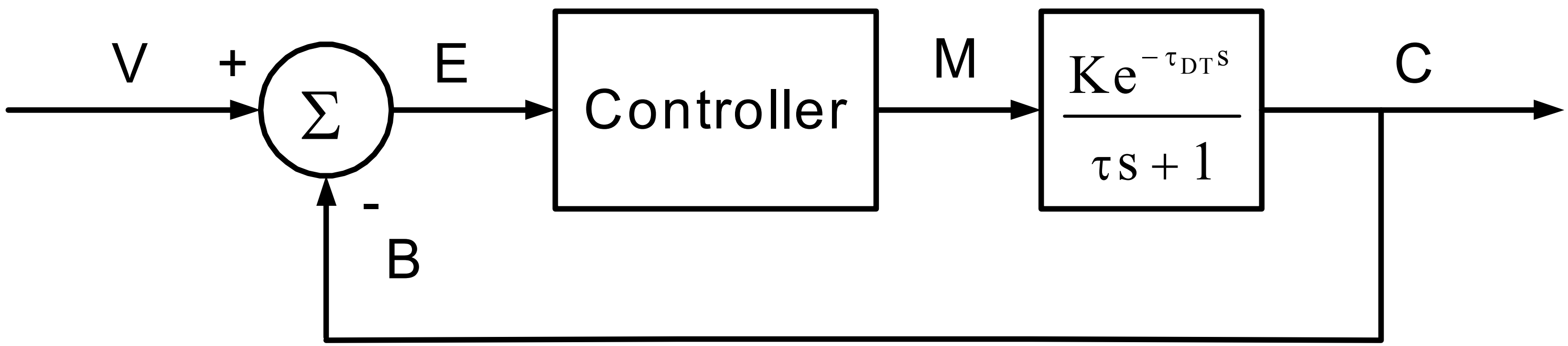
Control of a First-Order Process with Dead Time



The most commonly used model to describe the dynamics of chemical processes is the First-Order Plus Time Delay Model. By proper choice of τ_{DT} and τ, this model can be made to represent the dynamics of many industrial processes.

- 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

- 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.

$q_i(t) =$ input to dead-time element

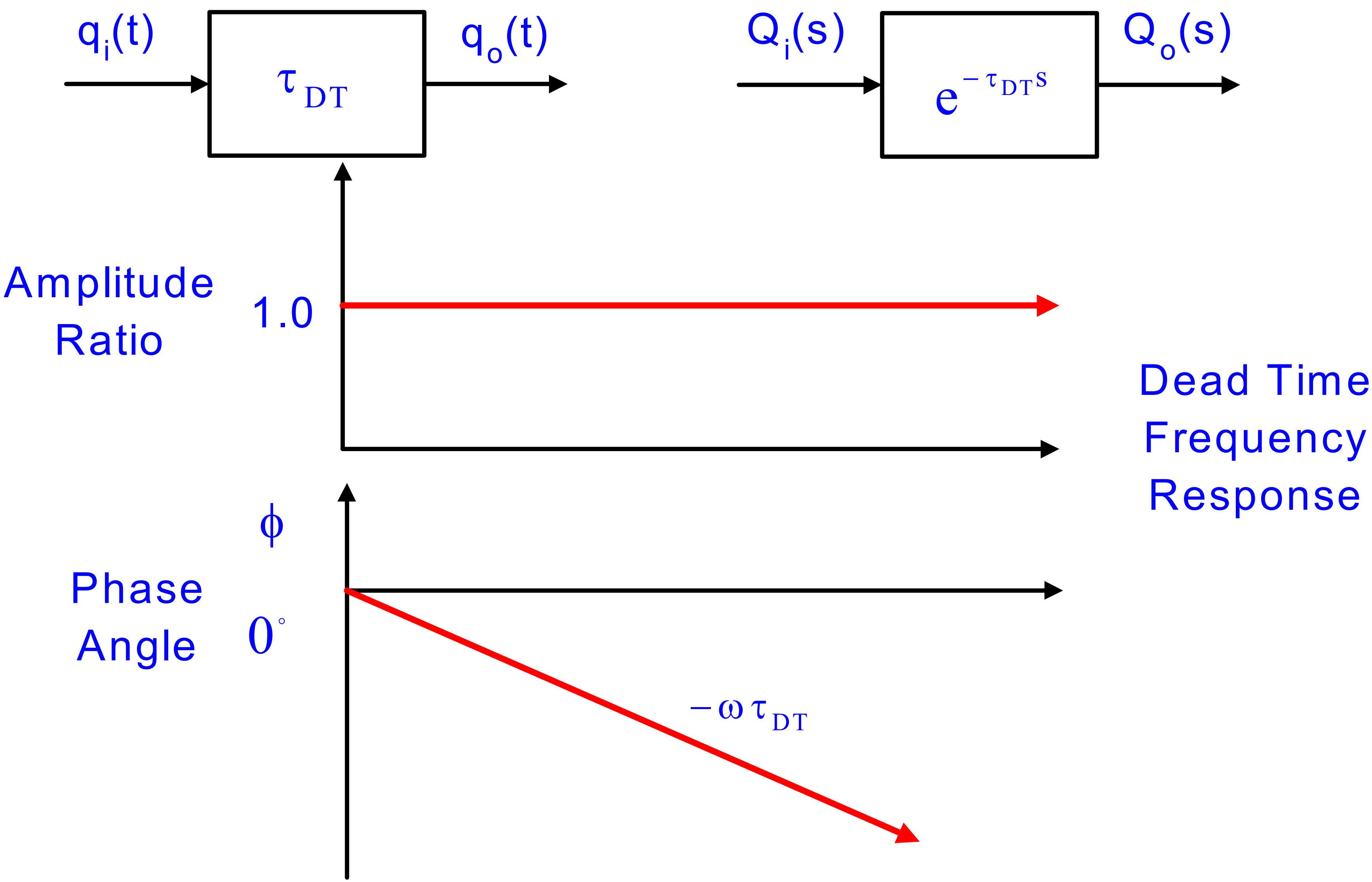
$q_o(t) =$ output of dead-time element $= q_i(t-\tau_{DT})u(t-\tau_{DT})$

$u(t-\tau_{DT}) = 1 \quad$ for $t \geq \tau_{DT}$

$u(t-\tau_{DT}) = 0 \quad$ for $t < \tau_{DT}$

Laplace Transform

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



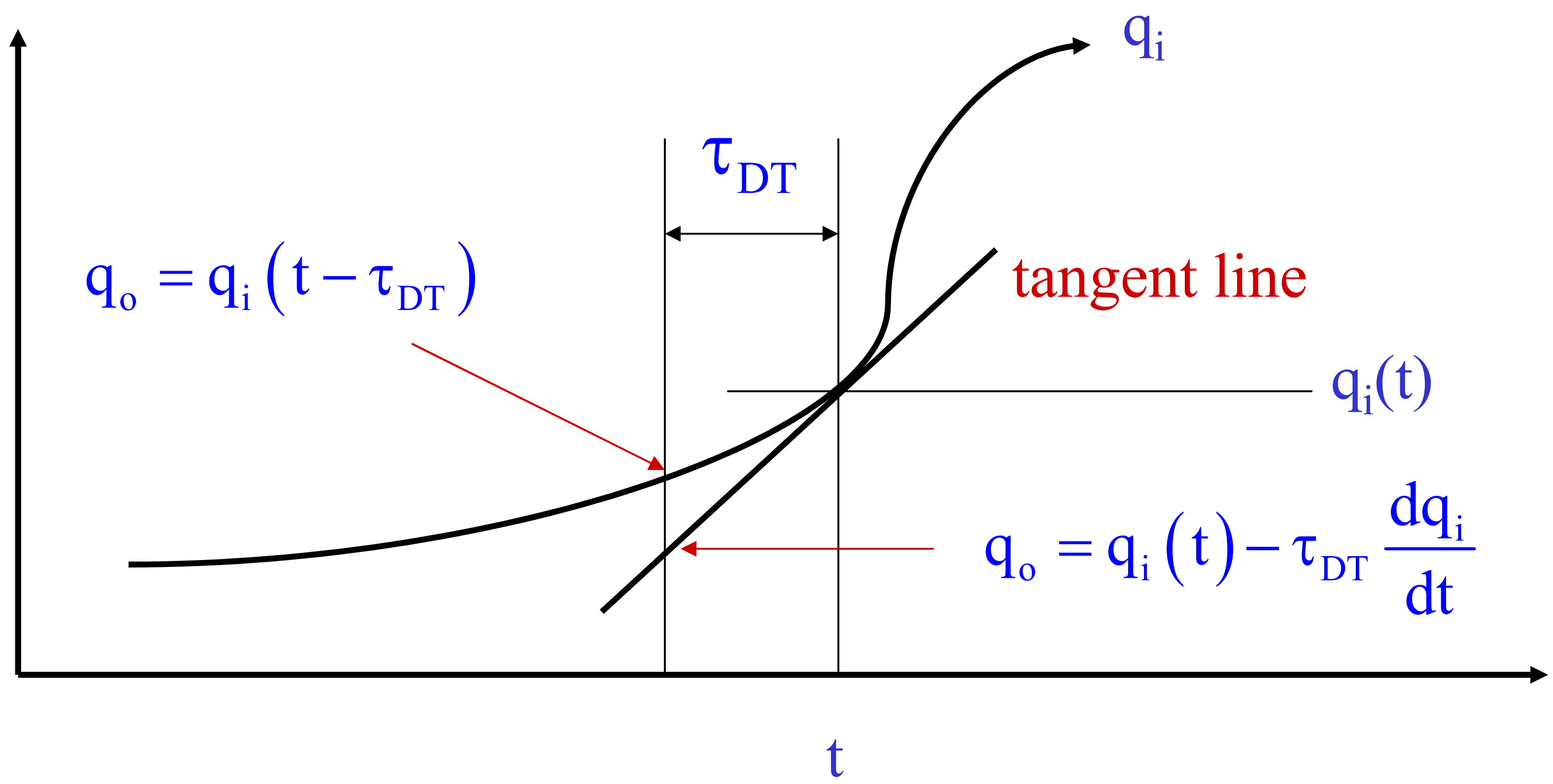
$q_i(t)$
τ_{DT}
$q_o(t)$
$Q_i(s)$
$e^{-\tau_{DT} s}$
$Q_o(s)$
Amplitude Ratio
1.0
Dead Time Frequency Response
ϕ
Phase Angle
0°
$-\omega \tau_{DT}$

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

$$\frac{Q_o}{Q_i}(s) = e^{-\tau_{DT}s} \approx 1 - \tau_{DT}s \qquad q_o(t) \approx q_i(t) - \tau_{DT}\frac{dq_i}{dt}$$

  - 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 for the range of frequencies present in $q_i(t)$, then the approximation is valid.

Dead-Time Graphical Approximation



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

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

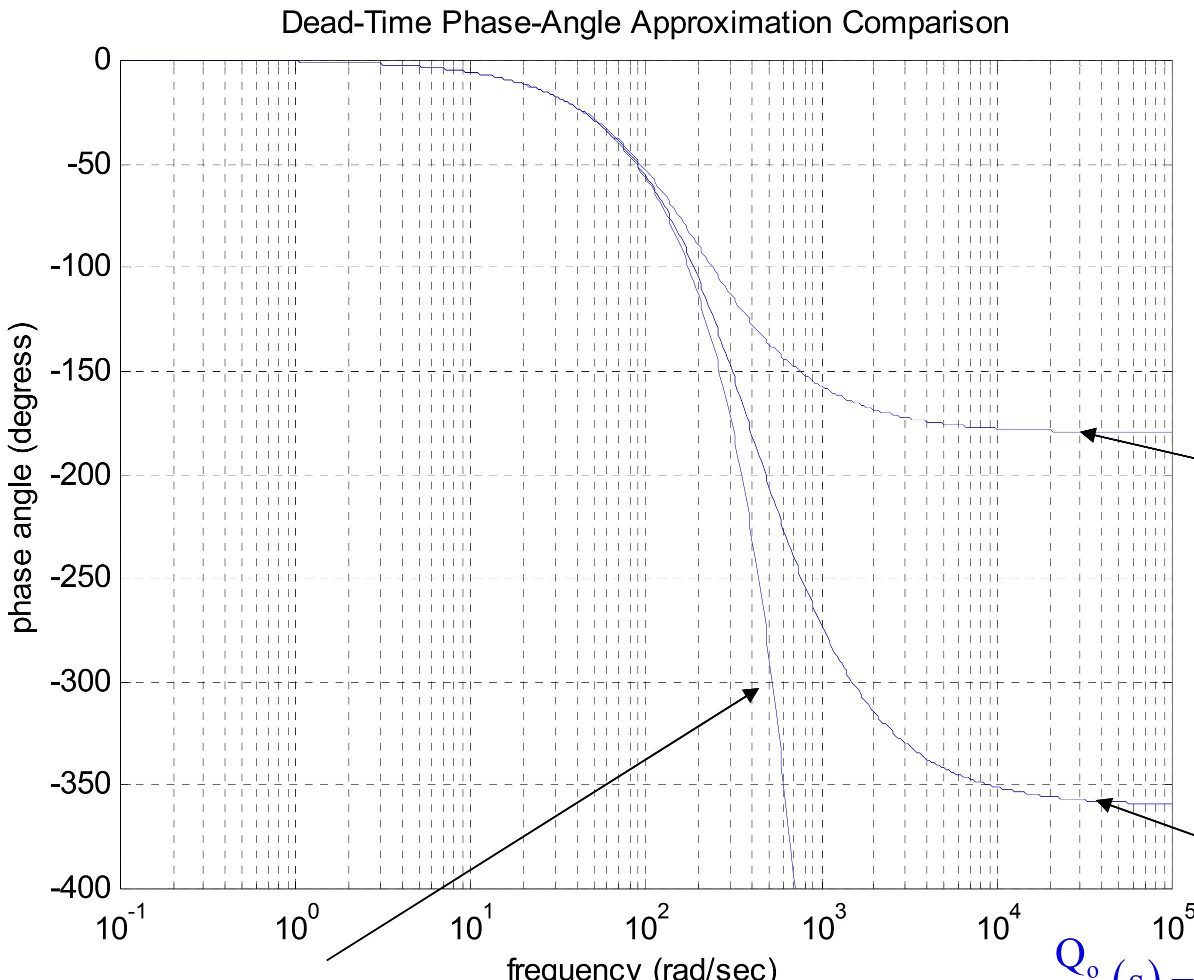
- In some cases, a very crude approximation given by a first-order lag is acceptable:

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

- 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



$$\tau_{dt} = 0.01$$

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

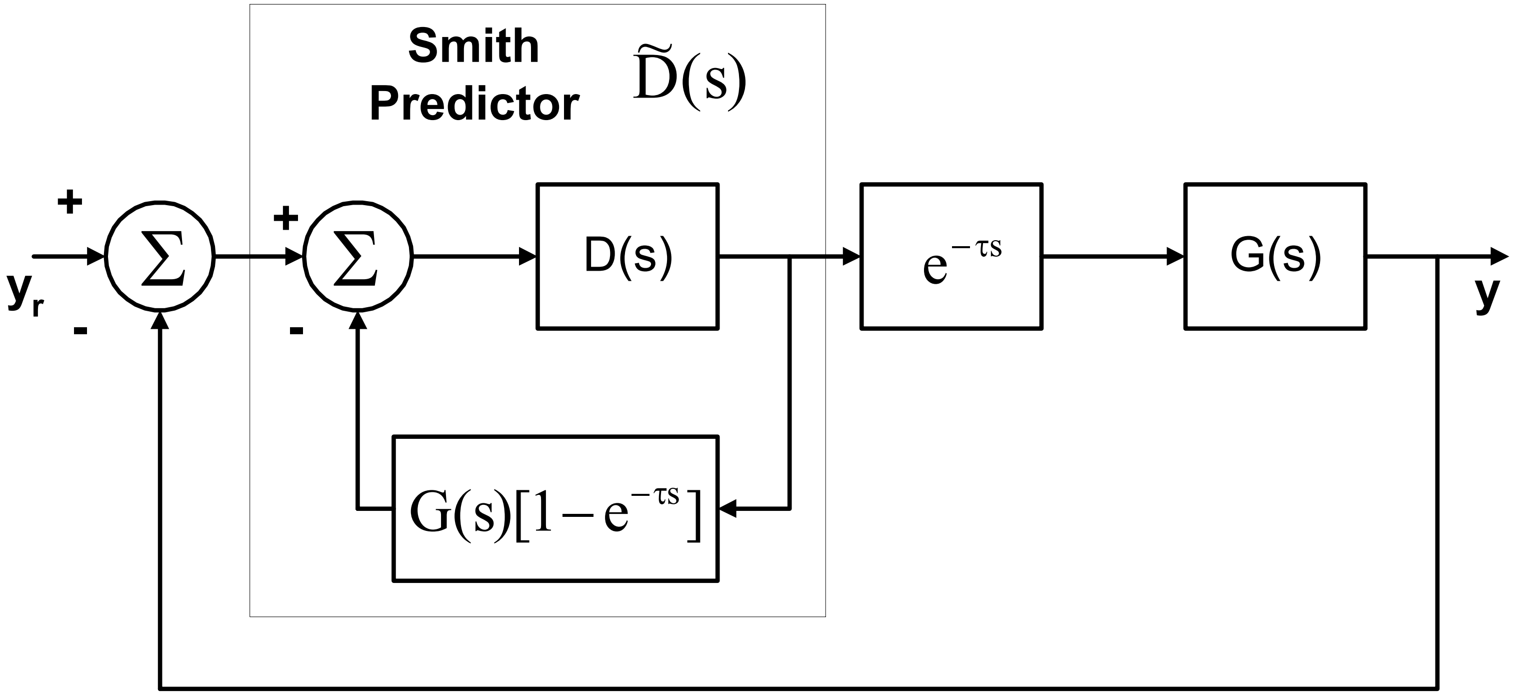
$$e^{-\tau_{dt} s} = 1 \angle -\omega \tau_{dt}$$

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

- 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



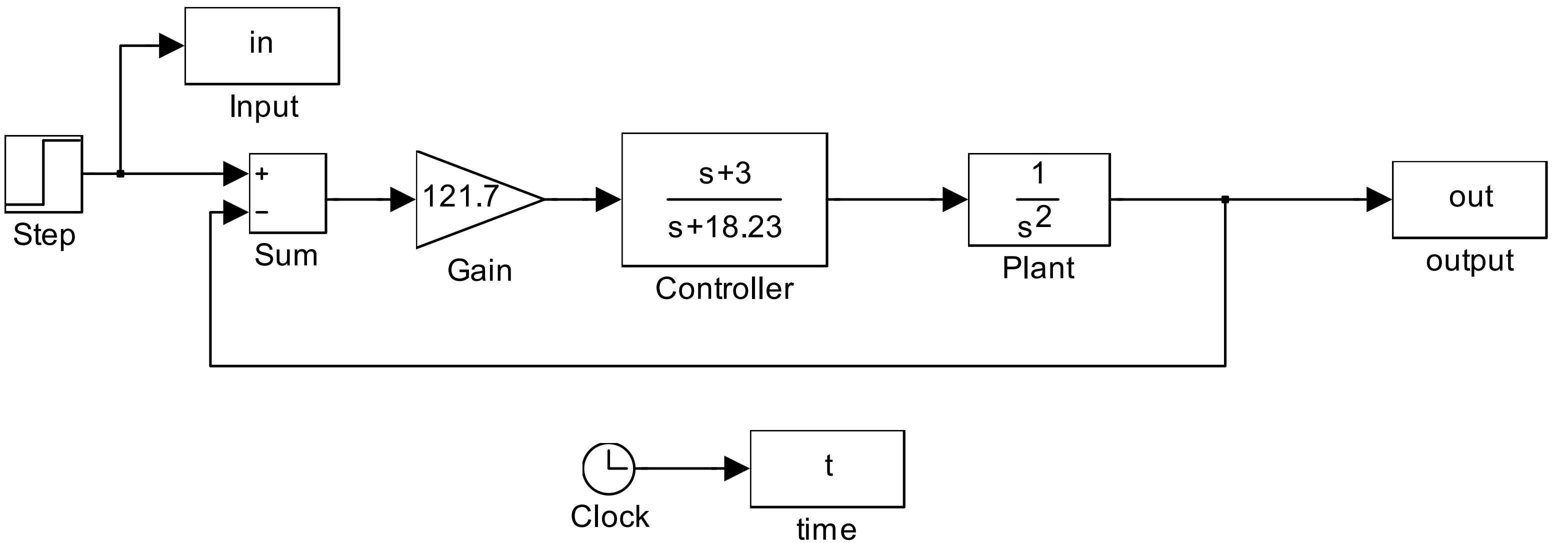
$$\tilde{D}(s) = \frac{D(s)}{1+(1-e^{-\tau s})D(s)G(s)} \qquad \frac{y}{y_r} = \frac{\tilde{D}(s)G(s)e^{-\tau s}}{1+\tilde{D}(s)G(s)e^{-\tau s}} = \frac{D(s)G(s)}{1+D(s)G(s)}e^{-\tau s}$$

- 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.

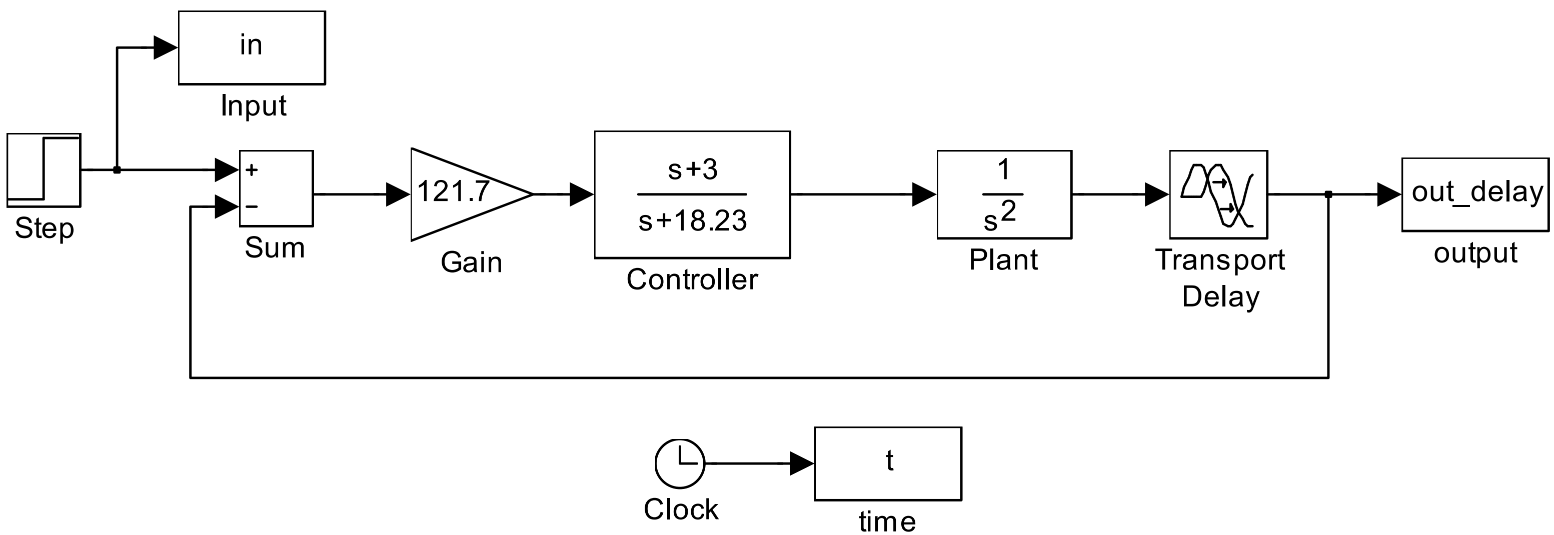
- <u>Implementation Issues</u>
  - You must know the plant transfer function and the time delay with reasonable accuracy.
  - You need a method of realizing the pure time delay that appears in the feedback loop, e.g., Pade approximation:

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

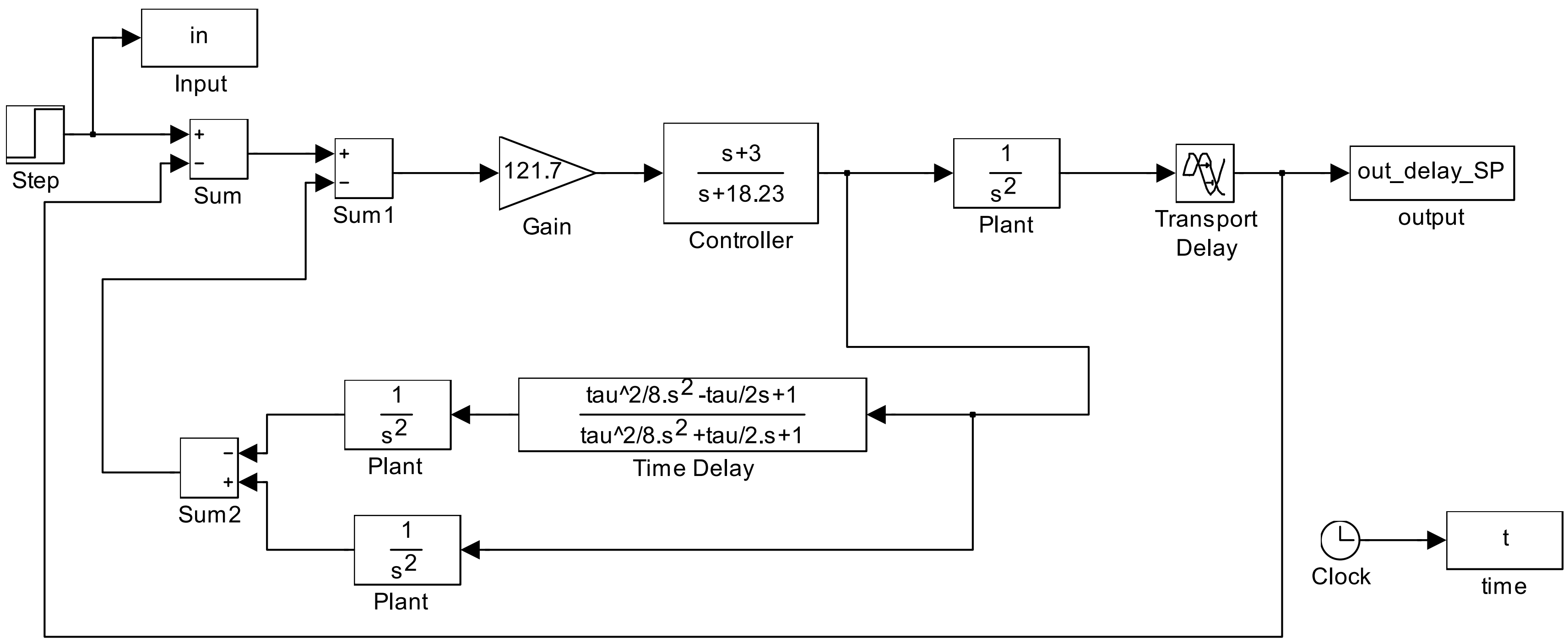
Example Problem



Basic Feedback Control System with Lead Compensator

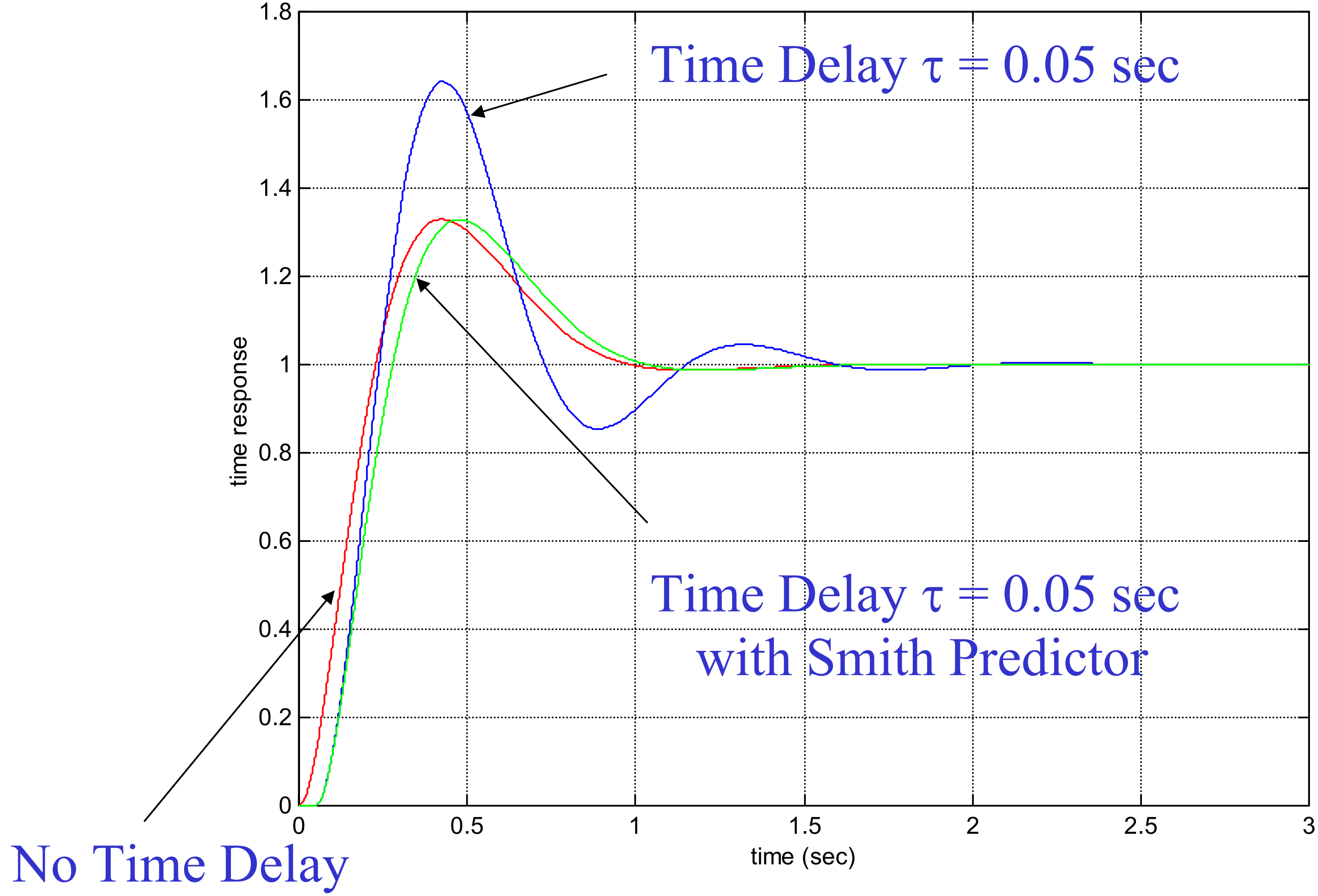


Basic Feedback Control System with Lead Compensator BUT with Time Delay $\tau = 0.05$ sec



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

System Step Responses



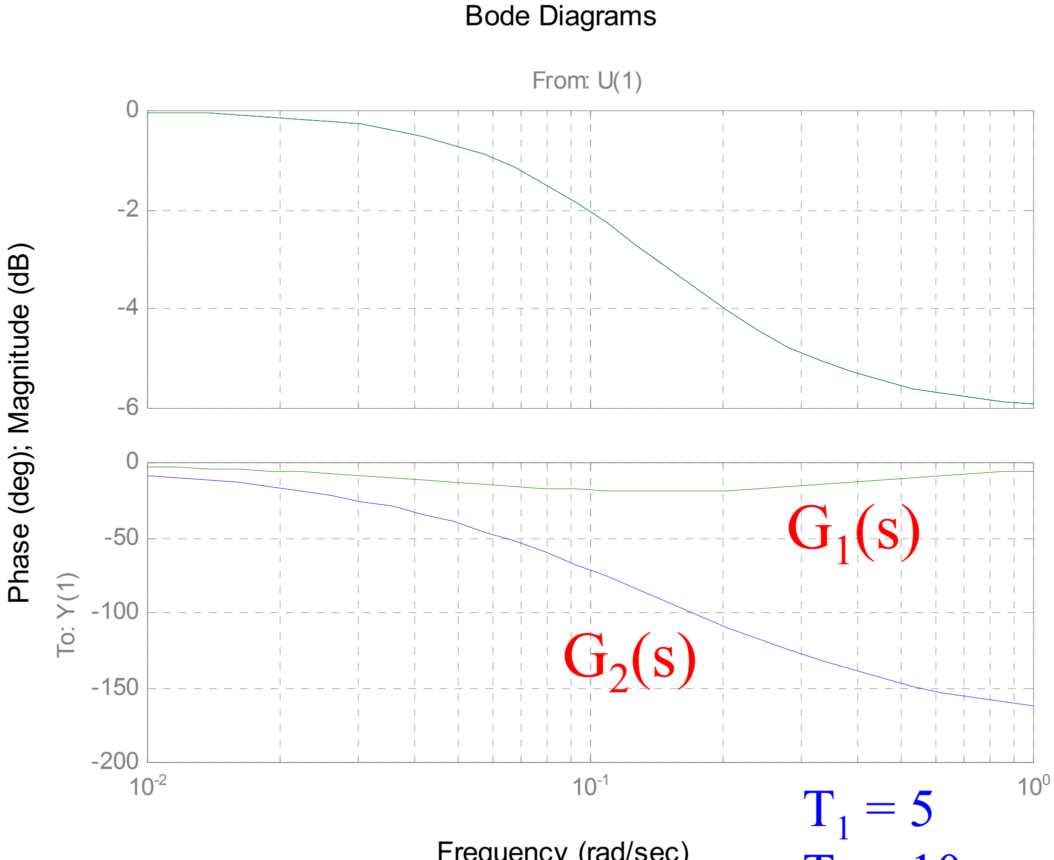
- Comments
  - The system with the Smith Predictor tracks reference variations with a time delay.
  - The Smith Predictor minimizes the effect of the DT on stability as model mismatching is bound to exist. This however still allows tighter control to be used.
  - What is the effect of a disturbance? 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:

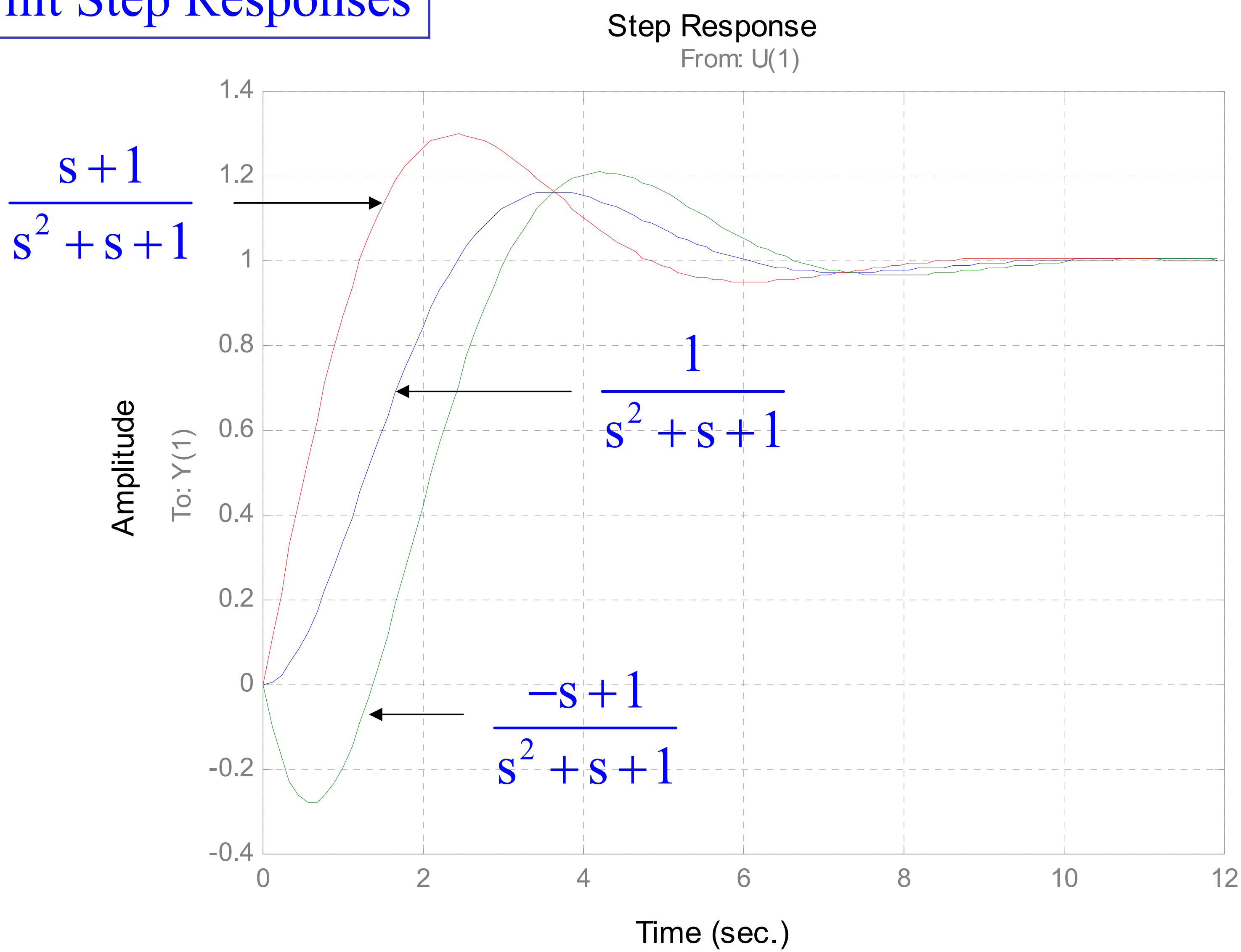
$$G(s) = \frac{1 - T_1 s}{1 + T_1 s}$$

- 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.

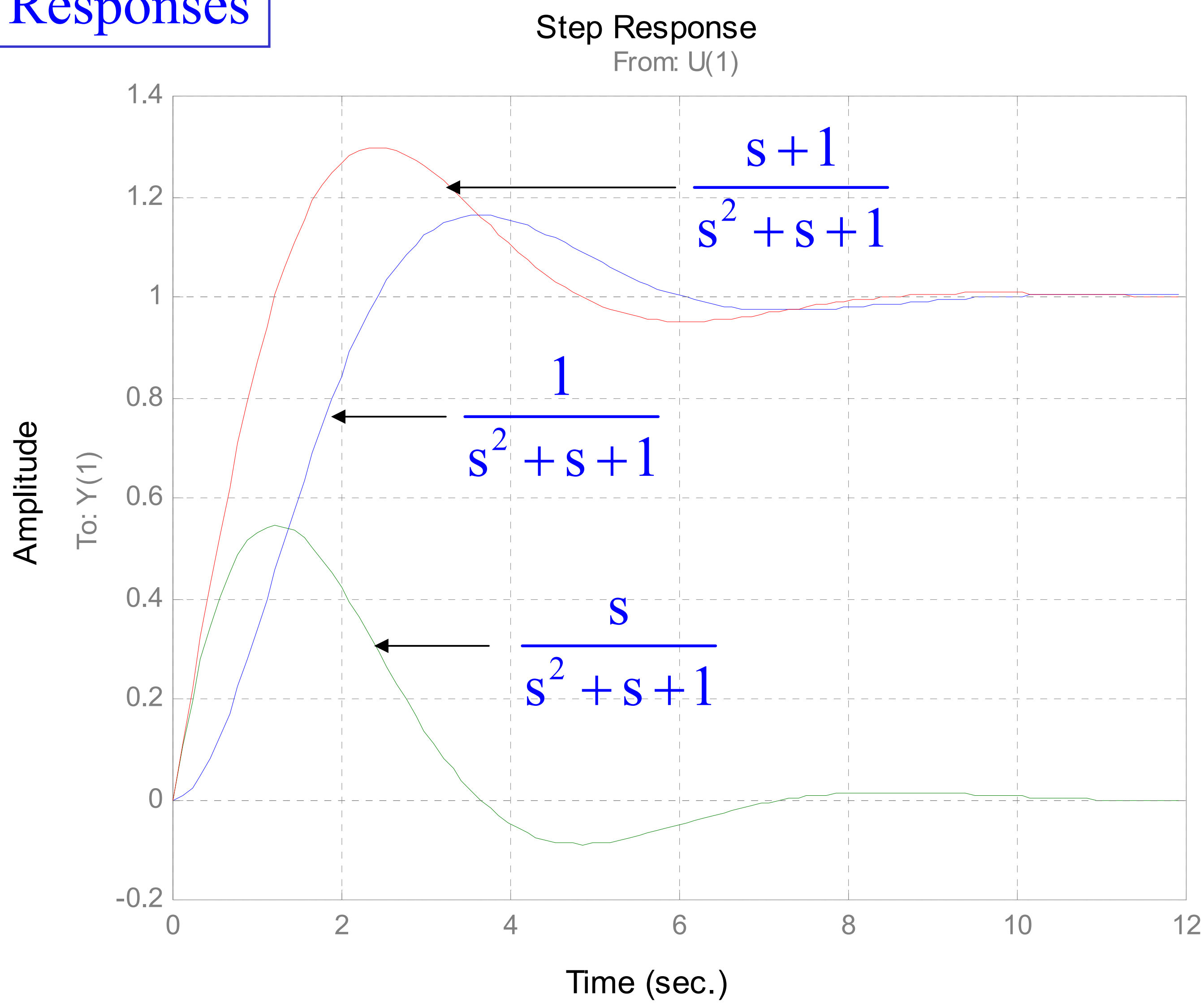
- This does not hold for a nonminimum-phase system.
- Nonminimum-phase systems may arise in two different ways:
  - When a system includes a nonminimum-phase element or elements
  - When there is an unstable minor loop
- For a minimum-phase system, the phase angle at $\omega = \infty$ becomes $-90°(q - p)$, 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: $e^{-\tau_{dt}s} = 1\angle - \omega\tau_{dt}$

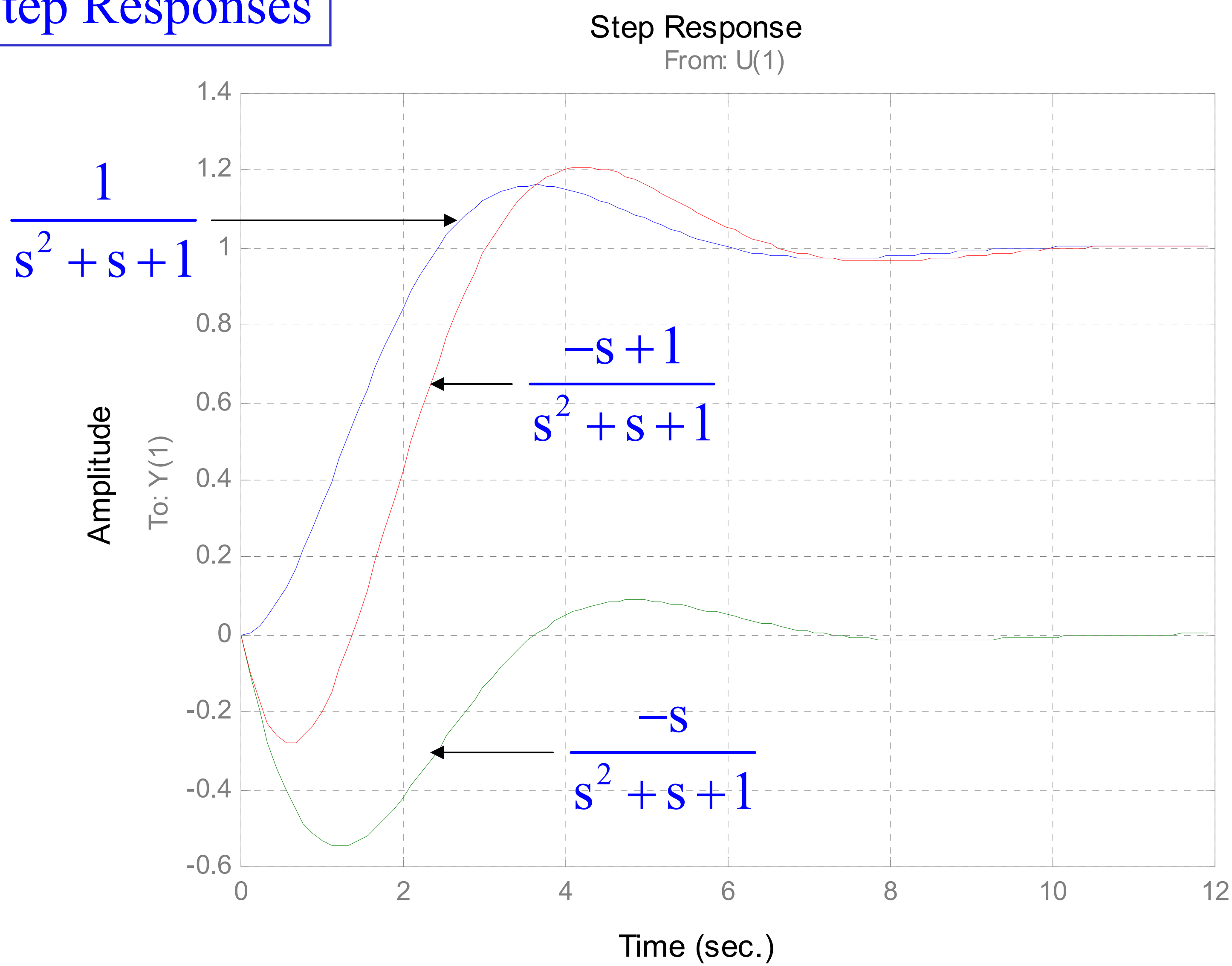
Unit Step Responses



Unit Step Responses



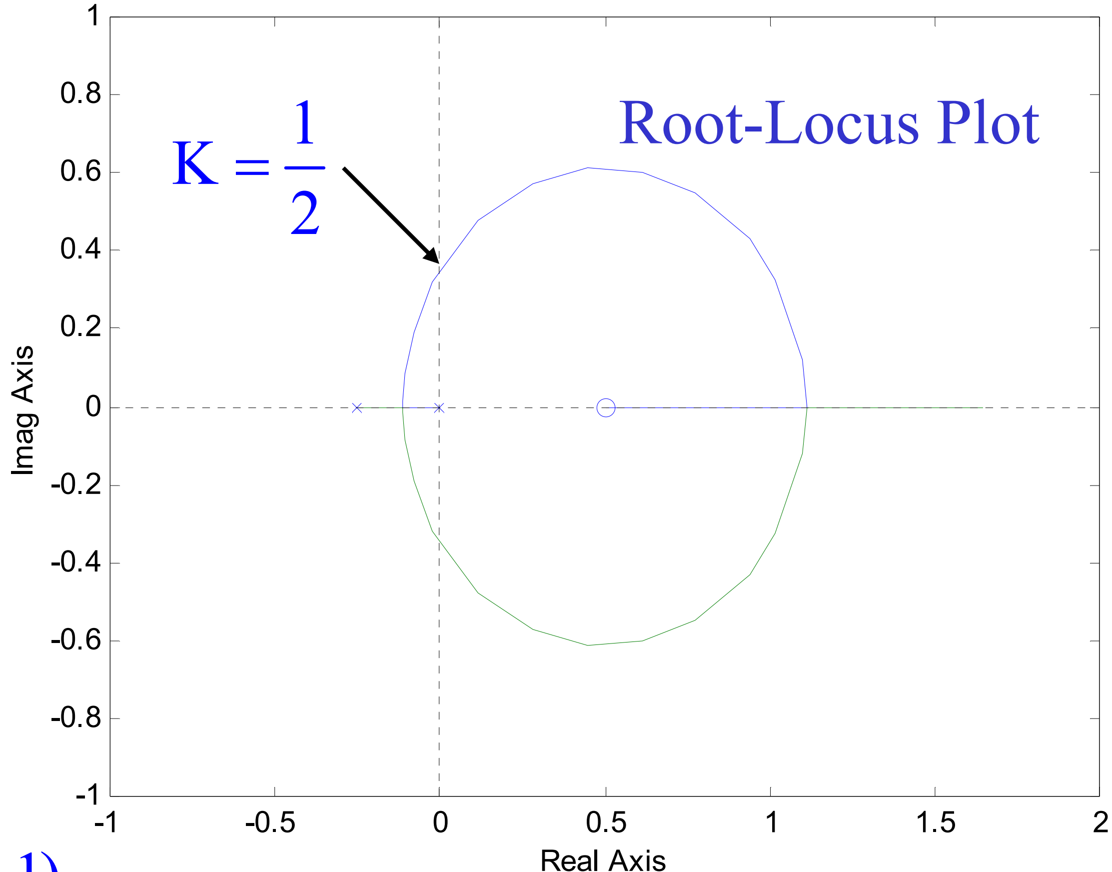
Unit Step Responses



- <u>Nonminimum-Phase Systems: Root-Locus View</u>
  - 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:

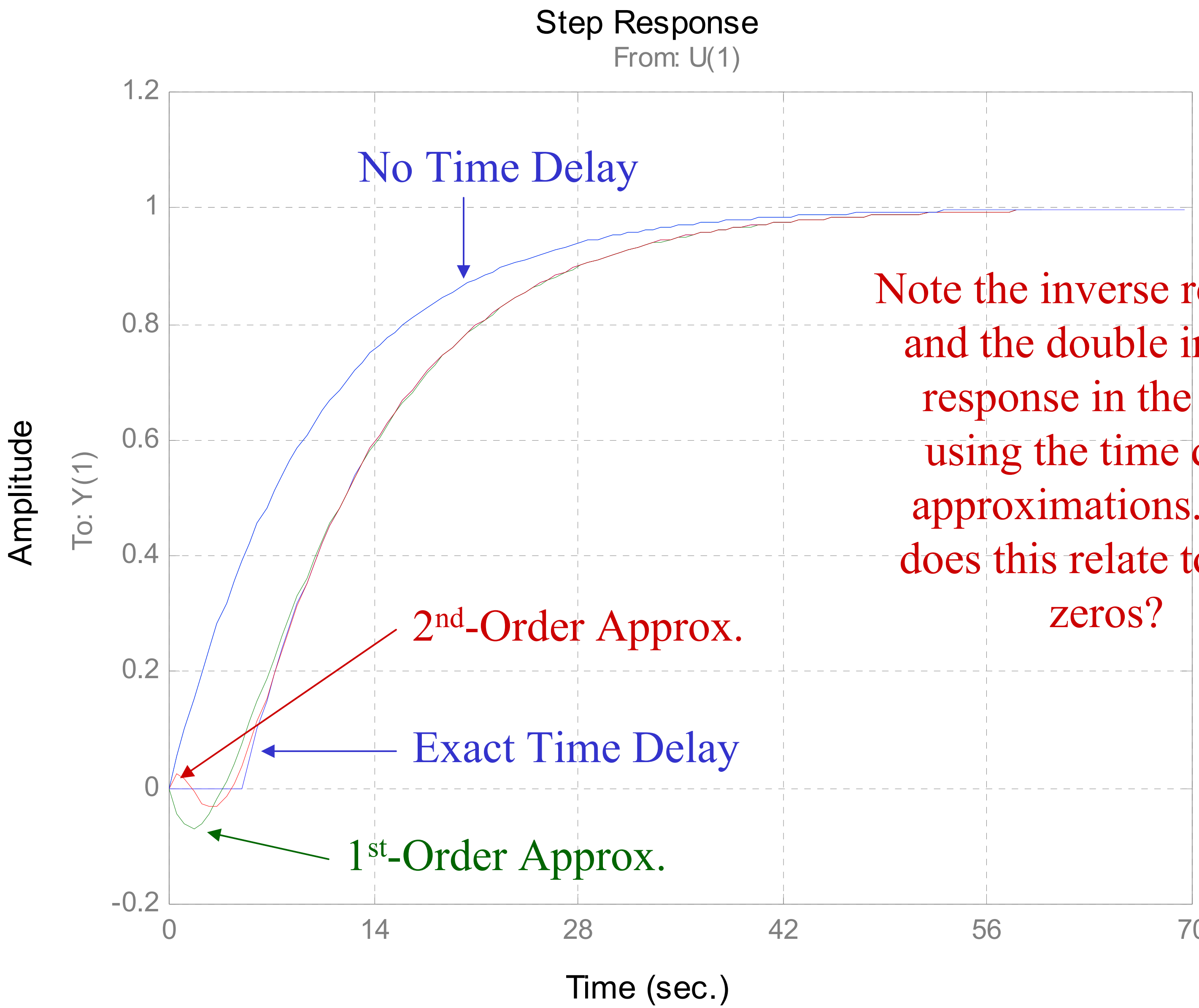
$$\angle G(s)H(s) = \angle \frac{-K(2s-1)}{s(4s+1)}$$

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

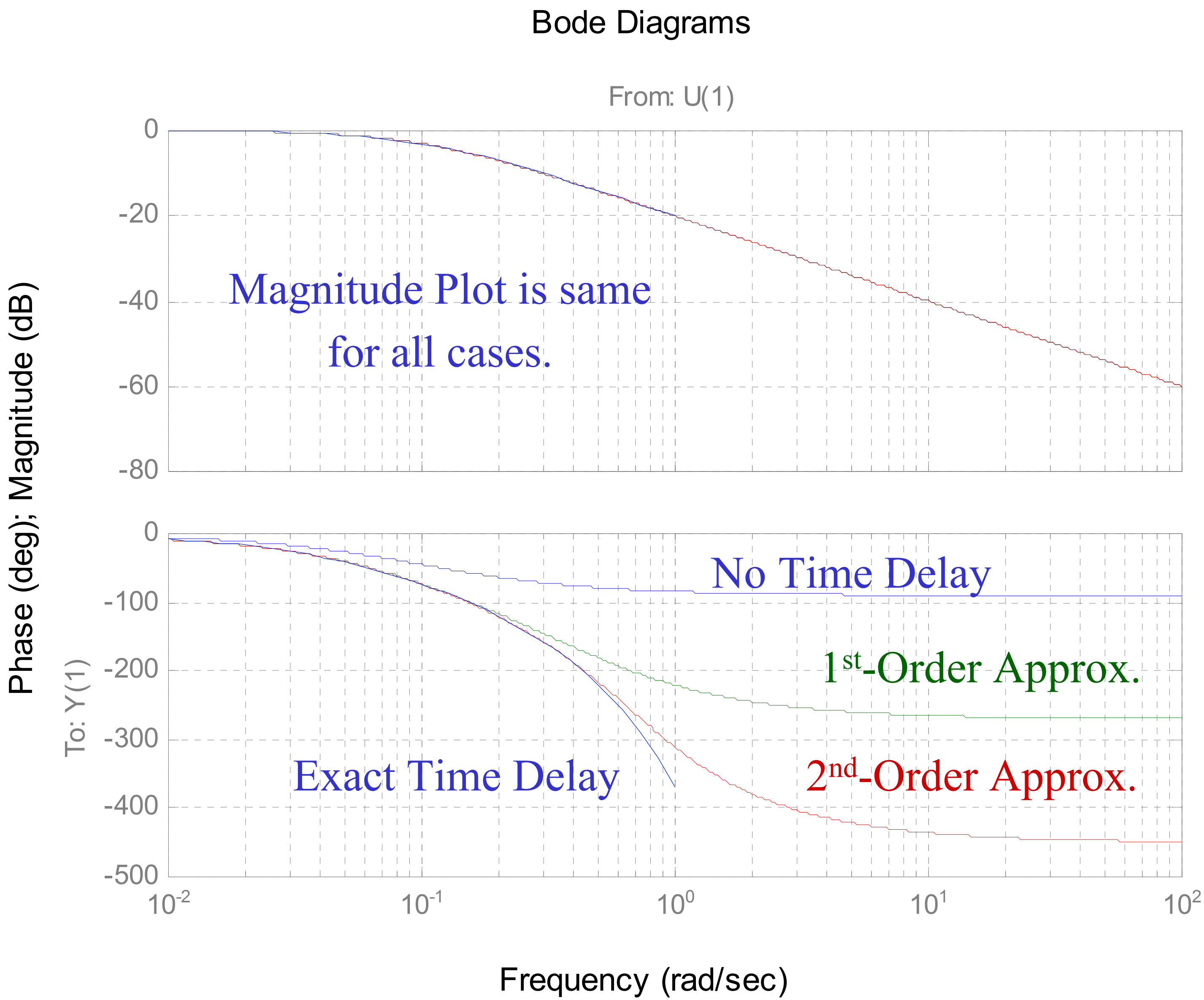
- Dynamic Response of a First-Order System with a Time Delay
  - The transfer function of a time delay combined with a first-order process is:

$$\frac{Ke^{-\tau_{DT}s}}{\tau s+1}$$

  - Consider the case with: $K = 1$, $\tau = 10$, $\tau_{DT} = 5$, and a unit step input at $t = 0$.
  - Simulate the step response with:
    - An exact representation of a time delay
    - A first-order Pade approximation of a time delay
    - A second-order Pade approximation of a time delay
  - Simulate the frequency response for the same cases.



Note the inverse response and the double inverse response in the plots using the time delay approximations. How does this relate to RHP zeros?


Bode Diagrams
From: U(1)
0
-20
-40
-60
-80
Magnitude Plot is same
for all cases.
0
-100
-200
-300
-400
-500
No Time Delay
1st-Order Approx.
Exact Time Delay
2nd-Order Approx.
Phase (deg); Magnitude (dB)
To: Y(1)
10-2
10-1
100
101
102
Frequency (rad/sec)

- Empirical Model
  - The most common plant test used to develop an empirical model is to make a step change in the manipulated input and observe the measured process output response.
  - Then a model is developed to provide the best match between the model output and the observed plant output.
  - Important Issues:
    - Selection of the proper input and output variables.
    - In step-response testing, we first bring the process to a consistent and desirable steady-state operating point, then make a step change in the input variable.
    - What should the magnitude of the step change be?

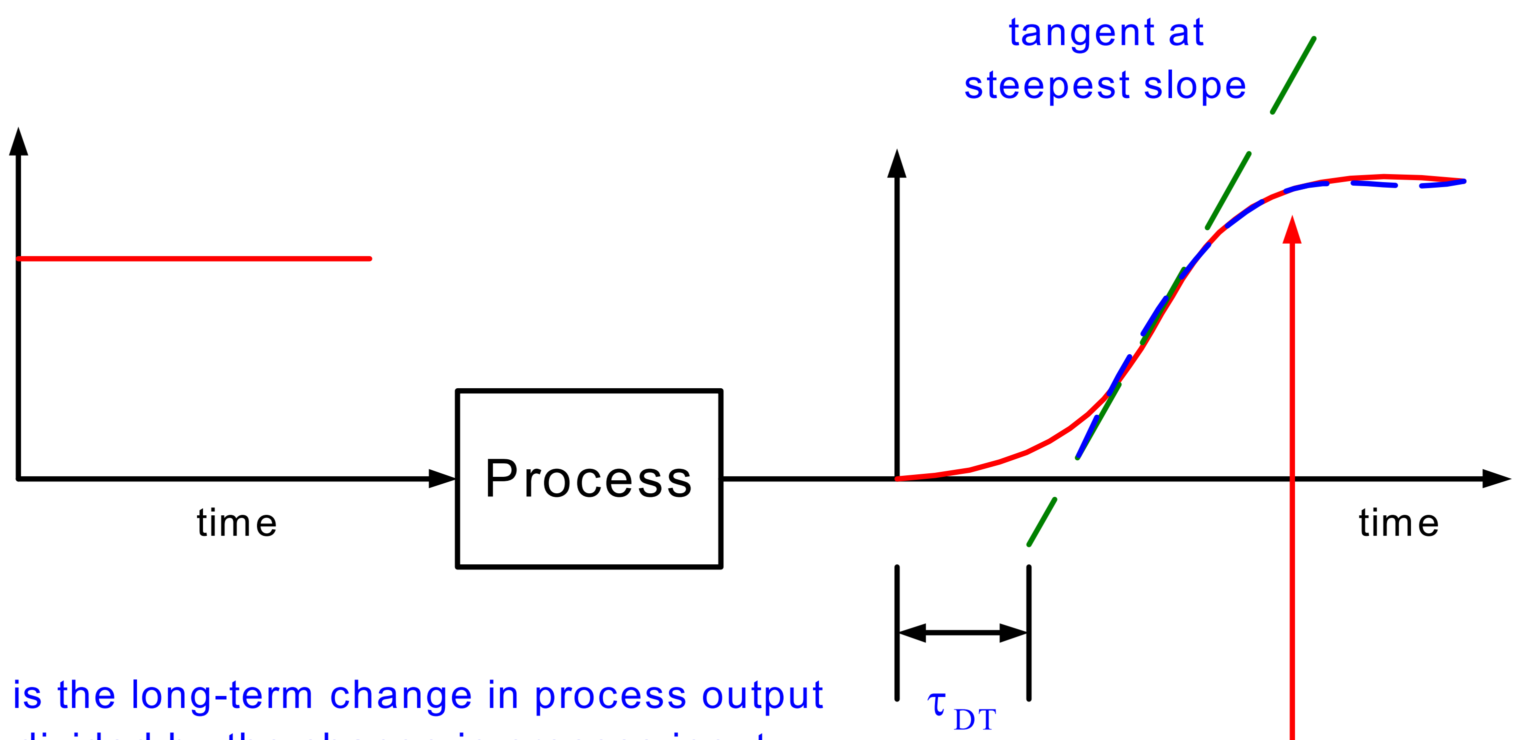
1. The magnitude of the step input must be large enough so that the output signal-to-noise ratio is high enough to obtain a good model.
2. If the magnitude of the step input is too large, nonlinear effects may dominate.

- Clearly there is a trade off.

– By far the most commonly used model for control-system design purposes, is the 1st-order plus time delay model.

$$\frac{Ke^{-\tau_{DT}s}}{\tau s+1}$$

– The three process parameters can be estimated by performing a single step test on the process input.



K is the long-term change in process output divided by the change in process input.

Estimate time constant from a semi-log plot of first-order response.

- If the process is already in existence, experimental step tests allow measurement of τ_{DT} and τ.
- At the process design stage, theoretical analysis allows estimation of these numbers if the process is characterized by a cascade of known 1st-order lags.
- Approximate the dead time with a 1st-order Pade approximation:

$$\frac{2-\tau_{DT}s}{2+\tau_{DT}s}$$

- Consider the open-loop transfer function:

$$\frac{Ke^{-\tau_{DT}s}}{\tau s+1} \approx \frac{K\left(\dfrac{2-\tau_{DT}s}{2+\tau_{DT}s}\right)}{\tau s+1} = G$$

- The closed-loop system transfer function is:

$$\frac{C}{V} = \frac{G}{1+G}$$

- The characteristic equation of the closed-loop system is:

$$1 + G(s) = 0$$

$$1 + \frac{K\left(\dfrac{2-\tau_{DT}s}{2+\tau_{DT}s}\right)}{\tau s + 1} = 0$$

$$\tau_{DT}\tau s^2 + \left(2\tau + \tau_{DT} - K\tau_{DT}\right)s + 2\left(K+1\right) = 0$$

- For what value of K will this system go unstable?

- The Routh Stability Criterion predicts that for stability:

$$-1 < K < 2\left(\frac{\tau}{\tau_{DT}}\right) + 1$$

- The gain value for marginal stability can be found precisely from the Nyquist criterion since we know the frequency response of a dead time exactly. For marginal stability, we require that (B/E)(iω) go precisely through the point $-1 = 1\angle 180°$. The phase angle part of the requirement can be stated as:

$$-\pi = -\omega_0 \tau_{DT} - \tan^{-1} \omega_0 \tau$$

- This fixes (for a given τ τ_{DT}) the frequency ω_0 at which (B/E)(iω) passes through the point $-1 = 1\angle 180°$.

– This equation has no analytical solution. Once ω_0 is found numerically, the gain K for marginal stability is obtained by requiring that:

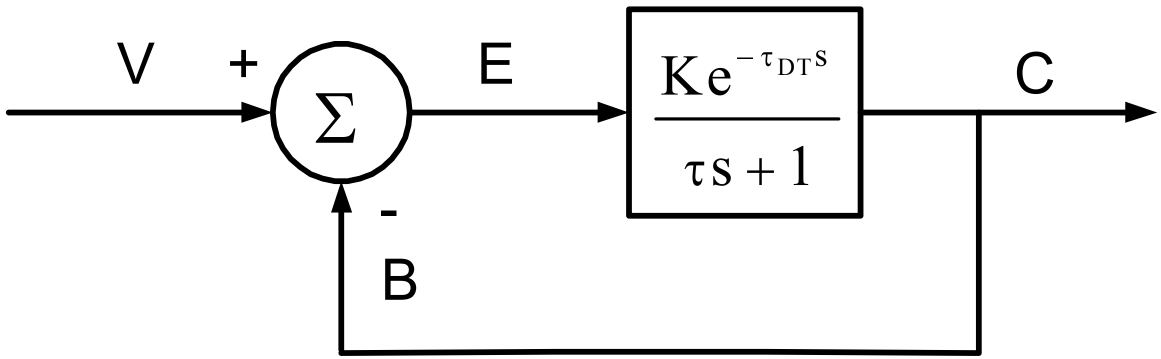
$$\left|\frac{B}{E}(i\omega)\right| = \frac{K}{\sqrt{(\omega_0\tau)^2 + 1}} = 1.0$$

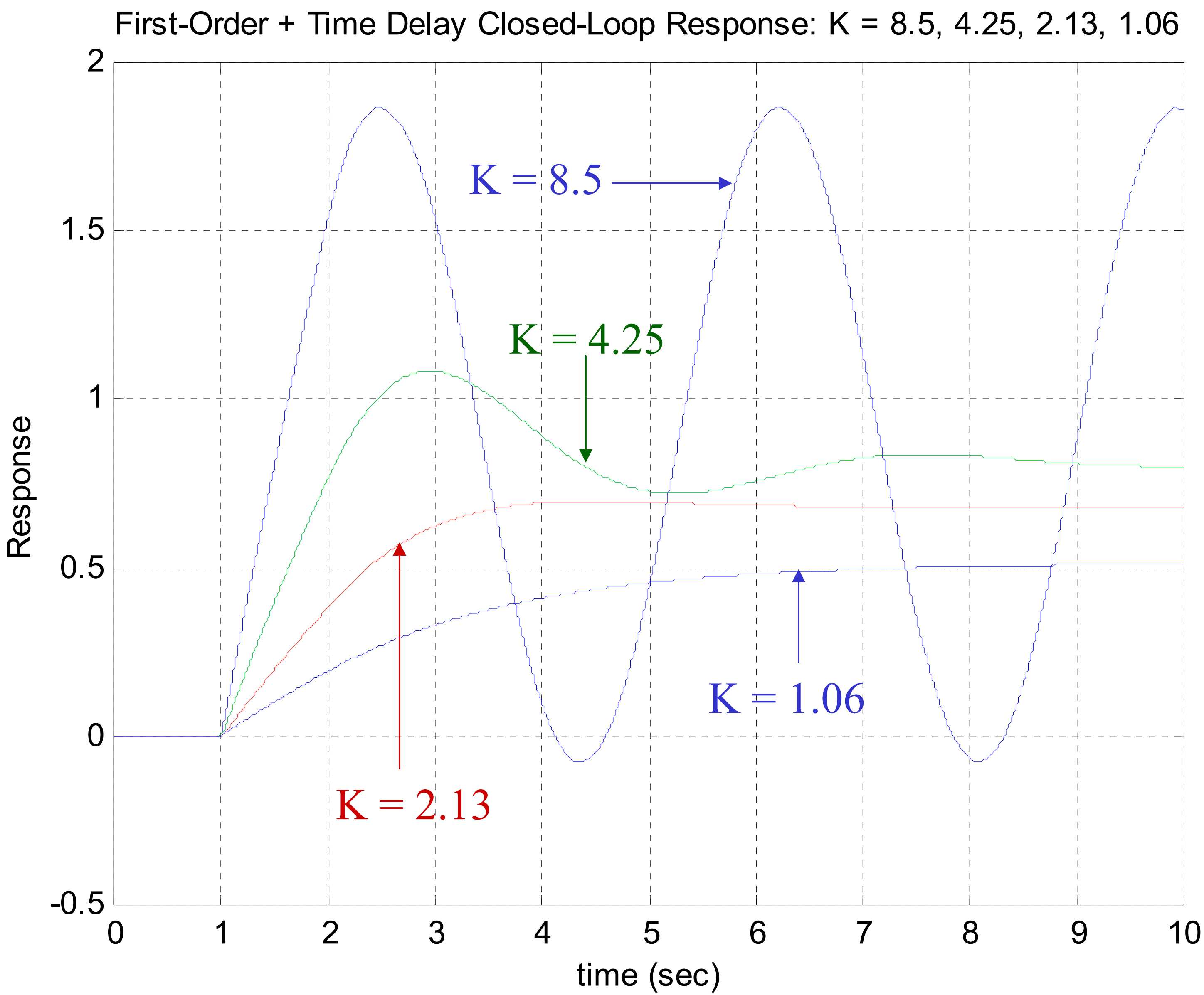
– A table shows results for a range of the most common values encountered for τ_{DT} / τ in modeling complex systems.

τ_{DT} / τ	$\omega_0\tau$	K
0.1	16.4	16.4
0.2	8.44	8.50
0.3	5.80	5.89
0.4	4.48	4.59
0.5	3.67	3.81
0.6	3.13	3.29
0.7	2.74	2.92
0.8	2.45	2.64
0.9	2.22	2.43
1.0	2.03	2.26

- The steady-state error is typical of proportional control. Design values of K must be less than those for marginal stability.
- A design criterion sometimes used in industrial process control is quarter-amplitude damping, wherein each cycle of transient oscillation is reduced to ¼ the amplitude of the previous cycle.
- A useful approximation for this behavior is a gain margin of 2.0 for the frequency response.
- If we apply this to the table of results for, say, $\tau_{DT} / \tau = 0.2$, we get a design gain value of 4.25, giving large steady-state errors.
- For this reason, processes of this type often use integral or proportional + integral control, which reduces steady-state errors without requiring large K values.

- <u>Exercise</u>:
  - For the closed-loop system below, evaluate the step response using:
    - $\tau_{DT} = 1$ sec
    - $\tau = 5$ sec
    - $K = 8.5, 4.25, 2.13, 1.06$

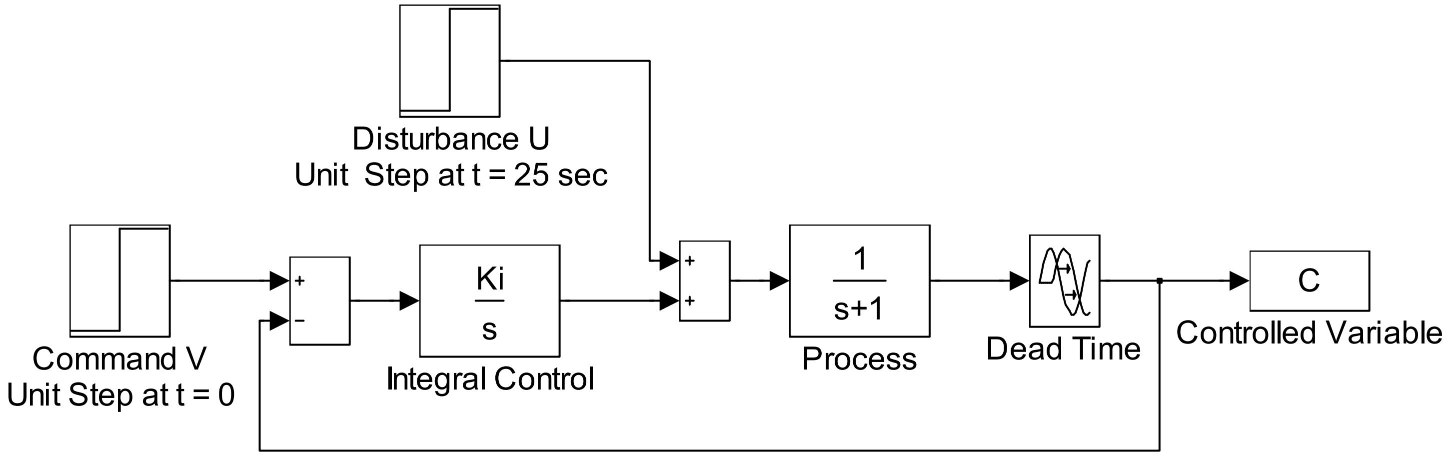



First-Order + Time Delay Closed-Loop Response: K = 8.5, 4.25, 2.13, 1.06
Response
time (sec)
K = 8.5
K = 4.25
K = 2.13
K = 1.06
2
1.5
1
0.5
0
-0.5
0
1
2
3
4
5
6
7
8
9
10

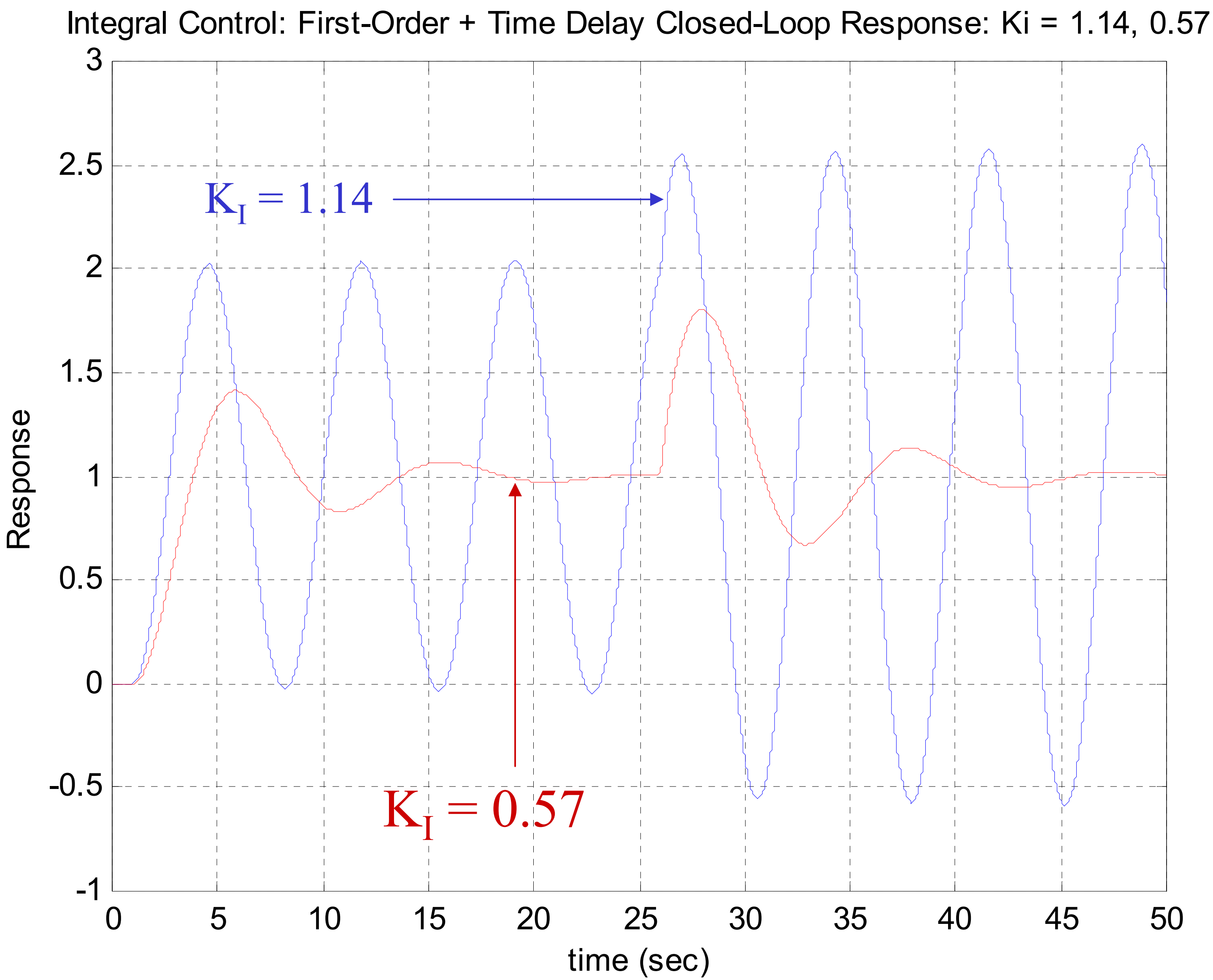
- <u>Consider Integral Control of a First-Order Process plus a Dead Time</u>
  - Proportional control was found to be difficult since loop gain was restricted by stability problems to low values, causing large steady-state error.
  - Integral control gives zero steady-state error (for both step commands and/or disturbances) for any loop gain and is thus an improvement.
  - The values of K for marginal stability are given in the following table.
  - Compared with proportional control, both loop gain and speed of response (ω_0 for a given τ) are lower. However, we do not depend on it to reduce steady-state error.

τ_{DT} / τ	$\omega_0 \tau$	K
0.1	3.11	10.2
0.2	2.16	5.16
0.3	1.74	3.49
0.4	1.48	2.65
0.5	1.31	2.15
0.6	1.18	1.81
0.7	1.07	1.57
0.8	0.99	1.39
0.9	0.92	1.25
1.0	0.86	1.14

- <u>Check Time-Domain Response</u>



  - Run simulations on the system for K_I = 1.14 (marginal stability) and for K_I = 0.57 (gain margin of 2.0).
  - Check response of C to both step inputs in V and U.
  - Note the well-damped response with zero steady-state error for both step commands and disturbances for K_I = 0.57.


Integral Control: First-Order + Time Delay Closed-Loop Response: Ki = 1.14, 0.57
Response
time (sec)
K_I = 1.14
K_I = 0.57
3
2.5
2
1.5
1
0.5
0
-0.5
-1
0
5
10
15
20
25
30
35
40
45
50