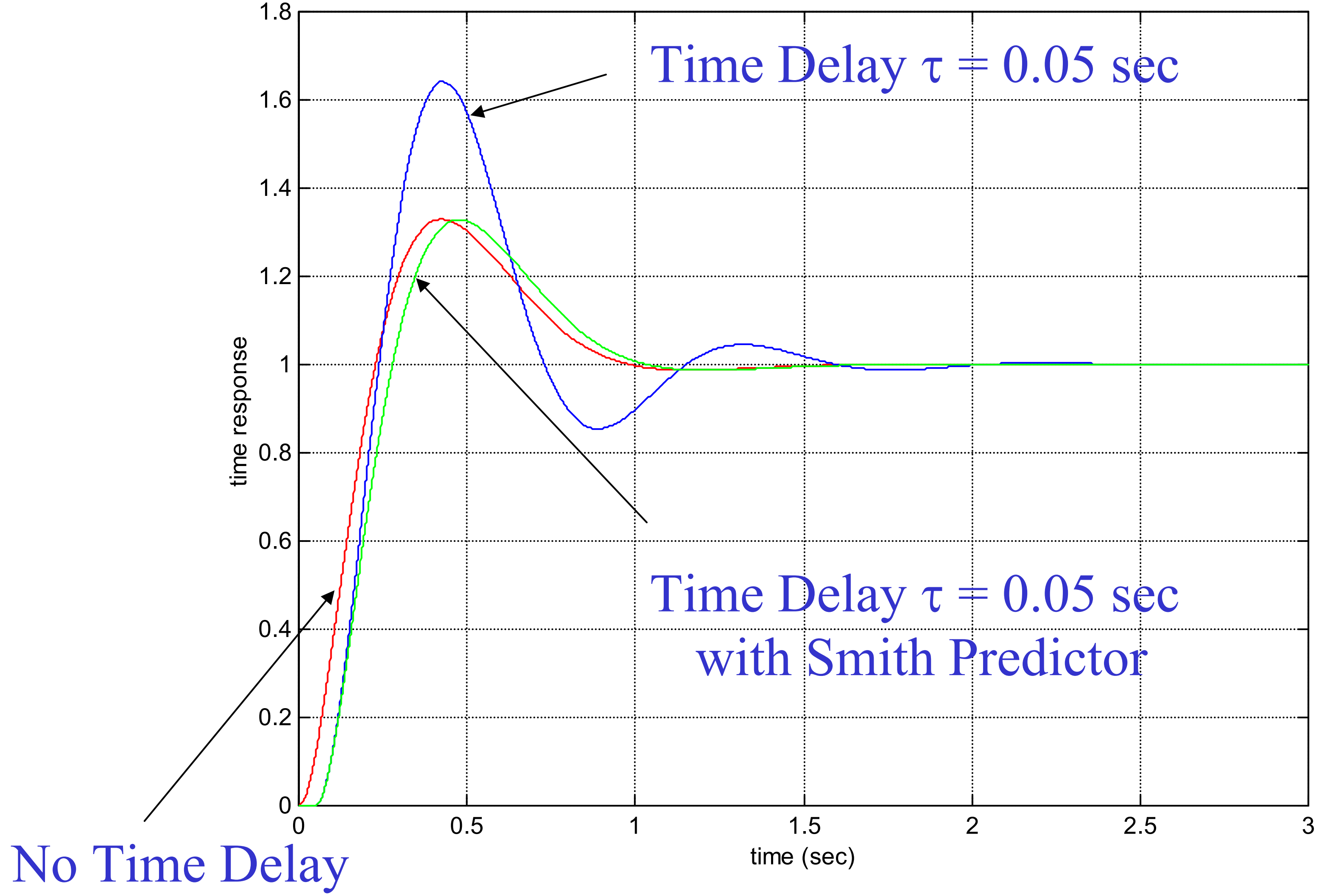


Basic Feedback Control System with Lead Compensator BUT with Time Delay $\tau = 0.05$ sec



Basic Feedback Control System with Lead Compensator
BUT with Time Delay $\tau = 0.05$ sec
AND Smith Predictor

System Step Responses



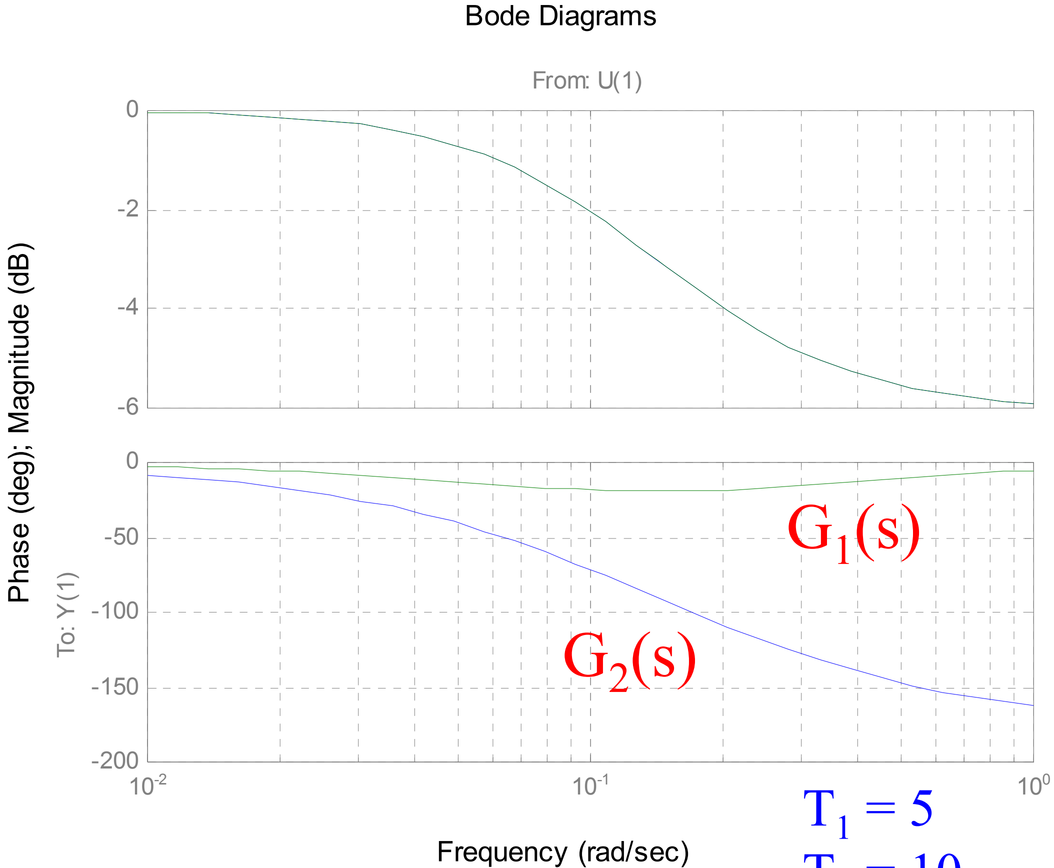
- Comments
  - The system with the Smith Predictor tracks reference variations with a time delay.
  - The Smith Predictor minimizes the effect of the DT on stability as model mismatching is bound to exist. This however still allows tighter control to be used.
  - What is the effect of a disturbance? If the disturbances are measurable, the regulation capabilities of the Smith Predictor can be improved by the addition of a feedforward controller.

- Minimum-Phase and Nonminimum-Phase Systems
  - Transfer functions having *neither* poles nor zeros in the RHP are *minimum-phase* transfer functions.
  - Transfer functions having *either* poles or zeros in the RHP are *nonminimum-phase* transfer functions.
  - For systems with the same magnitude characteristic, the range in phase angle of the minimum-phase transfer function is minimum among all such systems, while the range in phase angle of any nonminimum-phase transfer function is greater than this minimum.
  - For a minimum-phase system, the transfer function can be uniquely determined from the magnitude curve alone. For a nonminimum-phase system, this is not the case.

– Consider as an example the following two systems:

$$G_1(s) = \frac{1+T_1 s}{1+T_2 s} \qquad G_2(s) = \frac{1-T_1 s}{1+T_2 s} \qquad 0 < T_1 < T_2$$

A small amount of change in magnitude produces a small amount of change in the phase of $G_1(s)$ but a much larger change in the phase of $G_2(s)$.



$T_1 = 5$

$T_2 = 10$

- These two systems have the same magnitude characteristics, but they have different phase-angle characteristics.
- The two systems differ from each other by the factor:

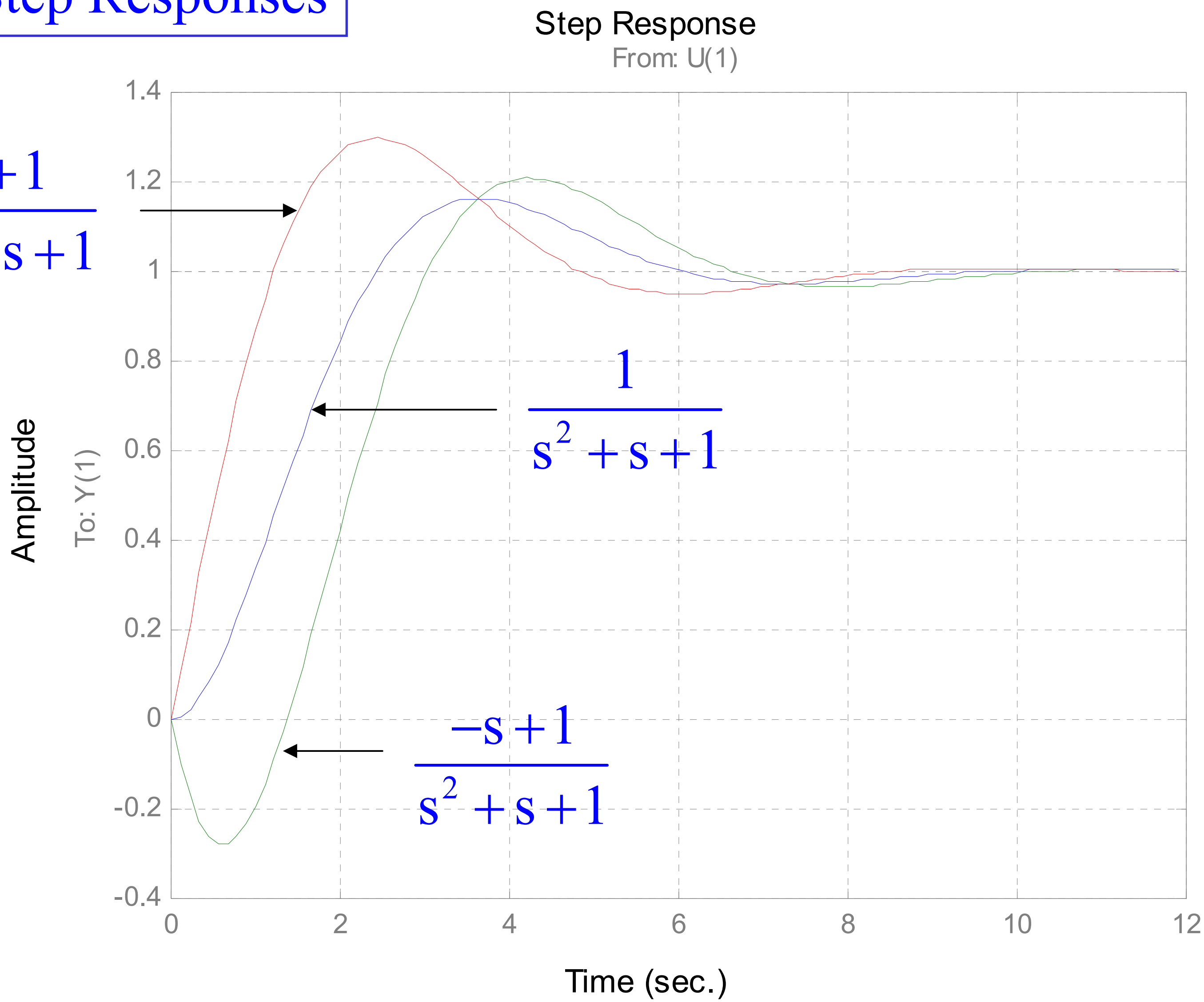
$$G(s) = \frac{1 - T_1 s}{1 + T_1 s}$$

- This factor has a magnitude of unity and a phase angle that varies from 0° to -180° as ω is increased from 0 to ∞.
- For the stable minimum-phase system, the magnitude and phase-angle characteristics are uniquely related. This means that if the magnitude curve is specified over the entire frequency range from zero to infinity, then the phase-angle curve is uniquely determined, and vice versa. This is called Bode's Gain-Phase relationship.

- This does not hold for a nonminimum-phase system.
- Nonminimum-phase systems may arise in two different ways:
  - When a system includes a nonminimum-phase element or elements
  - When there is an unstable minor loop
- For a minimum-phase system, the phase angle at $\omega = \infty$ becomes $-90°(q - p)$, where p and q are the degrees of the numerator and denominator polynomials of the transfer function, respectively.
- For a nonminimum-phase system, the phase angle at $\omega = \infty$ differs from $-90°(q - p)$.
- In either system, the slope of the log magnitude curve at $\omega = \infty$ is equal to $-20(q - p)$ dB/decade.

- It is therefore possible to detect whether a system is minimum phase by examining both the slope of the high-frequency asymptote of the log-magnitude curve and the phase angle at $\omega = \infty$. If the slope of the log-magnitude curve as $\omega \to \infty$ is $-20(q - p)$ dB/decade and the phase angle at $\omega = \infty$ is equal to $-90°(q - p)$, then the system is minimum phase.
- Nonminimum-phase systems are slow in response because of their faulty behavior at the start of the response.
- In most practical control systems, excessive phase lag should be carefully avoided. A common example of a nonminimum-phase element that may be present in a control system is transport lag: $e^{-\tau_{dt}s} = 1\angle -\omega\tau_{dt}$

Unit Step Responses

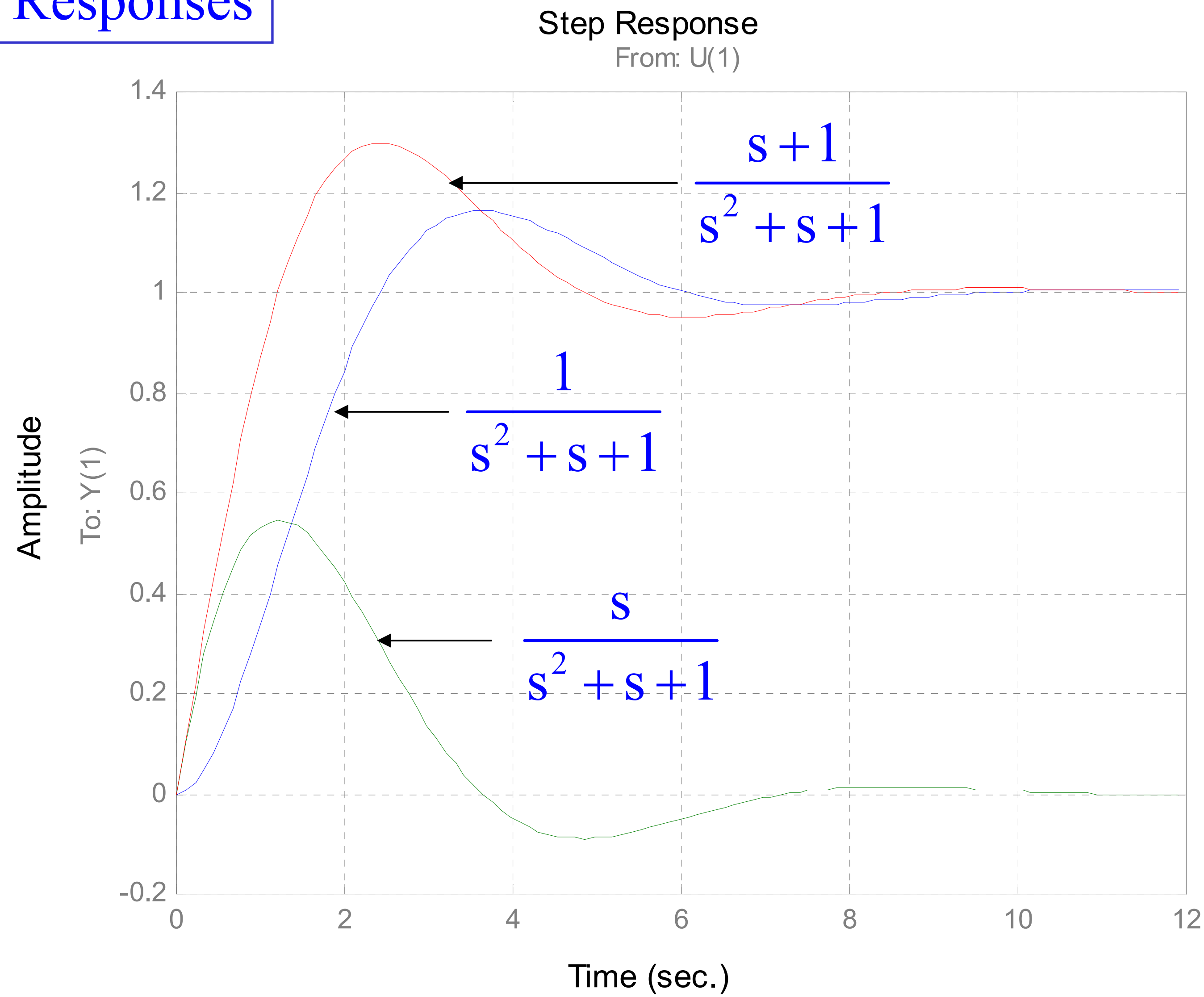


$$\frac{s+1}{s^2+s+1}$$

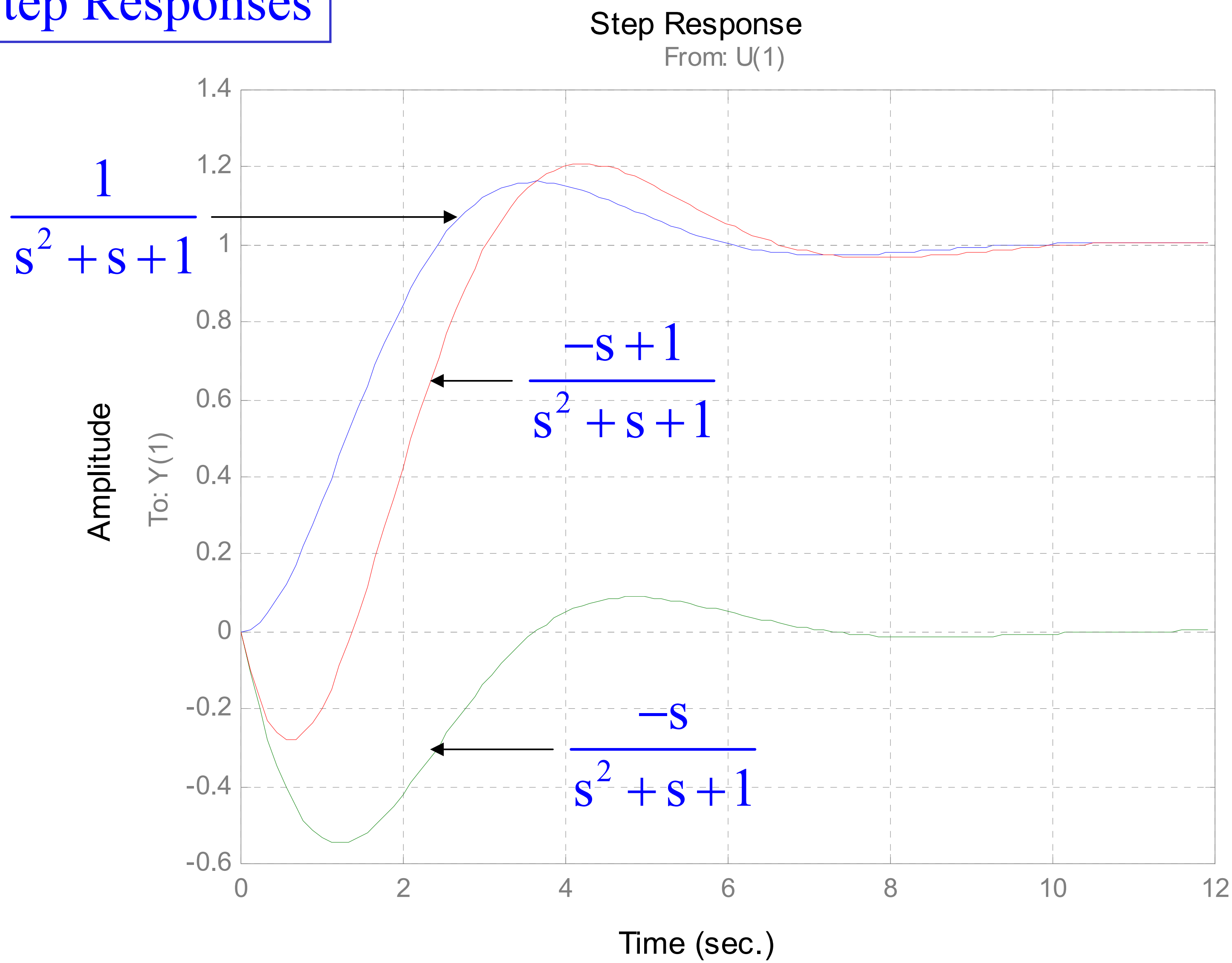
$$\frac{1}{s^2+s+1}$$

$$\frac{-s+1}{s^2+s+1}$$

Unit Step Responses



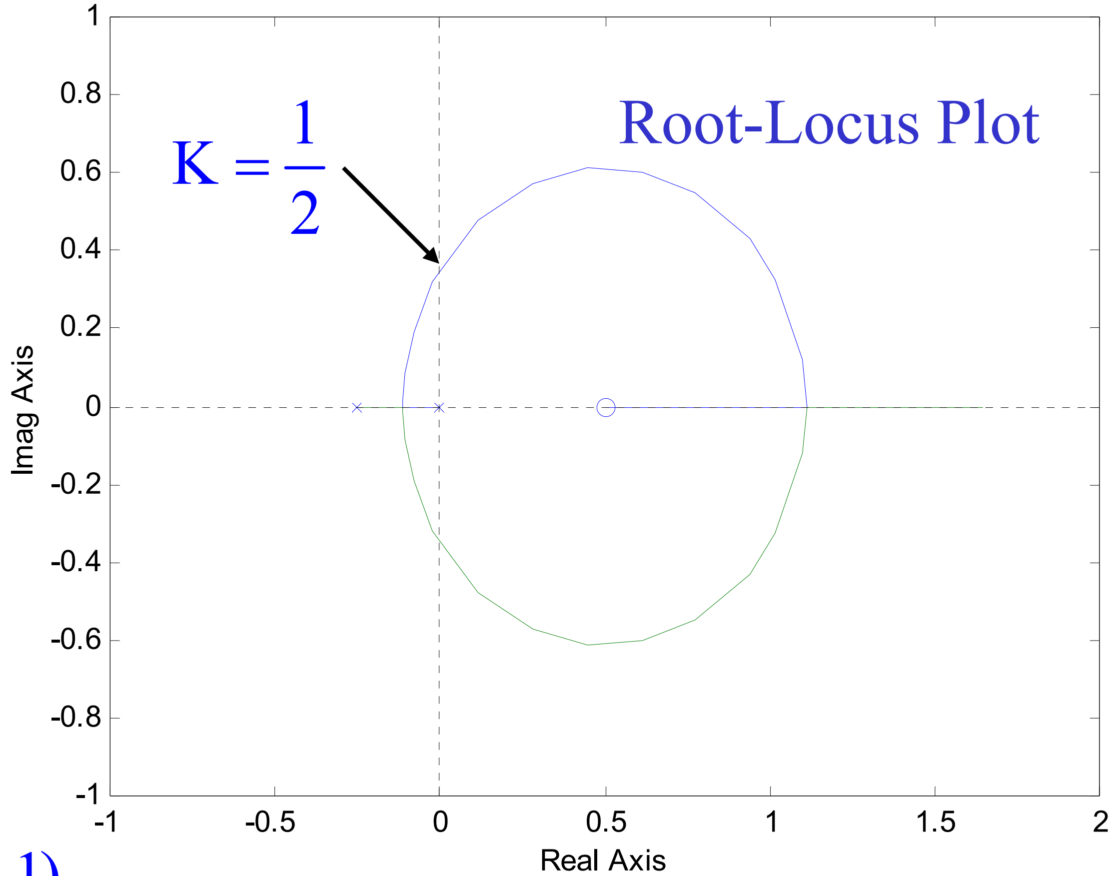
Unit Step Responses



- Nonminimum-Phase Systems: Root-Locus View
  - If all the poles and zeros of a system lie in the LHP, then the system is called *minimum phase*.
  - If at least one pole or zero lies in the RHP, then the system is called *nonminimum phase*.
  - The term nonminimum phase comes from the phase-shift characteristics of such a system when subjected to sinusoidal inputs.
  - Consider the open-loop transfer function:

$$G(s)H(s) = \frac{K(1-2s)}{s(4s+1)}$$

$$G(s)H(s) = \frac{K(1-2s)}{s(4s+1)}$$



Angle Condition:

$$\angle G(s)H(s) = \angle \frac{-K(2s-1)}{s(4s+1)}$$

$$= \angle \frac{K(2s-1)}{s(4s+1)} + 180° = \pm 180°(2k+1) \quad \text{or} \quad \angle \frac{K(2s-1)}{s(4s+1)} = 0°$$

Unmodeled Resonances

- Accurate modeling of the dynamic behavior of a mechanical system will result in a dynamic system of higher order than you probably would want to use for the design model.
- For example, consider a shaft that connects a drive motor to a load. Possibilities include:
  - Shaft has infinite stiffness (rigid)
  - Shaft has a stiffness represented by a spring constant that leads to a resonance in the model
  - Shaft is represented by a PDE that leads to an infinite number of resonances

- In most situations, the frequencies of these resonances will be orders of magnitude above the operating bandwidth of the control system and there will be enough natural damping present in the system to prevent any trouble.
- In applications that require the system to have a bandwidth that approaches the lowest resonance frequency, difficulties can arise.
- A control system based on a design model that does not account for the resonance may not provide enough loop attenuation to prevent oscillation and possible instability at or near the frequency of the resonance.

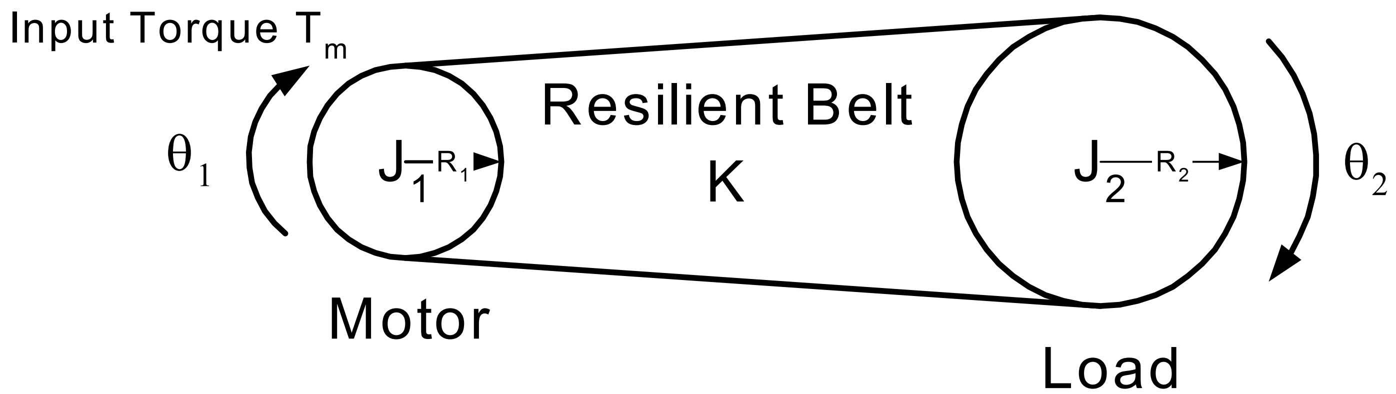
- If the precise nature of the resonances are known, they can be modeled and included in the design model.
- However, in many applications the frequencies of the poles (and neighboring zeros) of the resonances are not known with precision or may shift during the operation of the system. A small error in a resonance frequency, damping, or distance between the pole and zero might result in a compensator design that is even worse than a compensator that ignores the resonance phenomenon.

- Alternatives to including the resonances in the design model:
  - Increase the loop attenuation at frequencies above the desired operating point by adding one or more stages of low-pass filtering to the compensator. If the cut-off frequencies of the low-pass filters are well above the loop-gain crossover frequency, the additional phase lag introduced by these filters should not seriously compromise overall stability. Since the phase lag due to these filters starts to become effective well before their crossover frequency, however, it is important that the phase margin of the loop without filters be large enough to handle the additional phase shift caused by the filters.

– Cover the resonance with narrow-band noise, instead of attempting to include an accurate model of the actual resonance in the design model. This is accomplished by assuming the presence of a disturbance input *d* defined by: $\ddot{d} + 2\zeta\omega\dot{d} + d = w$

where *w* is white noise. The noise bandwidth (which is adjusted by the damping factor ζ) is chosen to be broad enough to encompass the entire range of possible resonance frequencies. The spectral density of the white noise is selected to produce as much an effect as that of the resonance. The differential equation of the narrow-band noise is used in the control system design model and prevents the compensator from relying upon the precise resonance frequency but rather places a broad notch in the vicinity of the resonance.

- Whatever method is chosen, it is important to evaluate the stability of the resulting design in the presence of resonances not included in the design model.
- Consider the resonance in a belt-driven servo-system.



- For simplicity assume:
  - Belt is modeled as an ideal spring
  - Both inertias are equal
  - No damping

Rigid Belt Case

$$R_1\theta_1 = R_2\theta_2$$

$$\left[J_1 + \left(\frac{R_1}{R_2}\right)^2 J_2\right]\ddot{\theta}_1 = T_m$$

- The equations of motion are:

$$J_1\ddot{\theta}_1 + 2KR_1^2\theta_1 - 2KR_1R_2\theta_2 = T_m$$
$$J_2\ddot{\theta}_2 + 2KR_2^2\theta_2 - 2KR_1R_2\theta_1 = 0$$

- Take the Laplace transform of the equations:

$$\begin{bmatrix} J_1s^2 + 2KR_1^2 & -2KR_1R_2 \\ -2KR_1R_2 & J_2s^2 + 2KR_2^2 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} T_m \\ 0 \end{bmatrix}$$

- Determine the transfer functions:

$$\frac{\theta_1}{T_m} = \frac{J_2s^2 + 2KR_2^2}{s^2\left[J_1J_2s^2 + 2K\left(R_1^2J_2 + R_2^2J_1\right)\right]} \qquad \frac{\theta_2}{T_m} = \frac{2KR_1R_2}{s^2\left[J_1J_2s^2 + 2K\left(R_1^2J_2 + R_2^2J_1\right)\right]}$$

Poles and Zeros of Transfer Functions

Definition of Poles and Zeros

- A pole of a transfer function *G(s)* is a value of *s* (real, imaginary, or complex) that makes the denominator of *G(s)* equal to zero.
- A zero of a transfer function *G(s)* is a value of *s* (real, imaginary, or complex) that makes the numerator of *G(s)* equal to zero.
- For Example:

$$G(s) = \frac{K(s+2)(s+10)}{s(s+1)(s+5)(s+15)^2}$$

Poles: 0, -1, -5, -15 (order 2)

Zeros: -2, -10, ∞ (order 3)

- Collocated Control System
  - All energy storage elements that exist in the system exist outside of the control loop.
  - For purely mechanical systems, separation between sensor and actuator is at most a rigid link.
- Non-Collocated Control System
  - At least one storage element exists inside the control loop.
  - For purely mechanical systems, separating link between sensor and actuator is flexible.

- Physical Interpretation of Poles and Zeros
  - Complex Poles of a collocated control system and those of a non-collocated control system are identical.
  - Complex Poles represent the resonant frequencies associated with the energy storage characteristics of the entire system.
  - Complex Poles, which are the natural frequencies of the system, are independent of the locations of sensors and actuators.
  - At a frequency of a complex pole, even if the system input is zero, there can be a nonzero output.

- Complex Poles represent the frequencies at which energy can freely transfer back and forth between the various internal energy storage elements of the system such that even in the absence of any external input, there can be nonzero output.
- Complex Poles correspond to the frequencies where the system behaves as an energy reservoir.
- Complex Zeros of the two control systems are quite different and they represent the resonant frequencies associated with the energy storage characteristics of a sub-portion of the system defined by artificial constraints imposed by the sensors and actuators.

- Complex Zeros correspond to the frequencies where the system behaves as an energy sink.
- Complex Zeros represent frequencies at which energy being applied by the input is completely trapped in the energy storage elements of a sub-portion of the original system such that no output can ever be detected at the point of measurement.
- Complex Zeros are the resonant frequencies of a subsystem constrained by the sensors and actuators.

- Collocated Control System

Transfer Function:

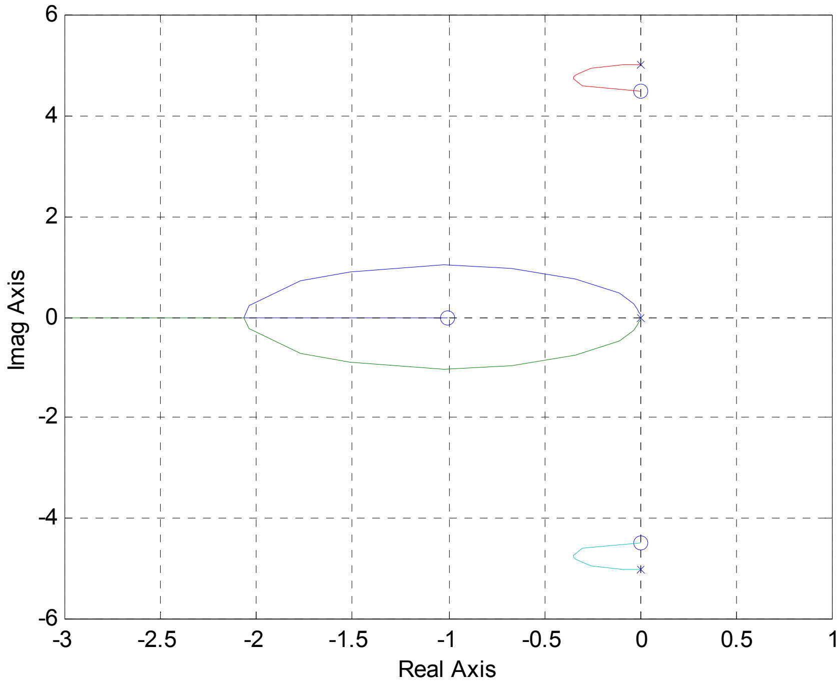
$$\frac{\theta_1}{T_m} = \frac{J_2 s^2 + 2KR_2^2}{s^2\left[J_1 J_2 s^2 + 2K\left(R_1^2 J_2 + R_2^2 J_1\right)\right]}$$

Poles:

$$0, 0, \pm i\sqrt{\frac{2K(R_1^2 J_2 + R_2^2 J_1)}{J_1 J_2}}$$

Zeros: $\pm i\sqrt{\frac{2KR_2^2}{J_2}}$

Root-Locus Plot with PD Controller

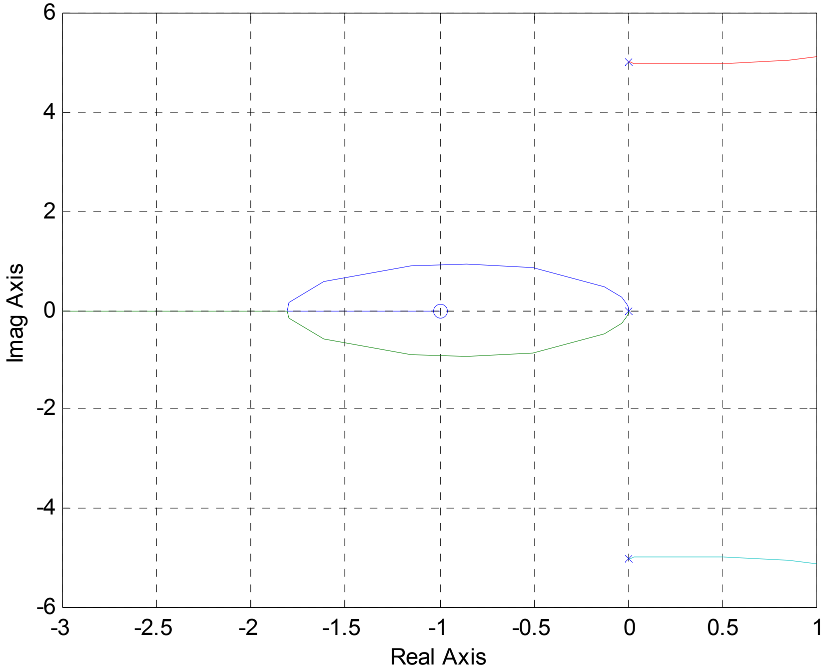


• Non-Collocated Control System

Transfer Function:

$$\frac{\theta_2}{T_m} = \frac{2KR_1R_2}{s^2\left[J_1J_2s^2 + 2K\left(R_1^2J_2 + R_2^2J_1\right)\right]}$$

Root-Locus Plot with PD Controller



Poles:

$$0, 0, \pm i\sqrt{\frac{2K(R_1^2J_2 + R_2^2J_1)}{J_1J_2}}$$

Zeros: None

- Reality is not so grim as this analysis would make it seem, otherwise one would never be able to stabilize systems with belts or flexible shafts.
- There will be friction in the motor and pulleys and structural damping in the resilient belt that are stabilizing.
- Moreover, the bandwidth of the amplifier that furnishes the voltage to the motor is likely to be lower than the resonance frequency and will provide additional attenuation at this frequency, thus further stabilizing the system.

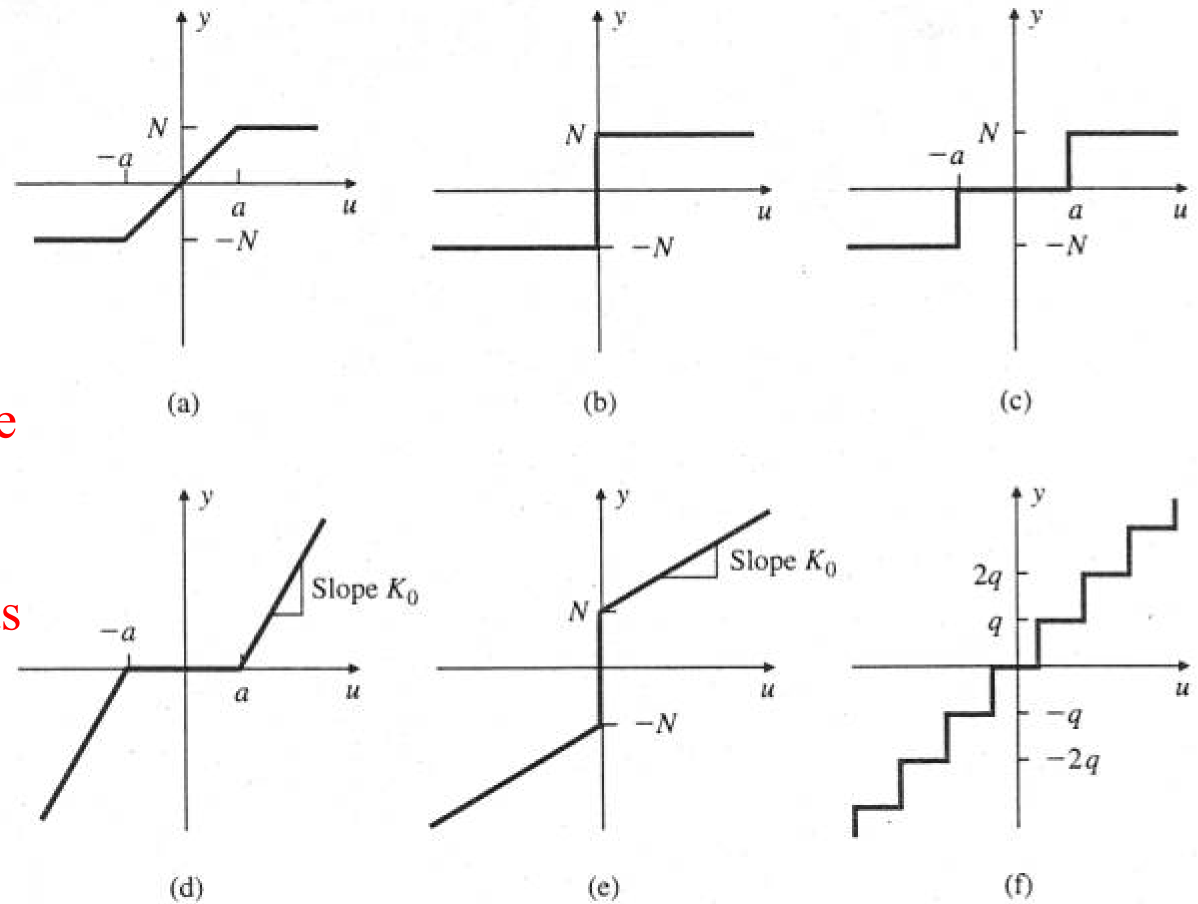
- If you do not want to rely on the uncertain characteristics of the hardware to stabilize the system, you must attend to the stabilization in your compensator design. A simple low-pass filter between the the control output and the input to the motor might do the job.
- Remember, if you want the system to operate with a bandwidth near the lowest resonance frequency, either you must include the resonance in the design model or be prepared to consider other measures to avoid the possible unfavorable consequences.

Saturation

- Nonlinear Systems
  - Every real control system is nonlinear and we use linear approximations to the real models.
  - There is one important category of nonlinear systems for which some significant analysis can be done: systems in which the nonlinearity has no dynamics and is well approximated as a gain that varies as the size of its input signal varies.
  - The behavior of systems containing such a nonlinearity can be quantitatively described by considering the nonlinear element as a varying, signal-dependent gain.

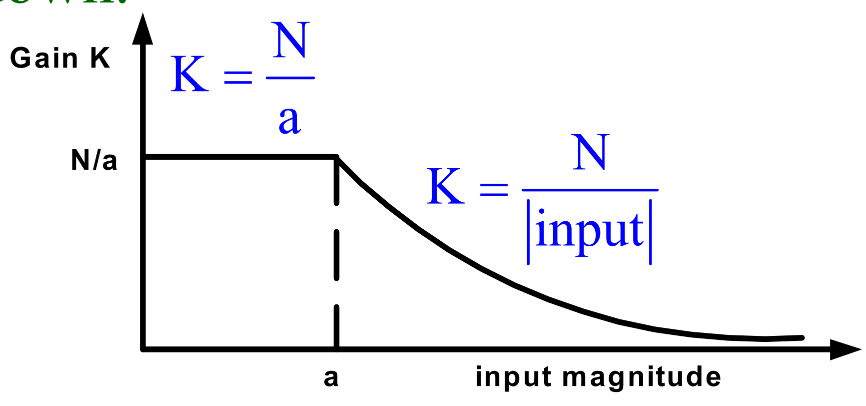
Nonlinear Elements with No Dynamics

(a) Saturation
(b) Relay
(c) Relay with Dead Zone
(d) Gain with Dead Zone
(e) Pre-loaded Spring or Coulomb plus Viscous Friction
(f) Quantization



- As an example, consider the *saturation element*. All actuators saturate at some level; if they did not, their output would increase to infinity, which is physically impossible.
- For the saturation element, it is clear that for input signals with magnitudes < *a*, the nonlinearity is linear with the gain *N/a*. However, for signals > *a*, the output size is bounded by *N*, while the input size can get much larger than *a*, so once the input exceeds *a*, the ratio of output to input goes down.

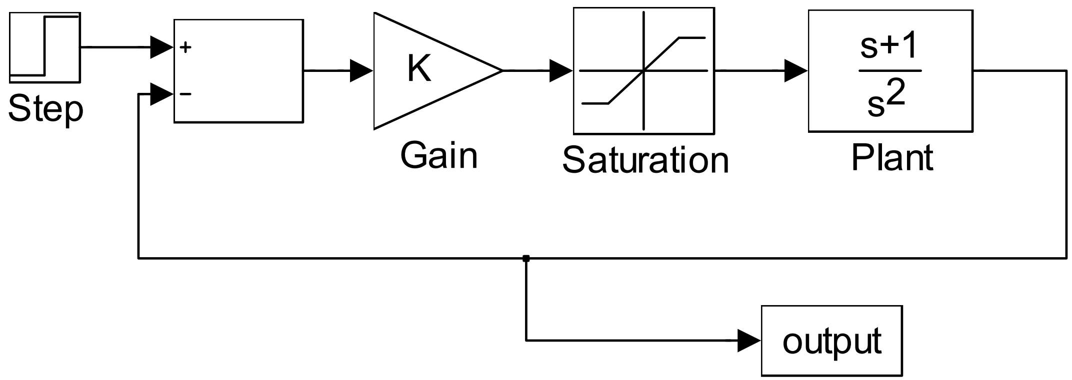
General Shape of the Effective Gain of Saturation

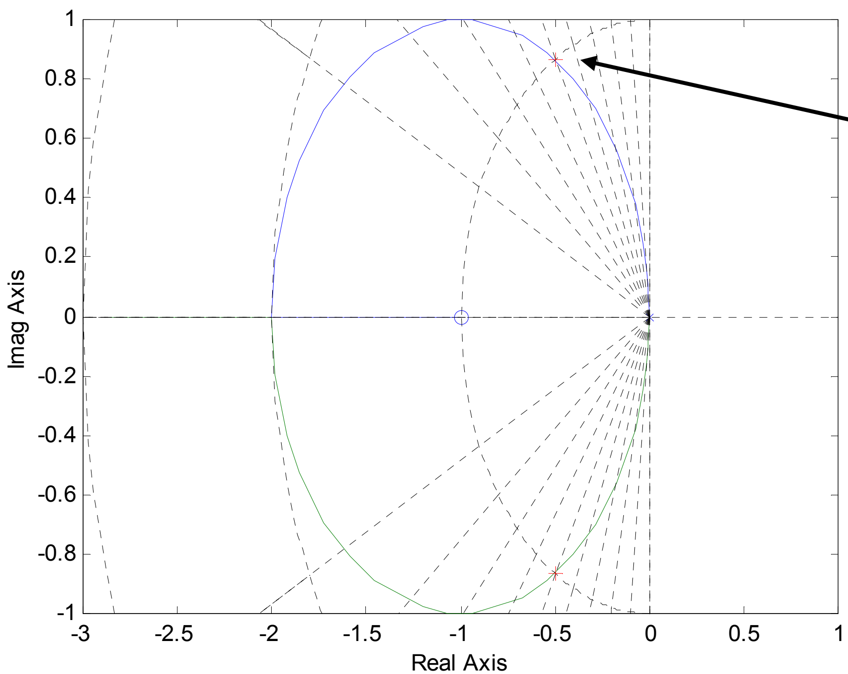


- An important aspect of control system design is *sizing the actuator*, which means picking the size, weight, power required, cost, and saturation level of the device.
- Generally, higher saturation levels require bigger, heavier, and more costly actuators.
- The key factor that enters into the sizing is the effect of the saturation on the control system's performance.

Dynamic System With Saturation

Saturation levels: ± 0.4





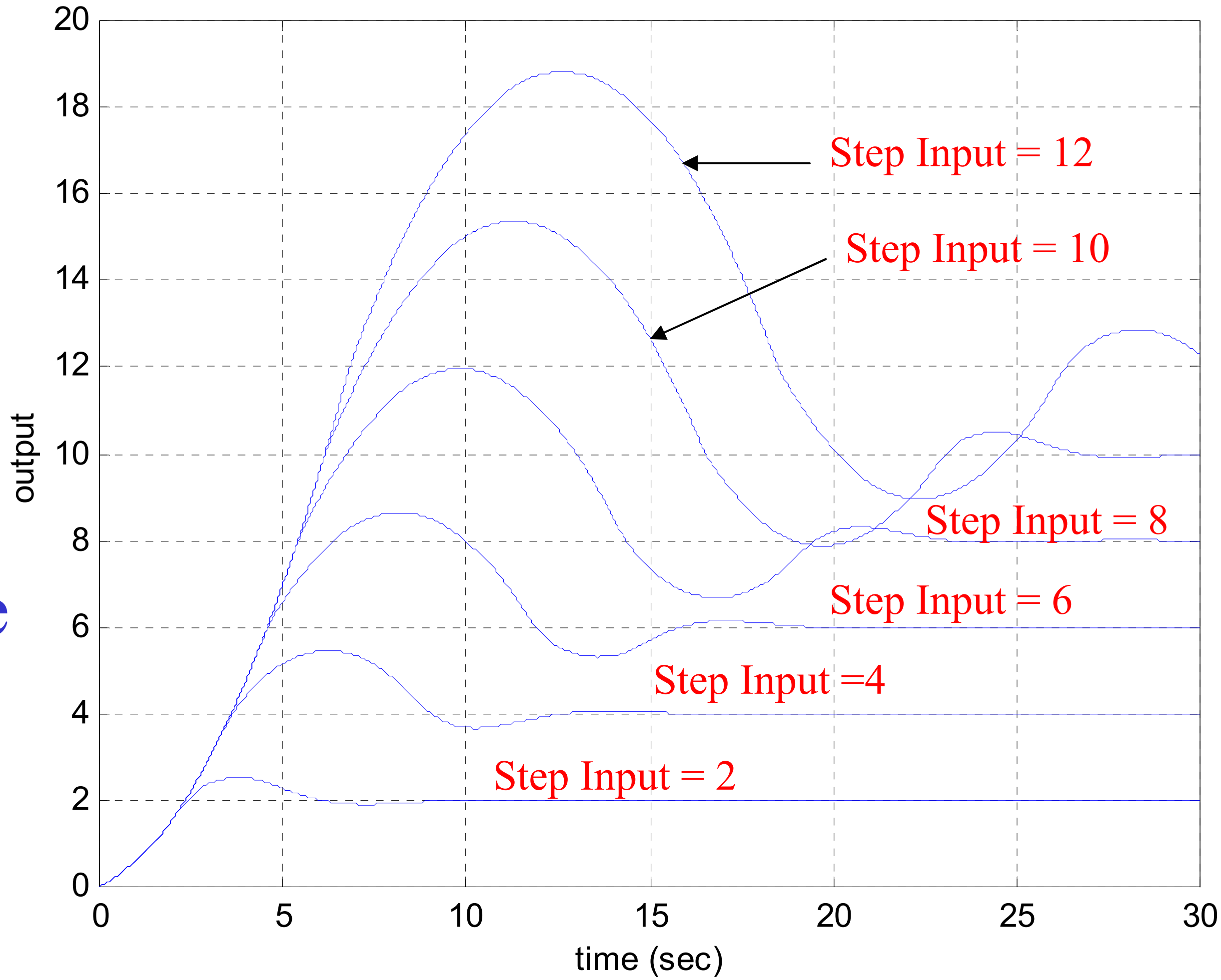
$\zeta = 0.5$, K = 1

Root-Locus Plot Without Saturation

As K is reduced, the roots move toward the origin of the s-plane with less and less damping.

Step-Response Results

$\zeta = 0.5, K = 1$

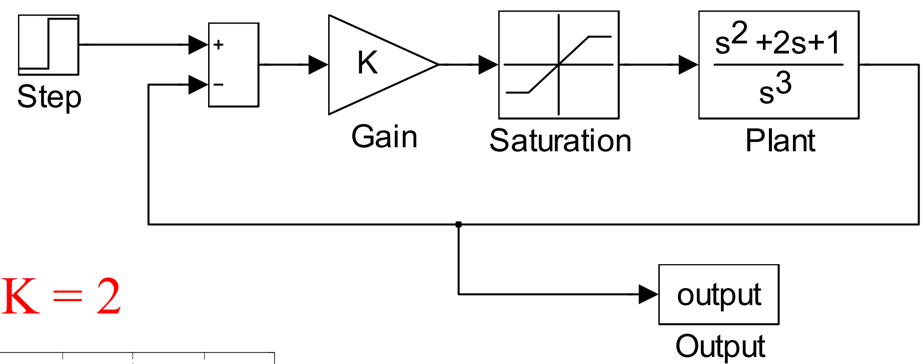


– Observations

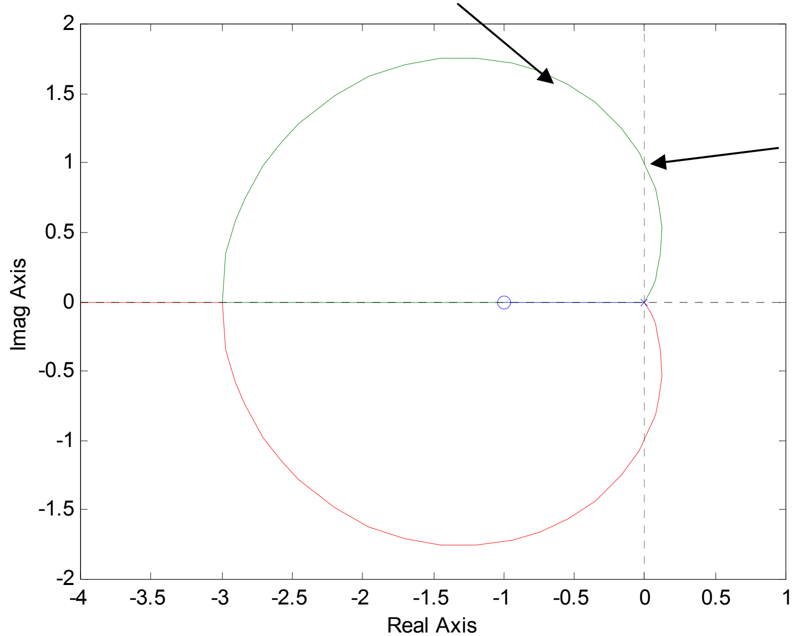
- As long as the signal entering the saturation remains less than 0.4, the system will be linear and should behave according to the roots at $\zeta = 0.5$.
- However, notice that as the input gets larger, the response has more and more overshoot and slower and slower recovery.
- This can be explained by noting that larger and larger input signals correspond to smaller and smaller effective gain K.
- From the root-locus plot, we see that as K decreases, the closed-loop poles move closer to the origin and have a smaller damping ζ.
- This results in the longer rise and settling times, increased overshoot, and greater oscillatory response.

– As another example, consider the block diagram below.

Saturation levels: ± 1



ζ =0.5, K = 2



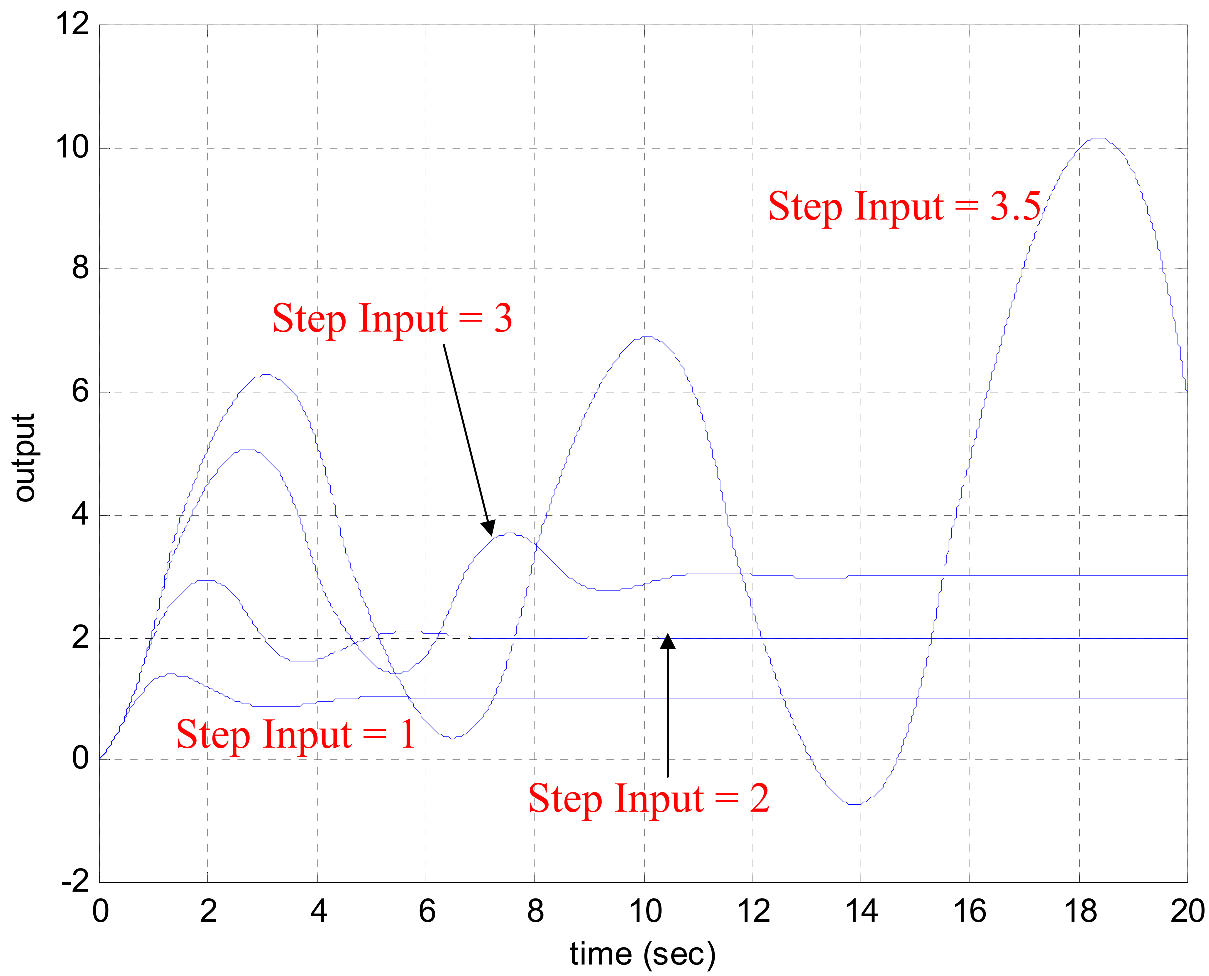
K = 1/2

Root-Locus Plot Without Saturation

System is stable for large gains but unstable for smaller gains

Step-Response Results

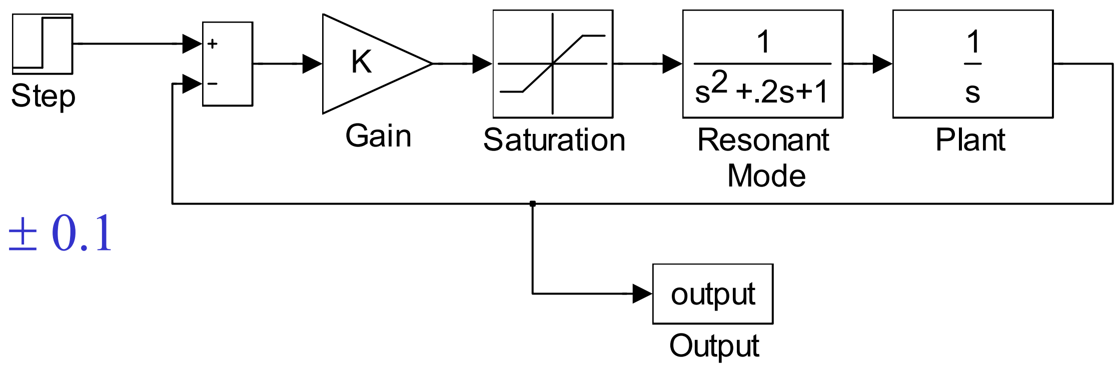
$\zeta = 0.5$, $K = 2$



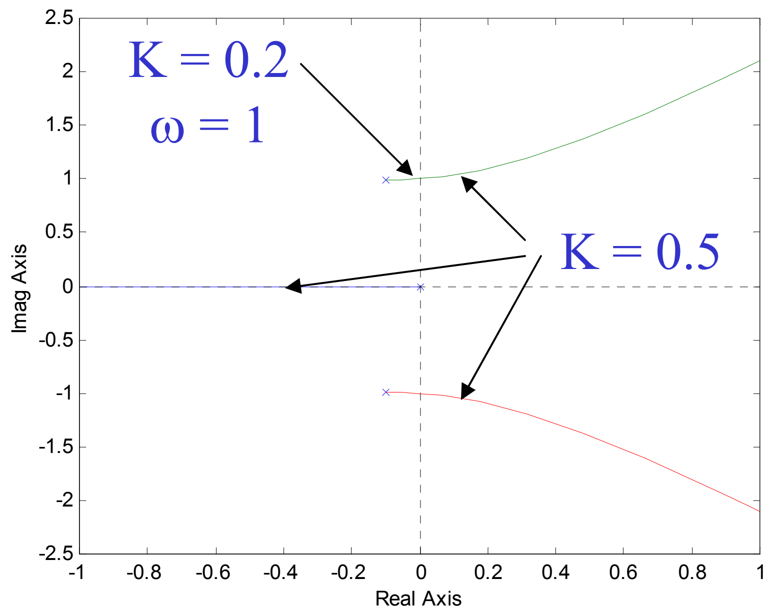
– Observations

- For K = 2, which corresponds to $\zeta = 0.5$ on the root locus, the system shows responses consistent with $\zeta = 0.5$ for small signals.
- As the signal strength is increased, the response becomes less well damped.
- As the signal strength is increased even more, the response becomes unstable.

– As a final example, consider the following block diagram.

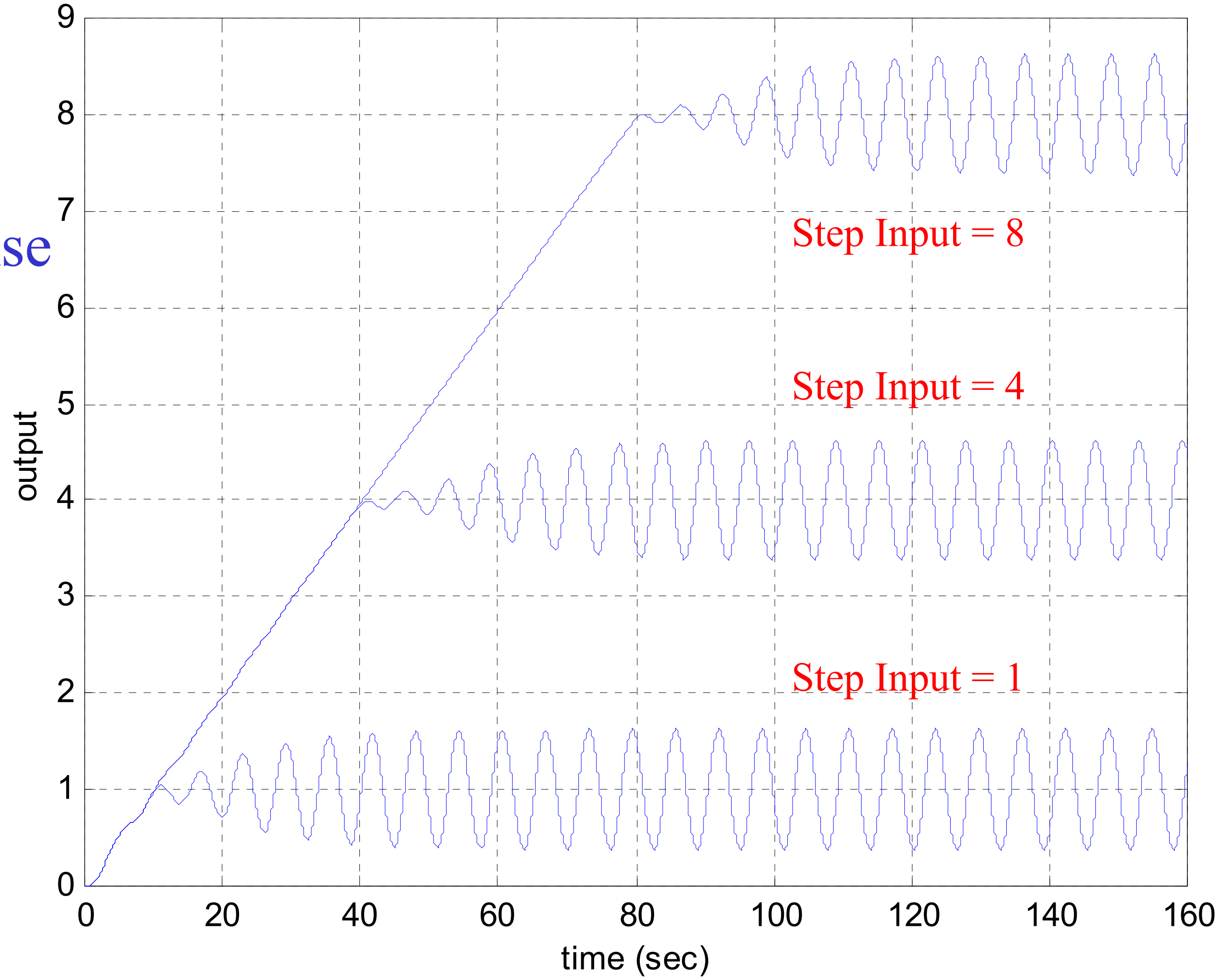


Saturation levels: ± 0.1



Root-Locus Plot Without Saturation

Step-Response Results



– Observations

- This system is typical of electromechanical control problems where the designer perhaps at first is not aware of the resonant mode corresponding to the denominator term $s^2 + 0.2s + 1$ ($\omega = 1$, $\zeta = 0.1$).
- A gain of $K = 0.5$ is enough to force the roots of the resonant mode into the RHP. At this gain our analysis predicts a system that is initially unstable, but becomes stable as the gain decreases.
- Thus we see that the response of the system with saturation builds up due to the instability until the magnitude is sufficiently large that the effective gain is lowered to $K = 0.2$ and then stops growing!

- The error builds up to a fixed amplitude and then starts to oscillate. The oscillations have a frequency of 1 rad/sec and hold constant amplitude at any DC equilibrium value (for the three different step inputs).
- The response always approaches a periodic solution of fixed amplitude known as a *limit cycle*, so-called because the response is cyclic and is approached in the limit as time grows large.
- In order to prevent the limit cycle, the root locus has to be modified by compensation so that no branches cross into the RHP. One common method to do this for a lightly-damped oscillatory mode is to place compensation zeros near the poles, but at a slightly lower frequency.

Backlash

- Gears and similar drive systems generally exhibit an effect called backlash.
- The two key phenomena associated with backlash are:
  - Hysteresis which occurs because the relative positions of the two halves of the backlash mechanism depend on the direction of motion.
  - Bounce which occurs when the two halves of the backlash mechanism impact after they have separated due to a change in direction. The amount of bounce depends on the coefficient of restitution of the two surfaces and the speed at which they impact.

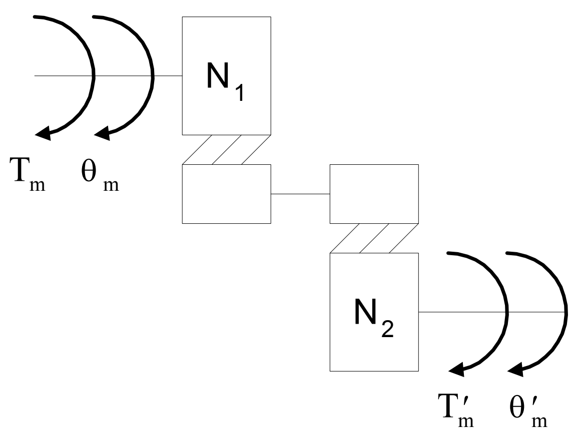
- As an example, consider a motor connected to a drive gear and that the driven gear is connected to an inertial load. The equations of motion are:

$$J_1\ddot{\theta}_1 + \tau = T_{in}$$

$$J_2\ddot{\theta}_2 - \tau = 0$$

  - τ is the torque transmitted through the gears

Gear Train Relations:



$$\frac{\theta_m}{\theta'_m} = \frac{N_2}{N_1} \equiv N$$

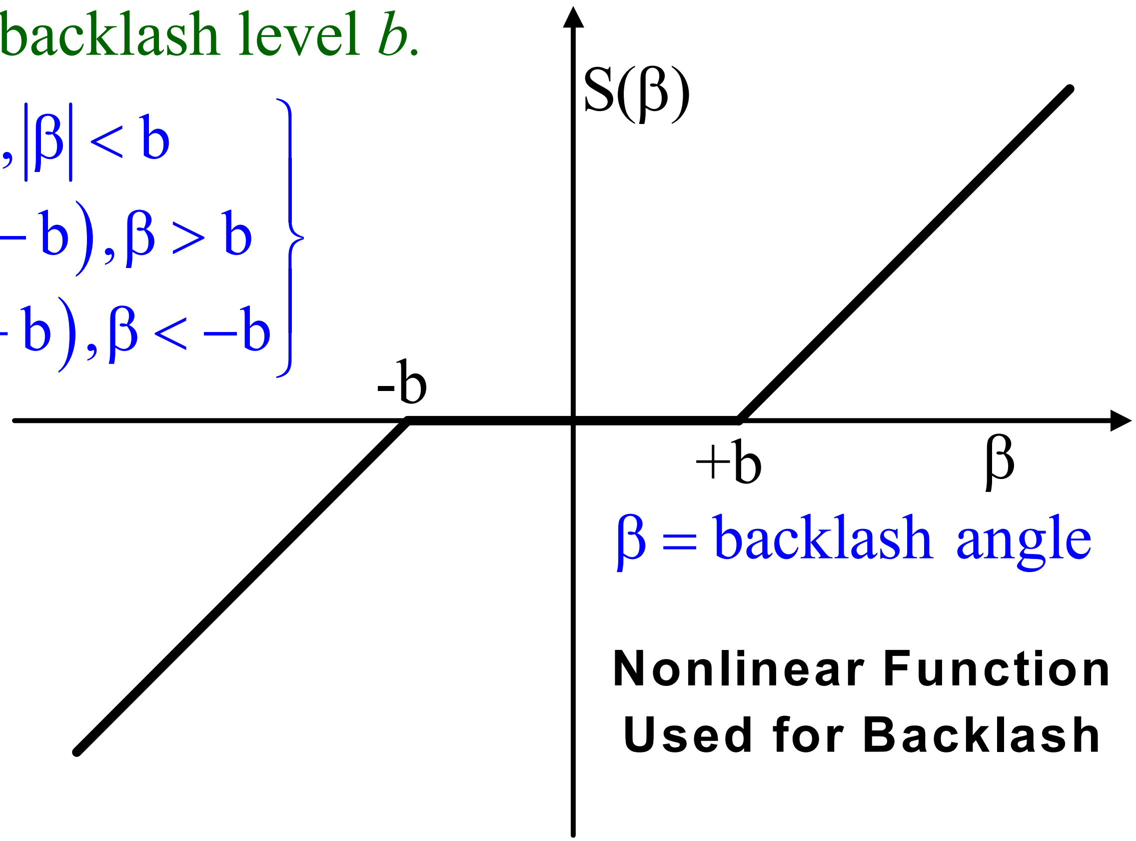
$$\frac{T_m}{T'_m} = \frac{N_1}{N_2} \equiv \frac{1}{N}$$

- When the gears are not engaged, the torque τ is zero and the two equations of motion are uncoupled.
- When the gears are engaged, however, the motion is constrained such that $\theta_1 = \theta_2$ and the torque τ is whatever it must be to maintain the constraint.
- Once the teeth are in contact, they will remain in contact until the relative motion of the gears changes in direction.
- The order of the system depends on whether or not the gears are engaged: if they are not engaged the system has two degrees of freedom and the system is 4^{th} order; if they are engaged, the system has only one degree of freedom and the system is 2^{nd} order.

– This paradox can be resolved by representing the effect of the gears by a highly nonlinear spring, in which the torque is zero for small angular displacements and becomes very large when the relative displacement exceeds the backlash level *b*.

$$\tau = f(\beta) = \begin{Bmatrix} 0, |\beta| < b \\ S(\beta - b), \beta > b \\ S(\beta + b), \beta < -b \end{Bmatrix}$$

$$\beta = \theta_1 - \theta_2$$



β = backlash angle

Nonlinear Function Used for Backlash

- The differential equations can now be written as:

$$\left.\begin{aligned} J_1\ddot{\theta}_1 + \tau &= T_{in} \\ J_2\ddot{\theta}_2 - \tau &= 0 \end{aligned}\right\}$$

$$\ddot{\theta}_1 = -\frac{1}{J_1} f(\beta) + \frac{T_{in}}{J_1}$$

$$\ddot{\theta}_2 = \frac{1}{J_2} f(\beta)$$

- The differential equation for β is obtained from the above equations noting that $\ddot{\beta} = \ddot{\theta}_1 - \ddot{\theta}_2$

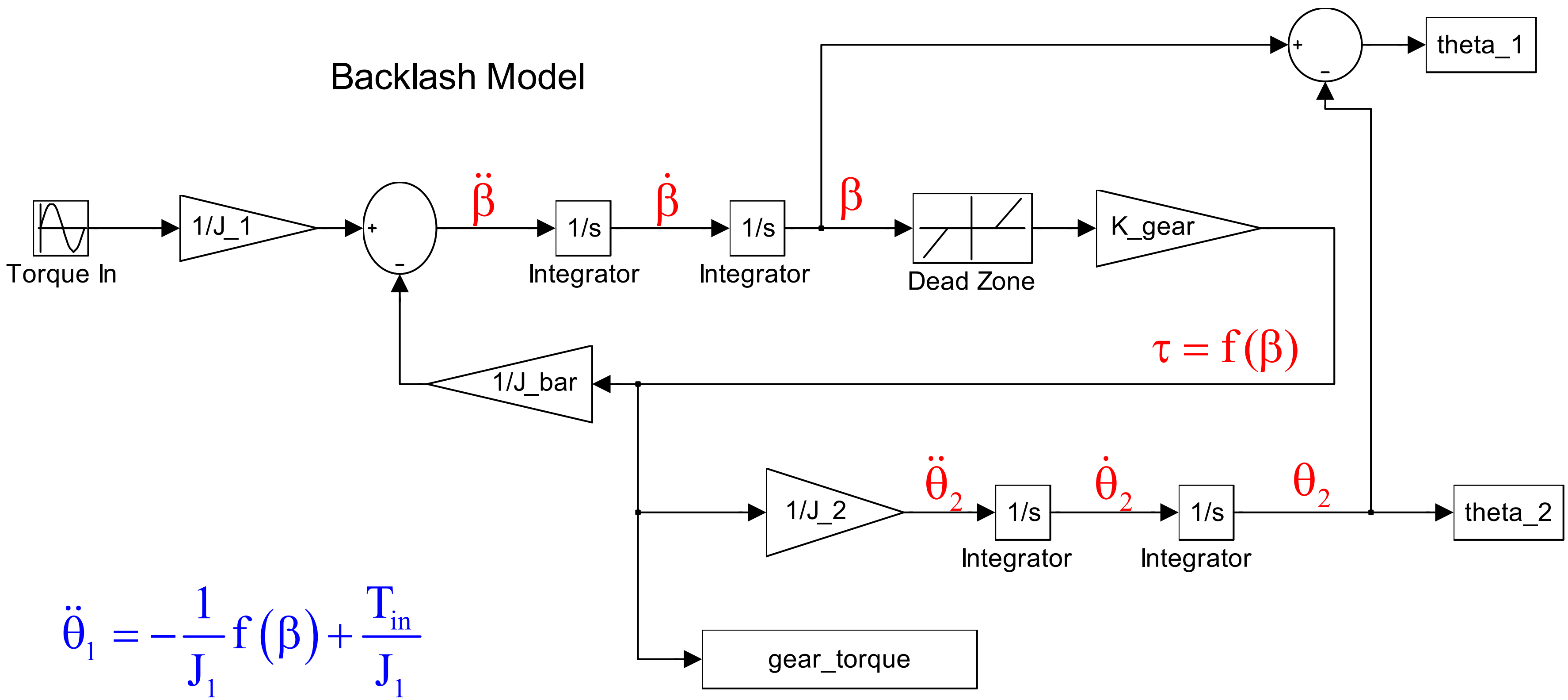
$$\ddot{\beta} = \left(\frac{-1}{J_1} + \frac{-1}{J_2}\right) f(\beta) + \frac{T_{in}}{J_1} = -\frac{1}{\overline{J}} f(\beta) + \frac{T_{in}}{J_1} \qquad \overline{J} = \frac{J_1 J_2}{J_1 + J_1}$$

- This equation defines the dynamics of a nonlinear oscillator.

$$\ddot{\beta} = \left(\frac{-1}{J_1} + \frac{-1}{J_2} \right) f(\beta) + \frac{T_{in}}{J_1} = -\frac{1}{\bar{J}} f(\beta) + \frac{T_{in}}{J_1}$$

$$\bar{J} = \frac{J_1 J_2}{J_1 + J_1} \qquad \beta = \theta_1 - \theta_2$$

Backlash Model



$$\ddot{\theta}_1 = -\frac{1}{J_1} f(\beta) + \frac{T_{in}}{J_1}$$

$$\ddot{\theta}_2 = \frac{1}{J_2} f(\beta)$$

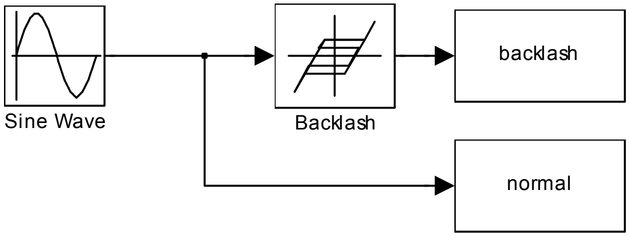
- MatLab Simulation of Backlash
  - The backlash block implements a system in which a change in input causes an equal change in output. However, when the input changes direction, an initial change in input has no effect on the output. The amount of side-to-side play in the system is referred to as the deadband. The deadband is centered about the output.
  - A system with play can be in one of three modes:
    - Disengaged: In this mode, the input does not drive the output and the output remains constant.
    - Engaged in a positive direction: In this mode, the input is increasing (has a positive slope) and the output is equal to the input minus half the deadband width.

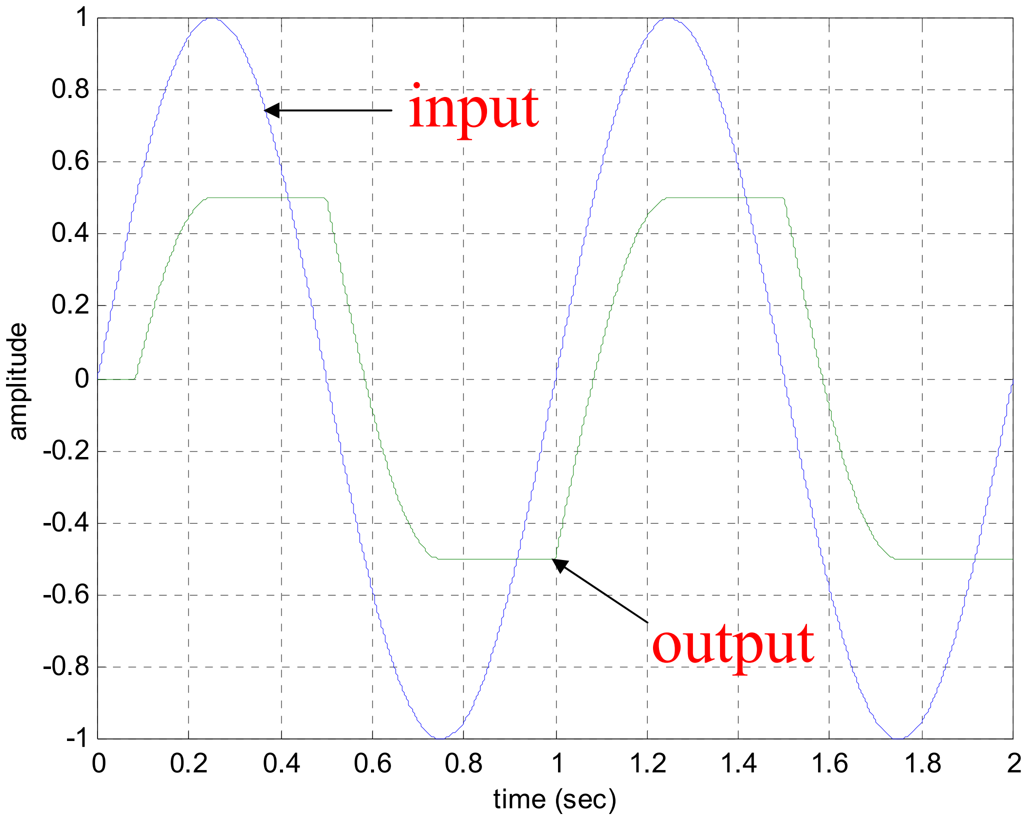
- Engaged in a negative direction: In this mode, the input is decreasing (has a negative slope) and the output is equal to the input plus half the deadband width.

– If the initial input is outside the deadband, the initial output parameter value determines if the block is engaged in a positive or negative direction and the output at the start of the simulation is the input plus or minus half the deadband width.

– This block can be used to model the meshing of two gears. The input and output are both shafts with a gear on one end, and the output shaft is driven by the input shaft. Extra space between the gear teeth introduces play. The width of this spacing is the deadband width parameter. If the system is disengaged initially, the output is defined by the initial output parameter.

– Consider the following example.





deadband width = 1
initial output = 0