Real Time Systems and Control Applications

Contents

NPCS → PIP

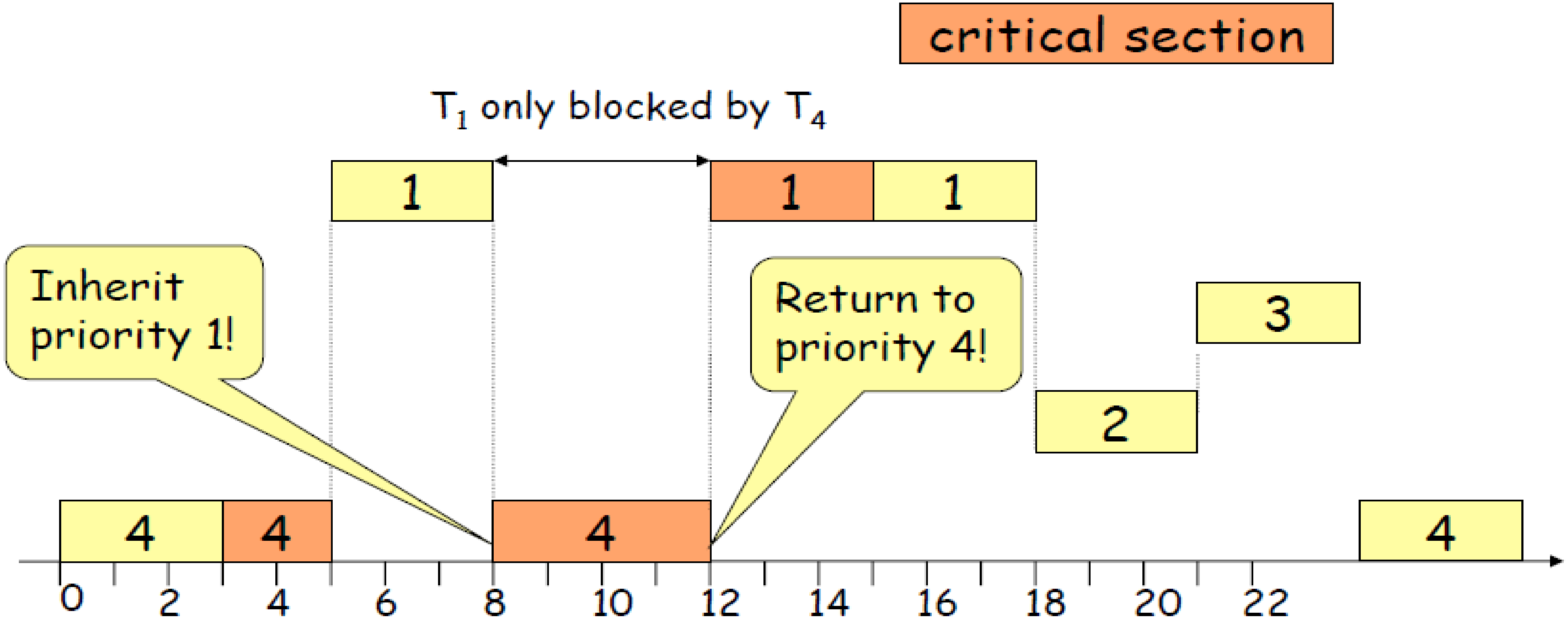
- NPCS: While a task holds or locks a resource it executes at a priority higher than the priorities of all tasks, and a task recovers its priority when resources are unlocked.

- Disadvantage of NPCS: A task can be blocked by a lower priority task, even without a resource conflict

- Priority Inheritance Protocol (PIP): Increase the priorities only upon resource contention

Solution --- Priority Inheritance

- Under priority inheritance the priority of the task that is in the critical section inherits the priority of a higher priority task when the higher priority task requests the resource.



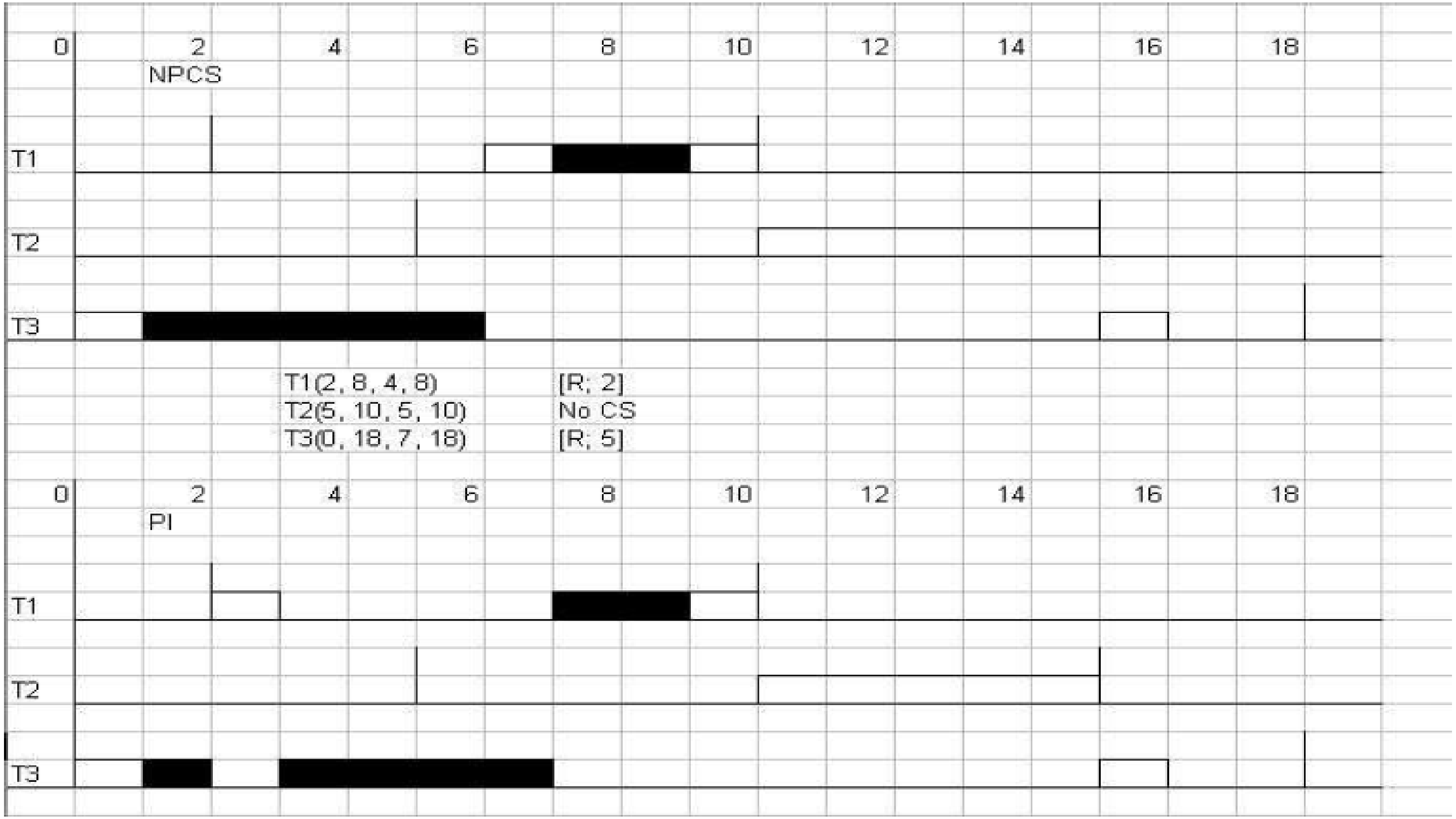
Priority Inheritance Rule

- When a task T1 is blocked due to non availability of a resource that it needs, the task T2 that holds the resource and consequently blocks T1, and T2 inherits the current priority of task T1.

- T2 executes at the inherited priority until it releases R.

- Upon the release of R, the priority of T2 returns to the priority that it held when it acquired the resource R.

NPCS v.s. PIP

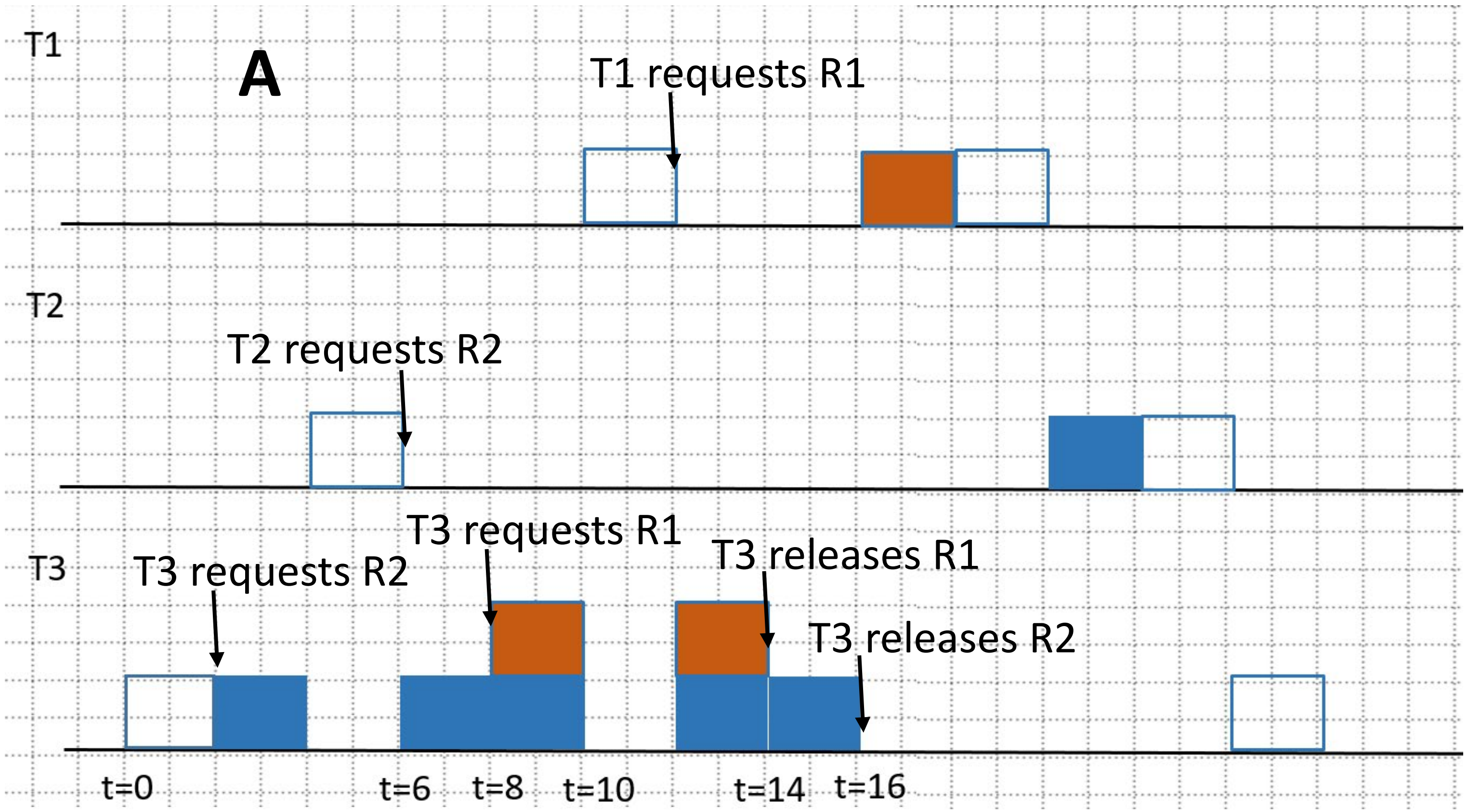


0 2 4 6 8 10 12 14 16 18
NPCS
T1
T2
T3
T1(2, 8, 4, 8)
T2(5, 10, 5, 10)
T3(0, 18, 7, 18)
[R; 2]
No CS
[R; 5]
0 2 4 6 8 10 12 14 16 18
PI
T1
T2
T3

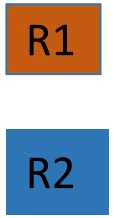
Priority Inheritance Protocol (PIP)

- If lower priority task TL blocks a higher priority task TH, priority(TL) ← priority(TH)
- When TL releases a resource, it returns to its normal priority if it doesn't block any task. Or it returns to the highest priority of the tasks waiting for a resource held by TL (example on next page)
- Transitive
  - T1 blocked by T2: priority(T2) ← priority(T1)
  - T2 blocked by T3: priority(T3) ← priority(T1)

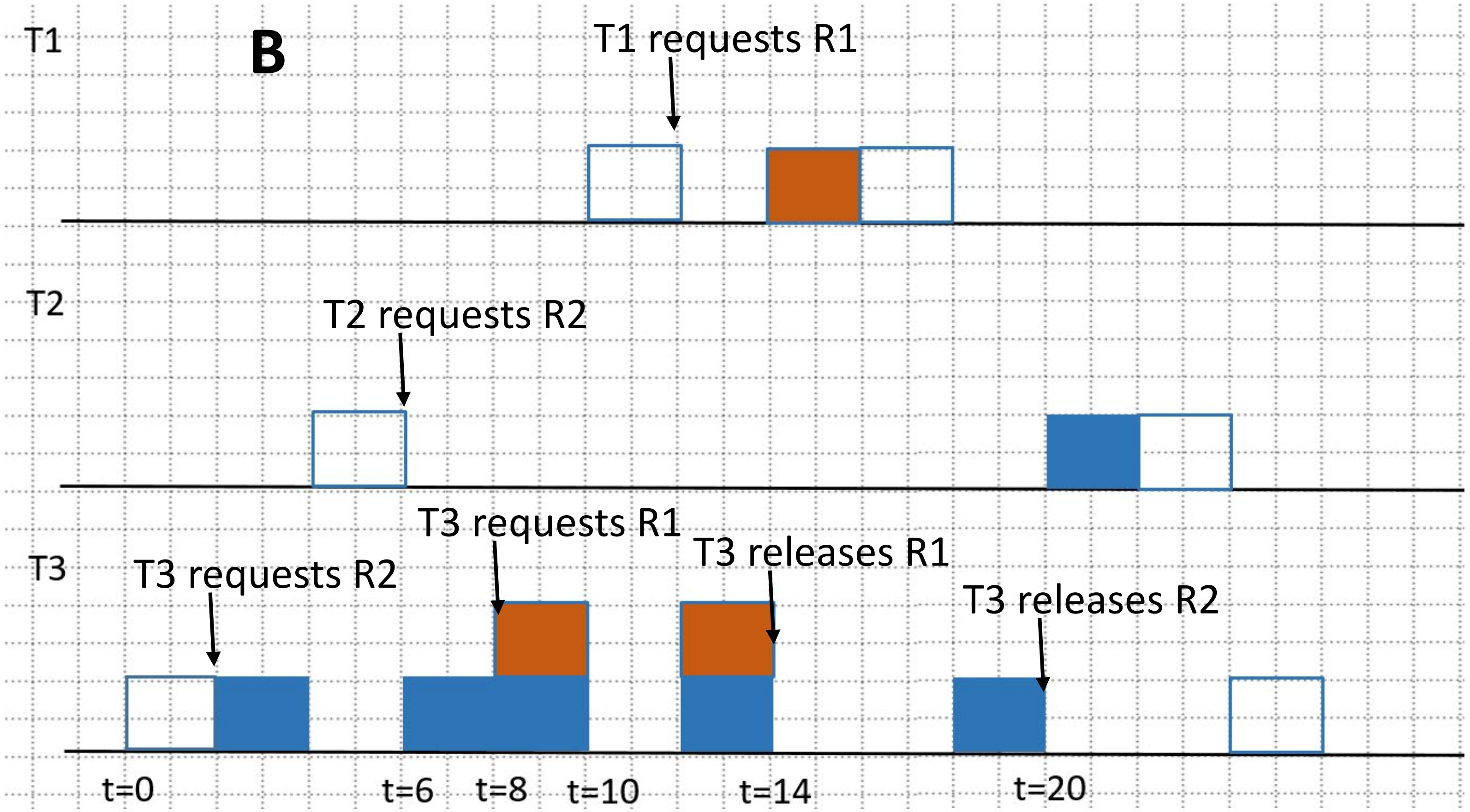
Which one of the scheduling output is correct? Assume T1 has priority high, T3 has priority low, and T2 has priority medium.



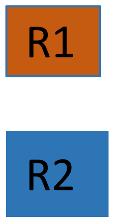
Resources R1 and R2.



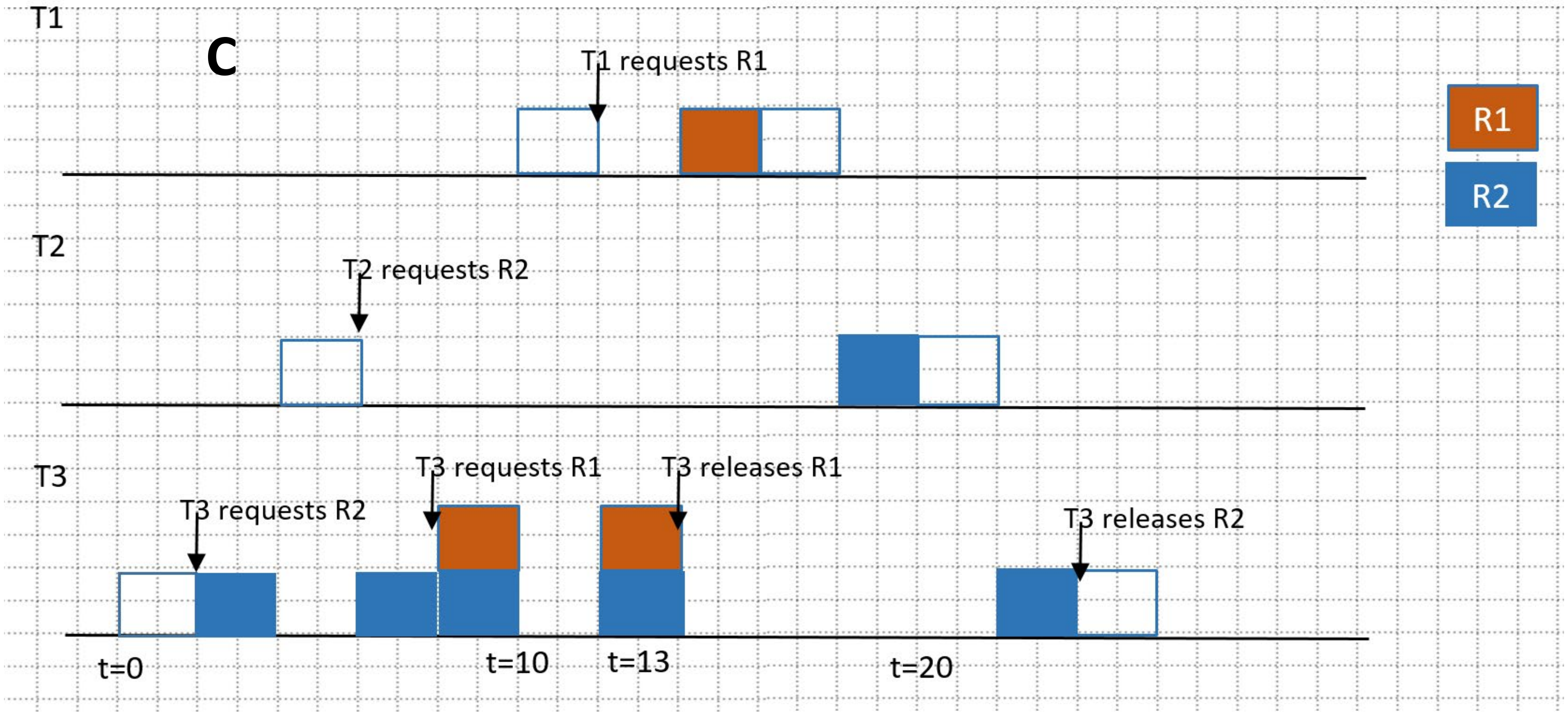
Which one of the scheduling output is correct? Assume T1 has priority high, T3 has priority low, and T2 has priority medium.



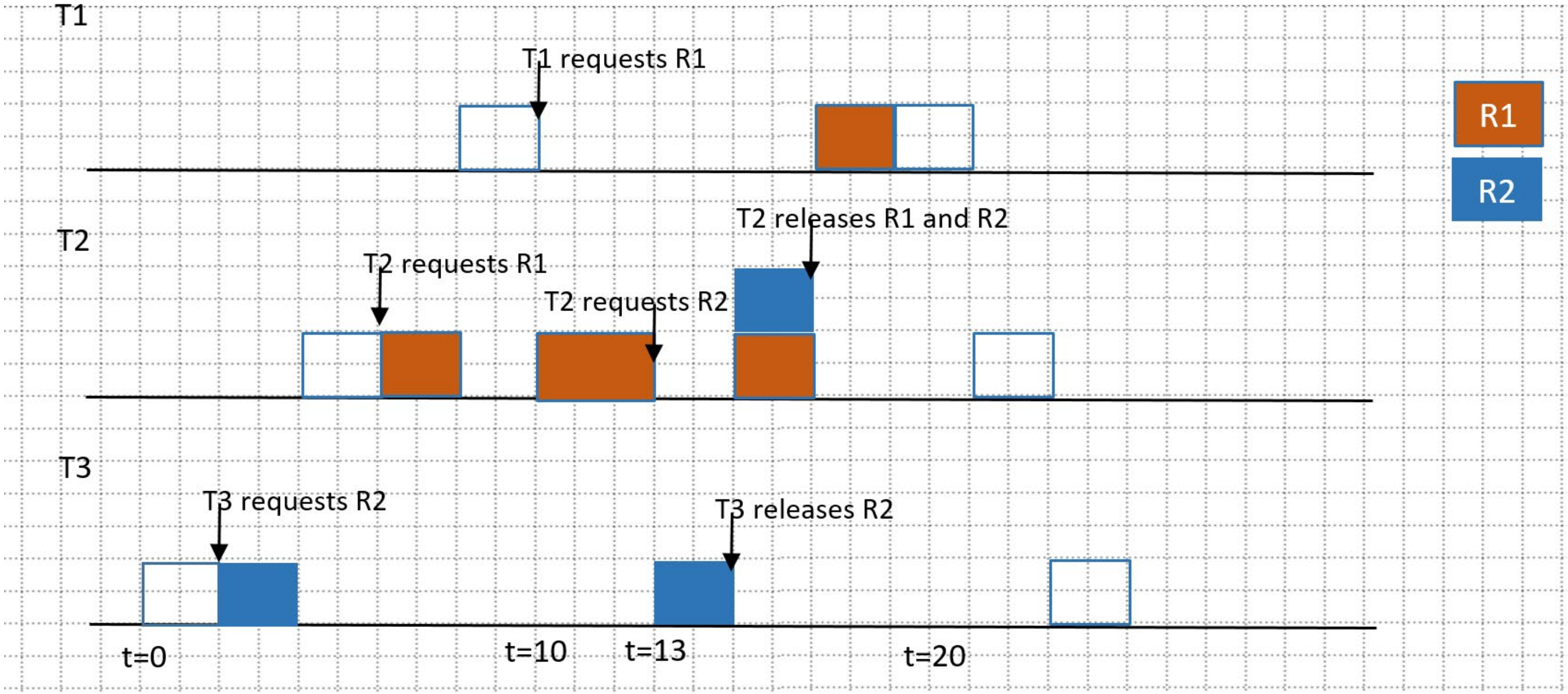
Resources R1 and R2.



Which one of the scheduling output is correct? Assume T1 has priority high, T3 has priority low, and T2 has priority medium.



Example of Transitive Inheritance



At t=10, T2 inherits T1’s priority, so Pr(T2)=H

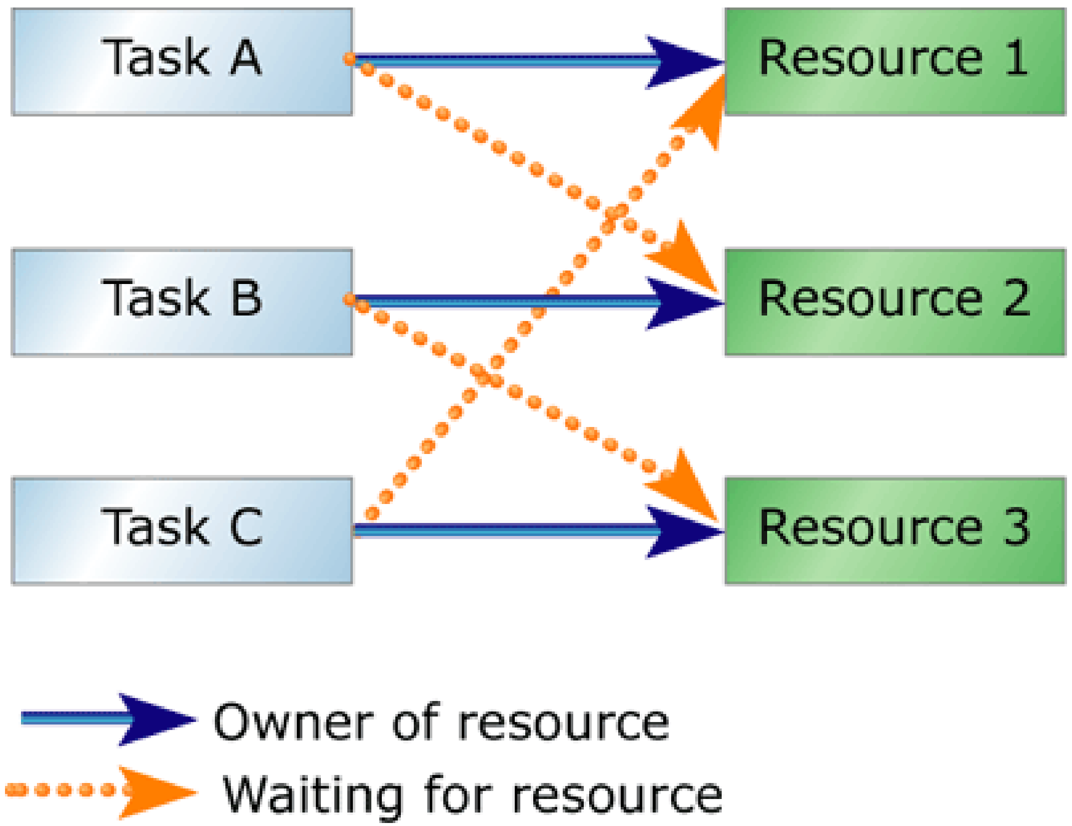
At t=13, T3 inherits T2’s priority, so Pr(T3)=H

Comments on PIP

- PIP avoids the disadvantage of NPCS
- PIP has most of the advantages of NPCS.
- However it does not avoid deadlocks.

- Question: Why PIP has the problem of deadlock?

Deadlock Scenario



What will happen in the scenario of deadlock?
(Assuming Task A has the highest priority and Task C has the lowest priority)

Task A requests R2 which is held by Task B, so B blocks A and inherits A's priority. Similarly, C blocks B and inherits B's priority. A blocks C and inherits C's priority. They will all set to be the highest priority, but still waiting for resources.

Priority Ceiling Protocol (PCP)

- Extend priority inheritance protocol to prevent deadlocks and to further reduce the blocking time.
- Assumptions:
  (1) Assigned priorities of all jobs are fixed
  (2) Resource requirements of all the tasks that will request a resource R is known

- **Priority ceiling of a resource R**:
  - ceiling(R) = highest priority among all the tasks that request R
  - Each resource has the fixed priority ceiling

- **Priority ceiling of a system**:
  - At any given time a set of resources are being used, the highest priority ceiling of the resource set is called the priority ceiling of the system.

Resource Allocation Rule in PCP

- When a task request resource R,
  - If R is held by another task, the request fails and the requesting task is blocked

  - If R is free then:
    1. If the requesting task's priority is higher than the current priority ceiling of the system, R is allocated to it.
    2. If the priority of the requesting task is not higher than the priority ceiling of the system, the request is denied and the task is blocked.

    Exception: If the requesting task is holding a resource $\mathrm{R_i}$ whose priority ceiling $\pi(\mathrm{R_i})$ is equal to the priority ceiling of the system, in which case the resource is allocated to the requesting task.

Priority Inheritance Rule in PCP

- When a task T1 gets blocked by T2, T2 inherits the priority of T1.

- T2 executes at the inherited priority until it releases the resource whose priority ceiling is equal to or higher than the inherited priority of T2. At this time, the priority of T2 resumes to the priority before T2 acquired the resource R.

More Comments About PCP

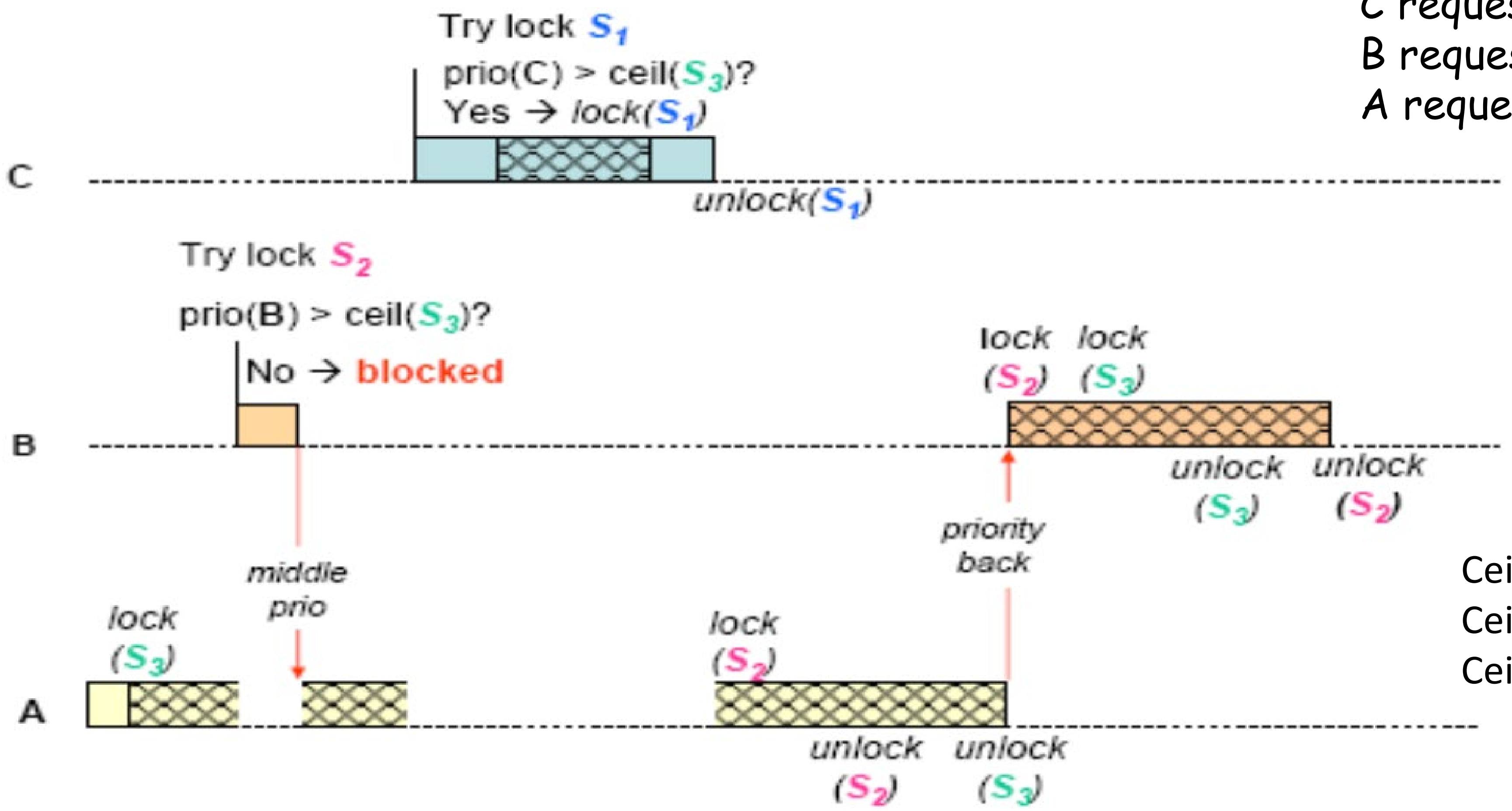
- Priority ceiling of a resource is fixed since we assume that we know all the resource requests. However, priority ceiling of the system is dynamically changed.

- When the priority of a task is updated?
  - When a task (T1) is blocked by another (T2) , we know that T2 is in its critical section which holds or will request a certain resource required by T1, so T2's priority is updated to the priority of T1.
  - When a task (T2) releases a resource, T2 resumes its priority.

- Only when a task has higher priority than the system's priority ceiling, the task will acquire a certain resource if it is available.

PCP

Assume C has a high priority (H), A has a low priority (L), and B has a medium priority (M).



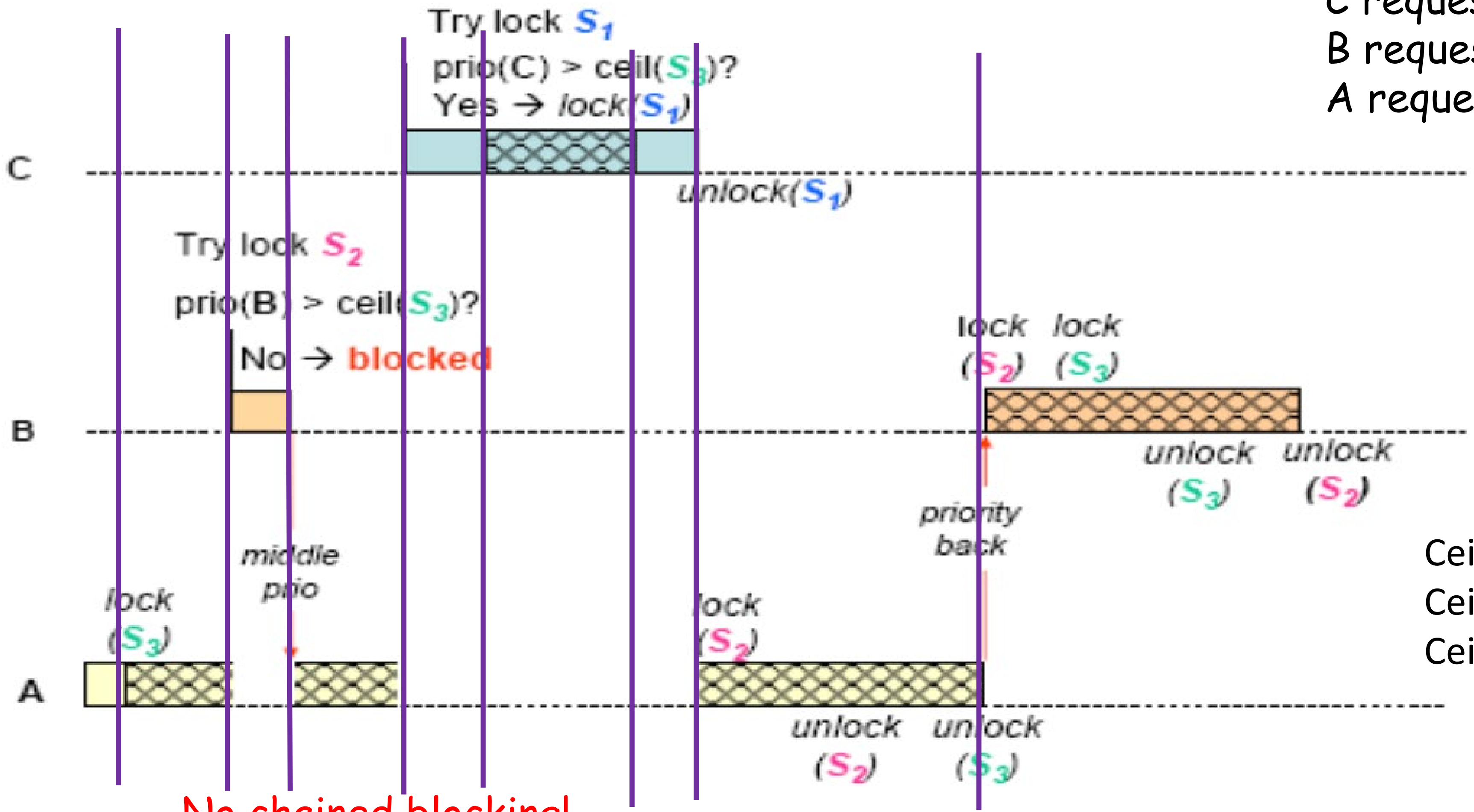
C requests S1
B requests S2 and S3
A requests S3 and S2

Ceiling(S1) = H (high)
Ceiling(S2) = M (medium)
Ceiling(S3) = M (medium)

No chained blocking!

PCP

Assume C has a high priority (H), A has a low priority (L), and B has a medium priority (M).



No chained blocking!

C requests S1
B requests S2 and S3
A requests S3 and S2

Ceiling(S1) = H (high)
Ceiling(S2) = M (medium)
Ceiling(S3) = M (medium)

Facts about Priority Ceiling

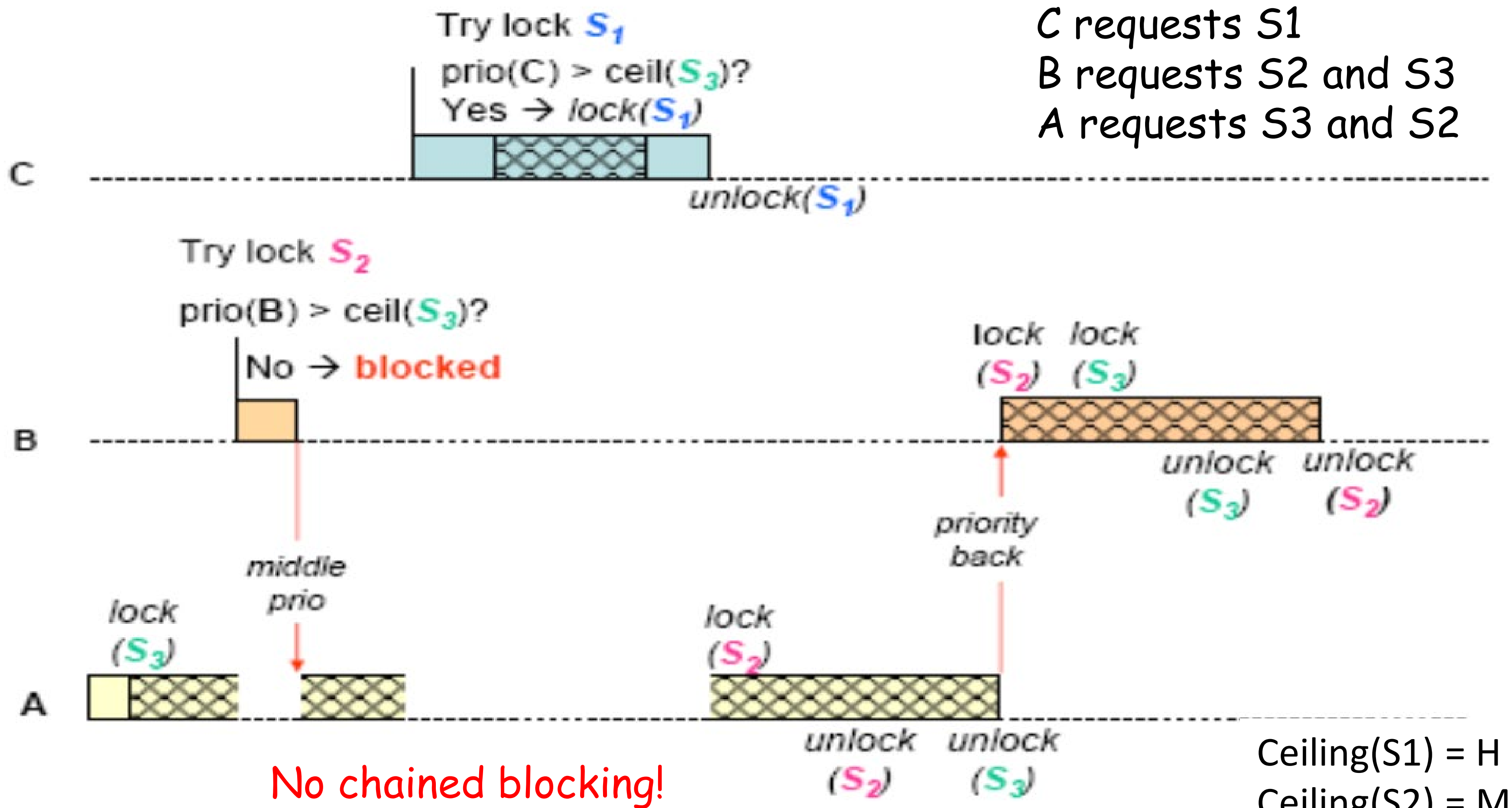
- A task can acquire a resource only if
  - the resource is free, AND
  - it has a higher priority than the priority ceiling of the system, or when the requesting task is holding the resource(s) whose priority ceiling is equal to the priority ceiling of the system.

- A task can be blocked by at most one critical section.

- Higher run-time overhead than Priority Inheritance Protocol

Revisit PCP Example

Assume C has a high priority (H), A has a low priority (L), and B has a medium priority (M).



C requests S1
B requests S2 and S3
A requests S3 and S2

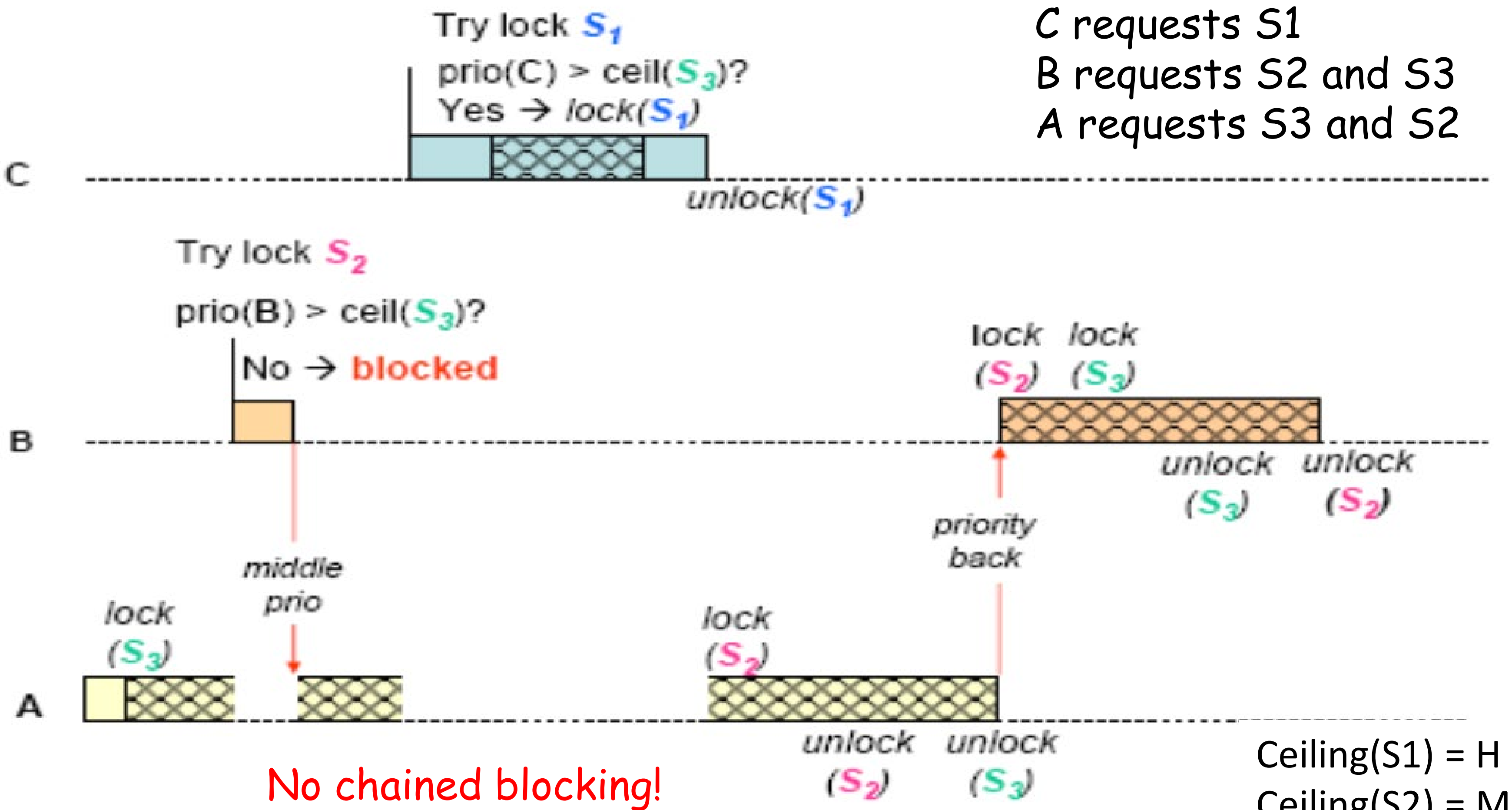
Q1:
At which time point, priority inheritance or priority update occurs?
How the priority of a task is computed?

No chained blocking!

Ceiling(S1) = H (high)
Ceiling(S2) = M (medium)
Ceiling(S3) = M (medium)

Revisit PCP Example

Assume C has a high priority (H), A has a low priority (L), and B has a medium priority (M).



C requests S1
B requests S2 and S3
A requests S3 and S2

Q2:
At which time point, the priority ceiling of the system changes?

No chained blocking!

Ceiling(S1) = H (high)
Ceiling(S2) = M (medium)
Ceiling(S3) = M (medium)

End

Real Time Systems and Control Applications

Contents

Practice Questions on PIP & PCP

Review Questions

- What is the major cause of Priority Inversion?
- What is the drawback of NPCS?
- What is the drawback of PIP?

True or False

- 1) In PCP, if a resource R is free, the resource requesting task will acquire the resource R.
- 2) The priority ceiling of a resource and the priority ceiling of the system is fixed, once the set of real time tasks and their requested resources are given.
- 3) When a task has the same priority to the system's priority ceiling, if the resource is free the task will acquire a certain resource.
- 4) When resource contention occurs in a real-time system, the priority celling protocal can be used with RM scheduler for resource access control.

Question

- Consider Set of 5 Tasks:
  - T1: (7, 20, 3, 20, **[R1, 1]**)

  T1 requests resource R1 one time unit after it is scheduled.
  - T2: (5, 22, 3, 22, **[R2, 1]**)

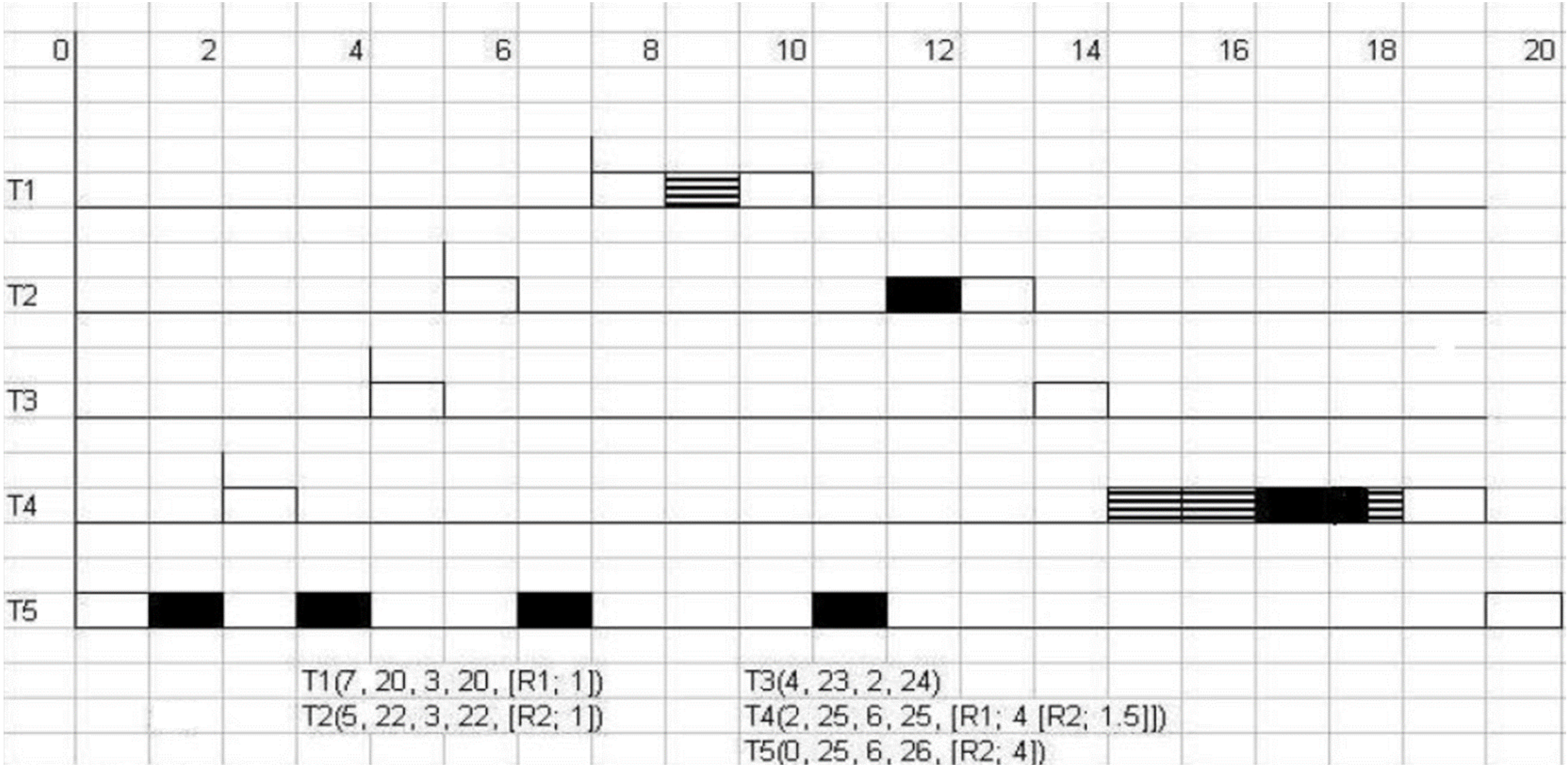
  T2 requests resource R2 one time unit after it is scheduled.
  - T3: (4, 23, 2, 24)

  No CS or resource requests in T3.
  - T4: (2, 25, 6, 25, **[R1, 4; [R2, 1.5]]**)

  T4 requests resource R1 one time unit after it is scheduled, and R2 is requested in CS for resource R1 (assume R2 is requested after 2 time units when R1 is allocated to T4).
  - T5: (0, 25, 6, 26, **[R2, 4]**)

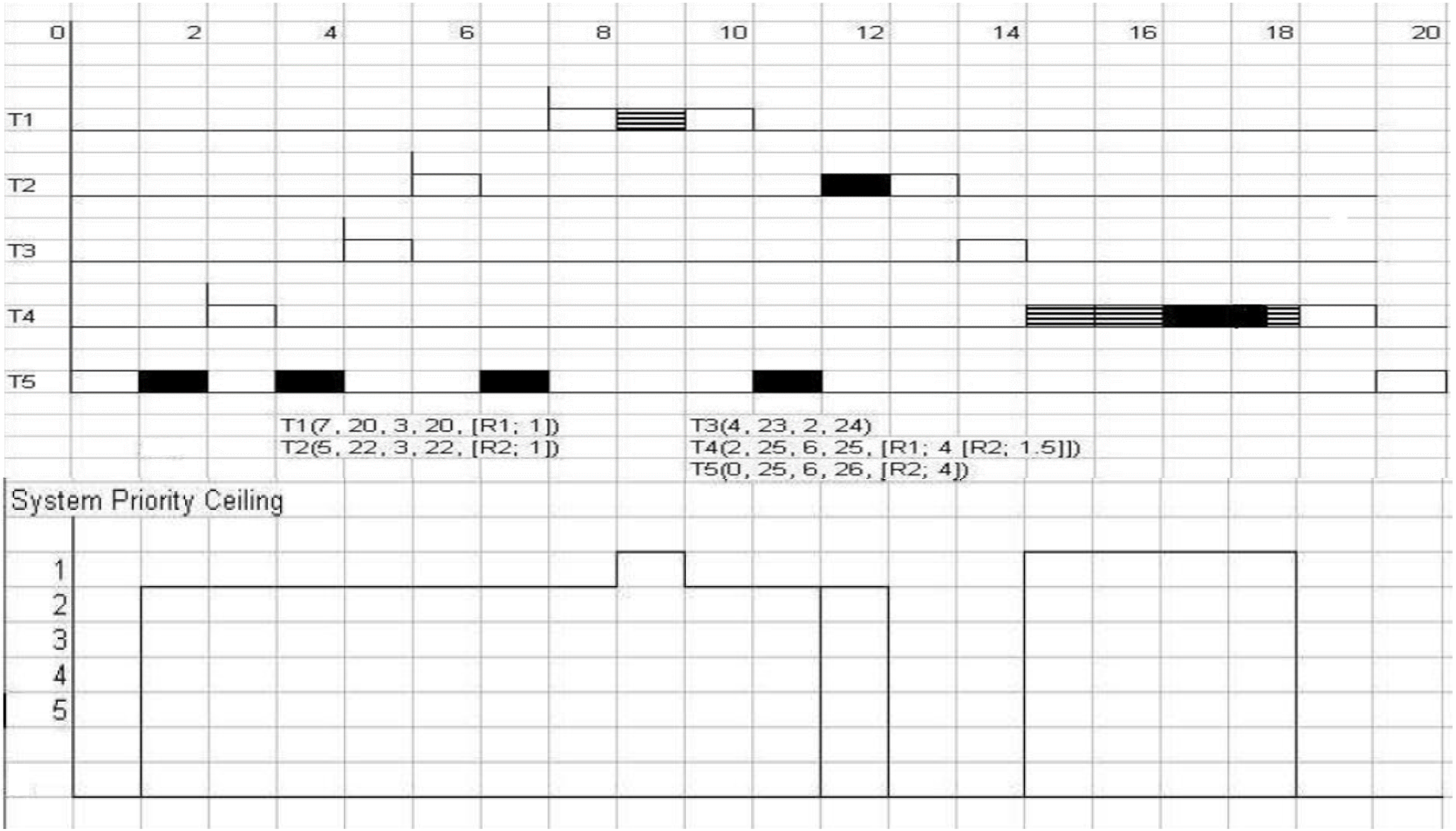
  T5 requests resource R2 one time unit after it is scheduled.

RM Schedule with PCP



Solution Steps

1. Determine priorities: T1>T2>T3>T4>=T5
2. Compute priority ceiling of resources: ceiling(R1) = Priority(T1), ceiling(R2) = Priority(T2)
3. Check the occurrence of events:
   (1) t=0, T5 is released (schedule T5)
   (2) t=1, T5 requests R2: Resource allocation decision: Check if priority of T5 > System priority celling: yes, so R2 is allocated and T5 is in CS.
   (3) t=2, T4 is released: Scheduling decision: Check if priority of T4 > T5: yes, T4 is scheduled
   (4) t=3, T4 requests R1: Resource allocation decision: Check if priority of T4 > System priority celling: no, system priority ceiling is Priority(T2), it is higher than Priority(T4), so R1 cannot be allocated. Now, priority inheritance occurs.
   ….
   (7) t=6, T2 requests R2: Resource allocation decision: Check if priority of T2 > System priority celling: no, they are the same, so R2 cannot be Allocated to T2, and T5 blocks T2.
   (8) t=7, T1 is released: Check if priority of T1 > T5: yes, so T1 is scheduled.
   (9) t=8, T1 requests R1: : Resource allocation decision: Check if priority of T1 > System priority celling: yes, so R1 is allocated to T1.
   ….
   (11) t=10, T1 is completed and T5 is in its critical section and with highest priority among the waiting tasks, so T5 is scheduled.
   ….



0 2 4 6 8 10 12 14 16 18 20
T1
T2
T3
T4
T5
T1(7, 20, 3, 20, [R1; 1])
T2(5, 22, 3, 22, [R2; 1])
T3(4, 23, 2, 24)
T4(2, 25, 6, 25, [R1; 4 [R2; 1.5]])
T5(0, 25, 6, 26, [R2; 4])
System Priority Ceiling
1
2
3
4
5

Question

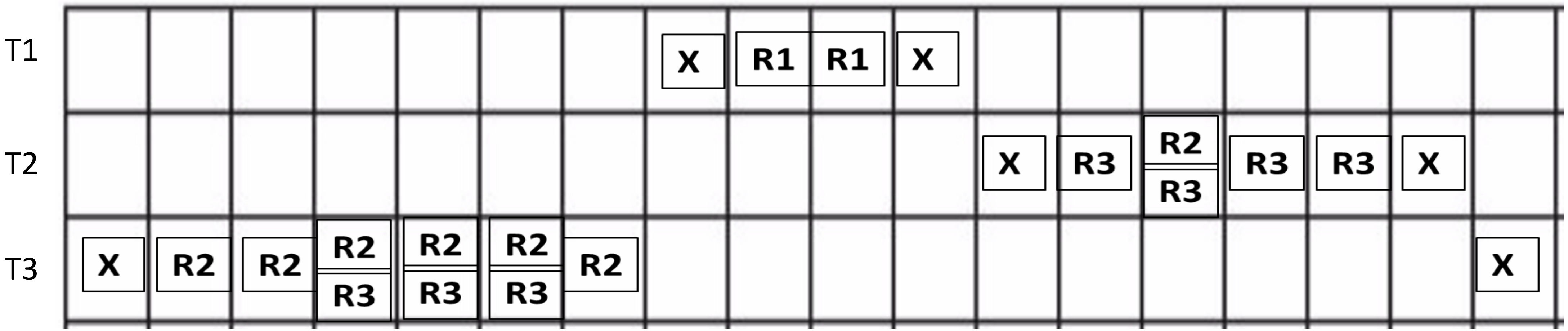
- Consider the following set of tasks (consider only one job in each task):
- T1(5; 10; 4; 10 [R1; 2])
- T2(2; 15; 6; 15 [R3; 4[R2 ; 1]])
- T3(0; 20; 8; 20 [R2; 6[R3; 3]])
- The resource R1 is required by T1 after it has executed for 1 time unit.
- The resource R2 is required by task T2 after it has executed for 2 time units and by Task T3 after it has executed 1 time unit.
- The resource R3 is required by task T2 after it has executed 1 time unit and by task T3 after it has executed 3 time units.

- **Show the schedule for these tasks based on RM algorithm and uses the NPCS, PIP, and PCP.**

Answer Using NPCS

T1(5; 10; 4; 10 [R1; 2])
T2(2; 15; 6; 15 [R3; 4[R2 ; 1]])
T3(0; 20; 8; 20 [R2; 6[R3; 3]])

Resource Requests
T1: (X) (R1) (R1) (X)
T2: (X) (R3) (R2 R3) (R3) (R3) (X)
T3: (X) (R2) (R2) (R2 R3) (R2 R3) (R2 R3) (R2) (X)



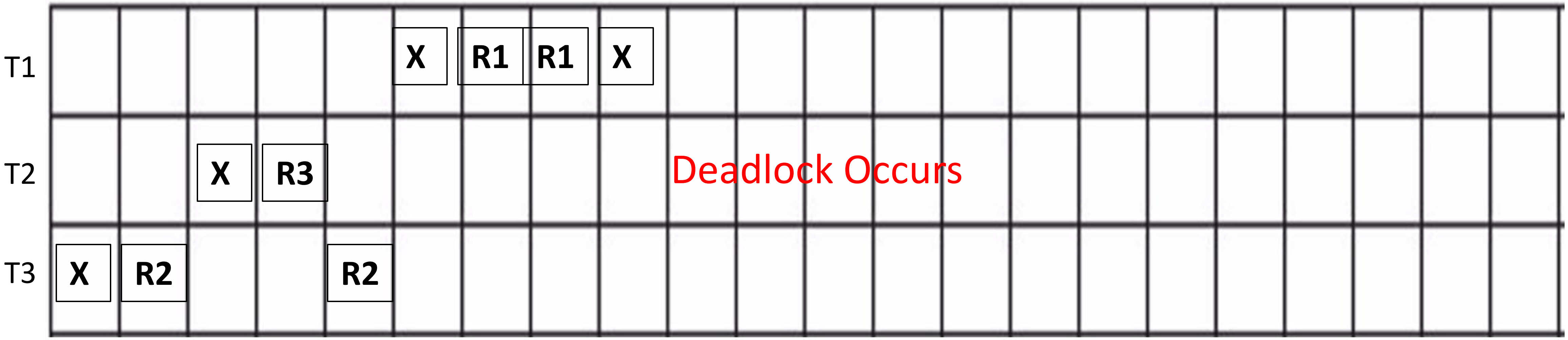
R1 | R2 | R3 Resource R1, R2, and R3 is allocated respectively

X No resource is allocated

Answer Using PIP

T1(5; 10; 4; 10 [R1; 2])
T2(2; 15; 6; 15 [R3; 4[R2 ; 1]])
T3(0; 20; 8; 20 [R2; 6[R3; 3]])

Resource Requests
T1: (X) (R1) (R1) (X)
T2: (X) (R3) (R2 R3) (R3) (R3) (X)
T3: (X) (R2) (R2) (R2 R3) (R2 R3) (R2 R3) (R2) (X)



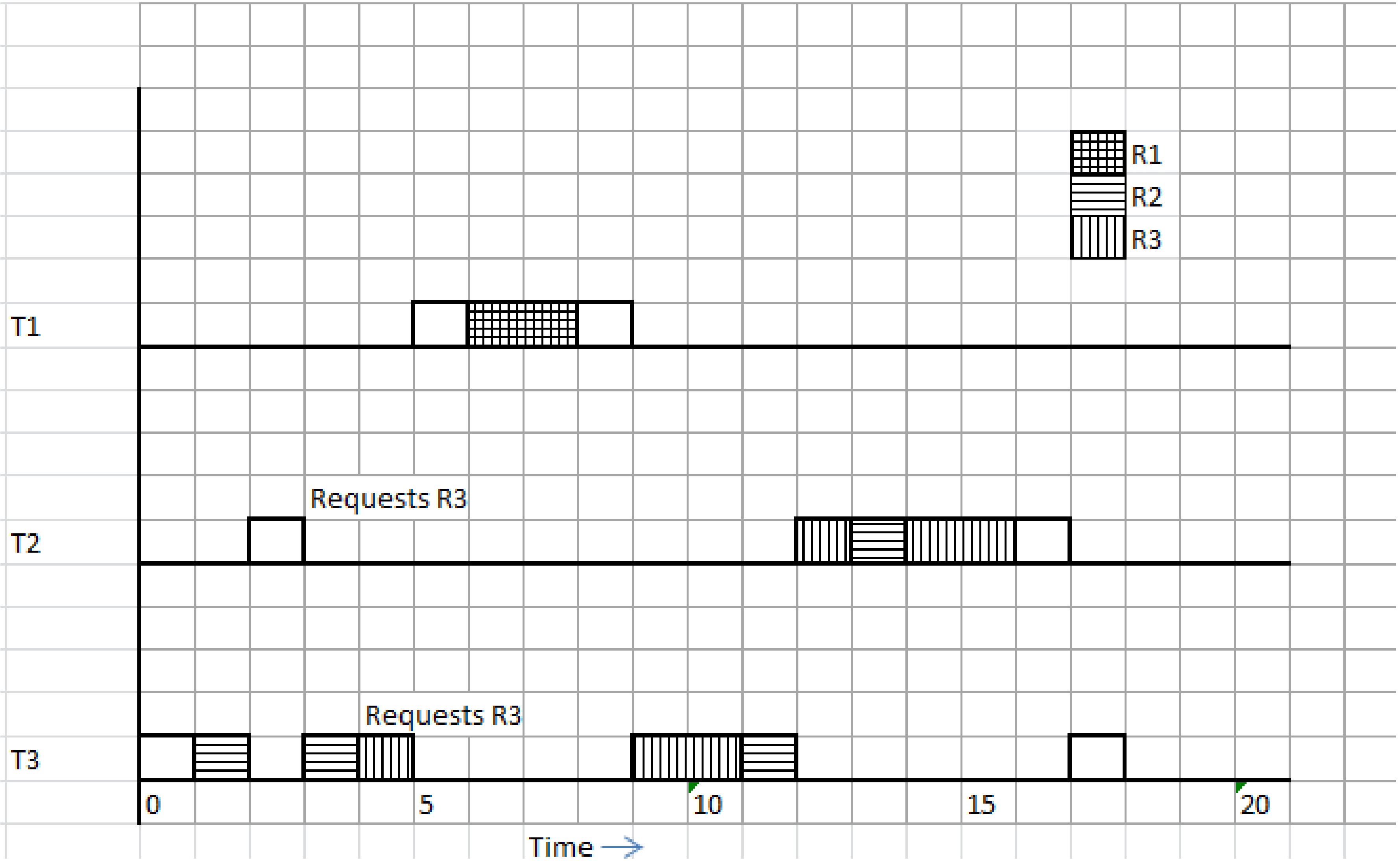
R1 R2 R3 Resource R1, R2, and R3 is allocated respectively

X No resource is allocated

Answer Using PCP

At time t=3 when T2 requests resource R3, but T3 is in critical section and will use R3, so priority inheritance occur, T3 increases its priority to that of T2.

At time t=5 when T1 arrives, T1's priority is higher than the priority ceiling of the system. So T1 is executed. At t=6, resource R1 is free and T1 has higher priority than the priority ceiling of the system. So T1 successfully holds R1.



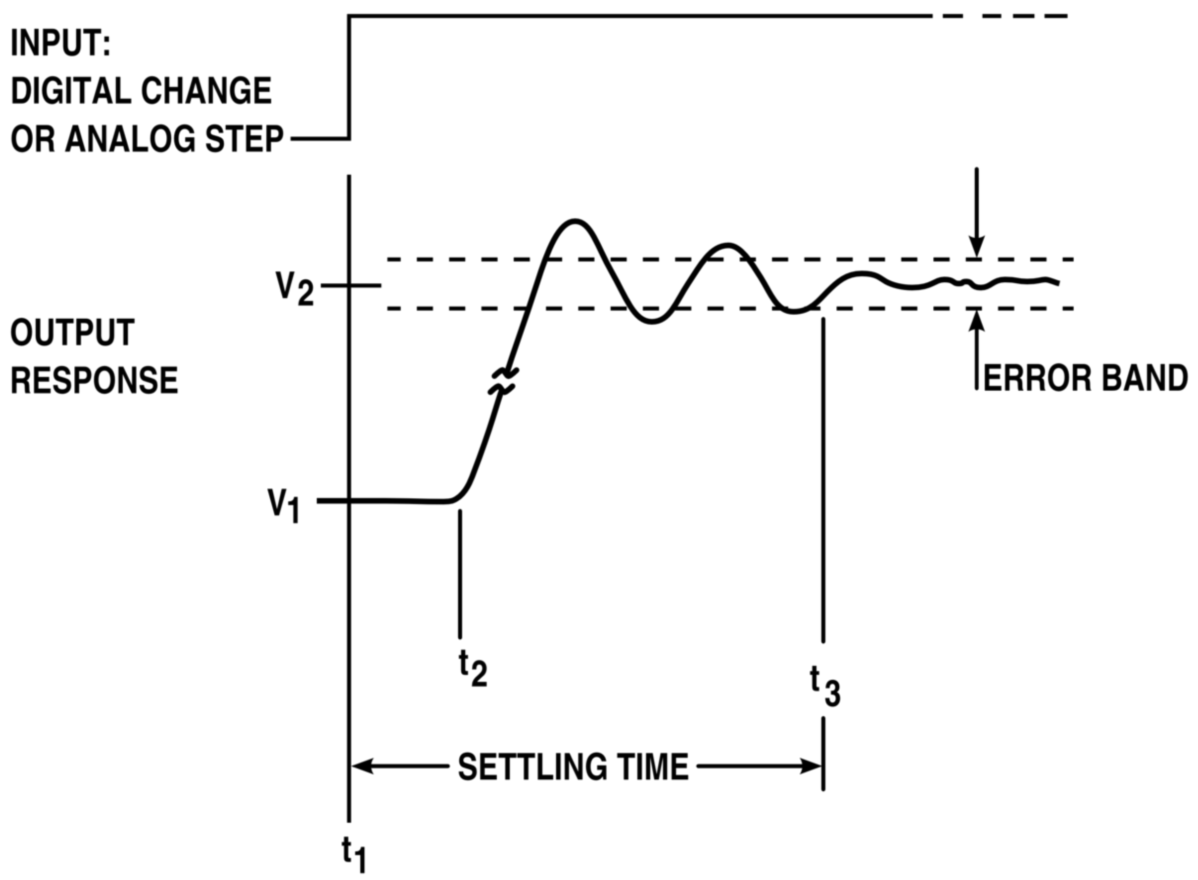
Real Time Systems and Control Applications

Contents

Control Systems

- What is a control system?
  - Desired output, desired performance with specified input
  - Performance: transient response, steady state error

- Types of Systems
  - Open Loop
  - Closed loop
  - Multi-loop



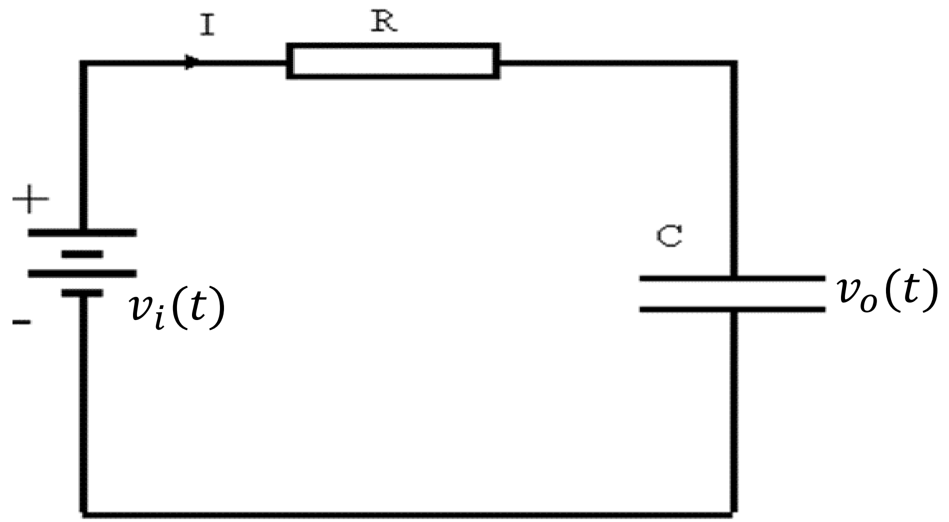
Time Domain v.s. Frequency Domain

- Actually, the input and output are in time domain, why we need Laplace Transform, and wish to investigate the system behavior in frequency domain?

- Consider a very simple circuit system:

$$RC\frac{dv_o(t)}{dt}+v_o(t)=v_i(t)$$

$$v_i(t)=1$$



Solving the ODE, we get

$RC\ dv_o(t) = [v_i(t) - v_o(t)]\ dt$ and $v_i(t) = 1\ for\ t > 0$

$$RC\ \frac{dv_o(t)}{[1 - v_o(t)]} = dt$$

$$\int RC\ \frac{dv_o(t)}{[1 - v_o(t)]} = \int dt$$

$$-RC \int \frac{d[1 - v_o(t)]}{[1 - v_o(t)]} = \int dt$$

$$-RC\ \ln\big(1 - v_o(t)\big) = t + C1$$

$$\ln\big(1 - v_o(t)\big) = \frac{t + C1}{-RC}$$

$$e^{-\frac{t+C1}{RC}} = 1 - v_o(t)$$

$$v_o(t) = 1 - e^{-\frac{t+C1}{RC}}$$

Time Domain v.s. Frequency Domain

- In time domain by solving differential equation, we have:

$$v_o(t) = 1 + k\, e^{-\frac{1}{RC}t}$$

Need initial condition of $v_o(0)$ to solve k. If $v_o(0)$ =0, k=-1.

Hence, $v_o(t) = 1 - e^{-\frac{1}{RC}t}$

- In frequency domain by Laplace Transform (LT), we have:

sRC$v_o(s) + v_o(s) = v_i(s)$ ➔ (sRC+1)$v_o(s)$=$v_i(s)$

Hence, $v_o(s) = \frac{\frac{1}{s}}{1+sRC} = \frac{1}{s+s^2RC}$

Inverse Laplace Transform: $L^{-1}\left(\frac{1}{s+s^2RC}\right) = 1 - e^{-\frac{1}{RC}t}$

Laplace Transform

Laplace transformation from the time domain to the frequency domain transforms differential equations into algebraic equations and convolution into multiplication.

Laplace Transform:

$$F(s) = \int_0^{\infty} f(t)\ e^{-st}\, dt$$

Inverse Laplace Transform:

$$f(t) = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} F(s)\ e^{st}\, ds$$

Table of Laplace Transforms: http://tutorial.math.lamar.edu/pdf/Laplace_Table.pdf

Recall An Example

- Unit step function $u(t) = \begin{cases} 1, \ t \geq 0 \\ 0, t < 0 \end{cases}$
- $\mathrm{U}(s) = \mathcal{L}\{u(t)\} = \int_0^\infty u(t)\, e^{-st}\, dt = \int_0^\infty e^{-st}\, dt = -\frac{1}{s}\ e^{-st}|_0^\infty$

$$\mathrm{U}(s) = -\frac{1}{\mathrm{s}}(0-1) = \frac{1}{s}$$

Note that Laplace transforms of various functions are found in tables. Similarly, tables exist for inverse Laplace transforms.

Laplace Transform of Derivatives and Integrals

If $\mathcal{L}[f(t)] = F(s)$, then

$$\mathcal{L}[f'(t)] = sF(s) - f(0)$$

and

$$\mathcal{L}\left[\int_0^t f(t)\,dt\right] = \frac{F(s)}{s}$$

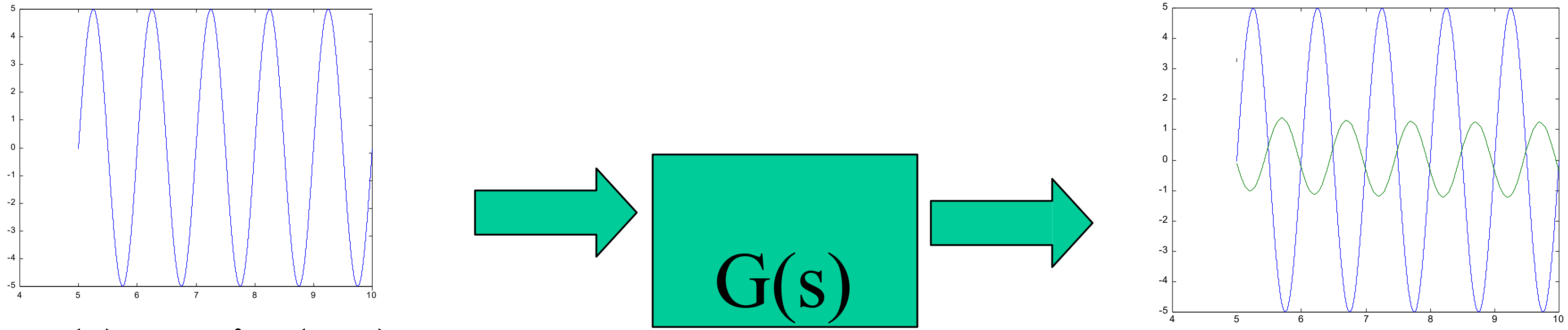
For higher order derivatives

$$\mathcal{L}[f''(t)] = s^2F(s) - sf(0) - f'(0)$$

A convenient method to transform differential equations to algebraic equations.

Frequency Response

For a sinusoidal input, the output of a linear system is also a sinusoidal. However, the output will have a **different magnitude** and will be subject to **a phase shift**.

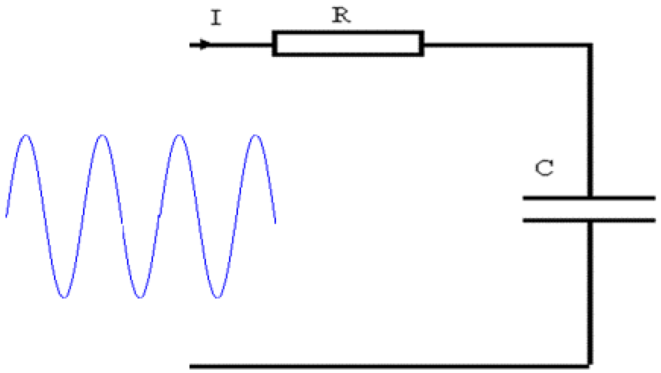


$$V_{in}(t) = \sin(\omega t)$$

$$vel(t) = a_3 \sin(\omega t + \phi)$$

Another Form of System Model in Frequency Domain

- Replace s with jw in transfer function $G(s)$ to get $\mathrm{G}(jw)$



- $\mathrm{G}(jw) = \dfrac{1}{1+jwRC}$

- So $|\mathrm{G}(jw)| = \left|\dfrac{1}{1+jwRC}\right| = \dfrac{1}{\sqrt{1+(wRC)^2}}$

An expression for the gain where frequency is a variable. Allows calculation of gain at any frequency.

$$\angle \mathrm{G}(jw) = \tan^{-1}(-wRC)$$

It is also possible to calculate the phase shift of the output relative to the phase angle of the input.

Why s Is Substituted With jw In Transfer Function?

Laplace transform:

$$G(s) = \mathcal{L}\{f(t)\} = \int_0^{\infty} f(t)\ e^{-st}\, dt$$

Fourier transform:

$$H(s) = \mathcal{F}\{f(t)\} = \int_0^{\infty} f(t)\ e^{-jwt}\, dt$$

The only difference between Laplace transform and Fourier transform is the variable substitution s=jw. Hence, $H(s) = G(s)|_{s=jw} = G(jw)$

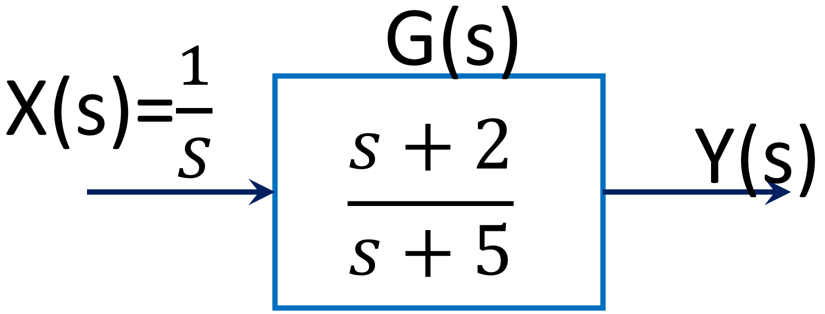
Time Response

- Why is time response important?
- Temperature control - how long it takes to reach a new steady state?
- Want 200 deg but it goes to 250 before settling down, will it be acceptable?
- How about cruise control at 100 km/h but actual speed varies between 80 and 120?

Next, let us examine the parameters that control the time behaviour of systems.

Consider First Order Systems

$$Y(s) = \frac{s+2}{s(s+5)} = \frac{2/5}{s} + \frac{3/5}{s+5}$$



$$y(t) = 2/5 + 3/5\ e^{-5t}$$

Forced response   Natural response

The output response of a system consists of:

(1) a natural (**transient**) response: $3/5\ e^{-5t}$

(2) a forced response (**steady state**) response: $2/5$

Poles and Zeros

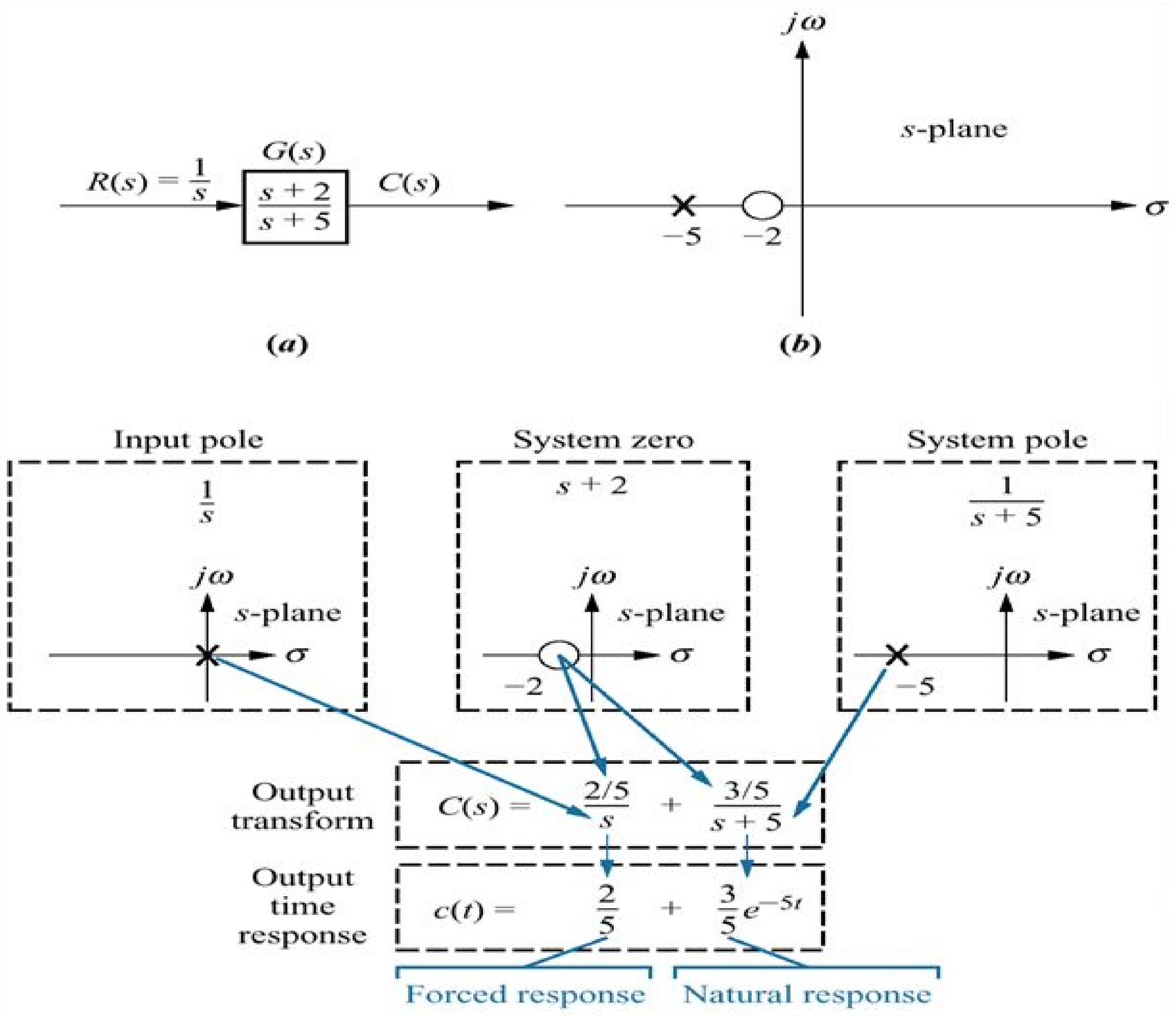
- In previous example: $Y(s) = \frac{s+2}{s(s+5)}$

s = 0, s = - 5 are poles and s = - 2 is the zero of the transfer function.

(1) A pole at origin generated a step function at the output.

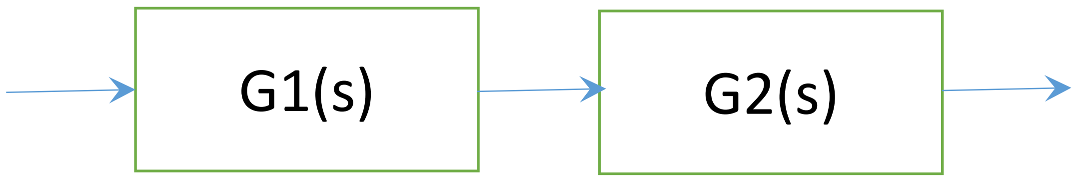
(2) The pole at -5 generated transient response e^{-5t}. The further to the left a pole is on the negative real axis, the faster the exponential transient response will decay to zero.

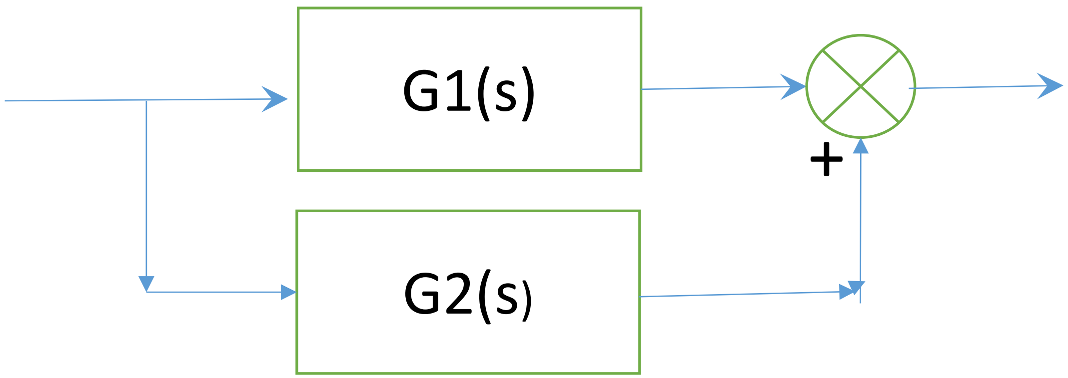
(3) The zeros and poles generate the amplitude for both the forced as well as the natural responses.

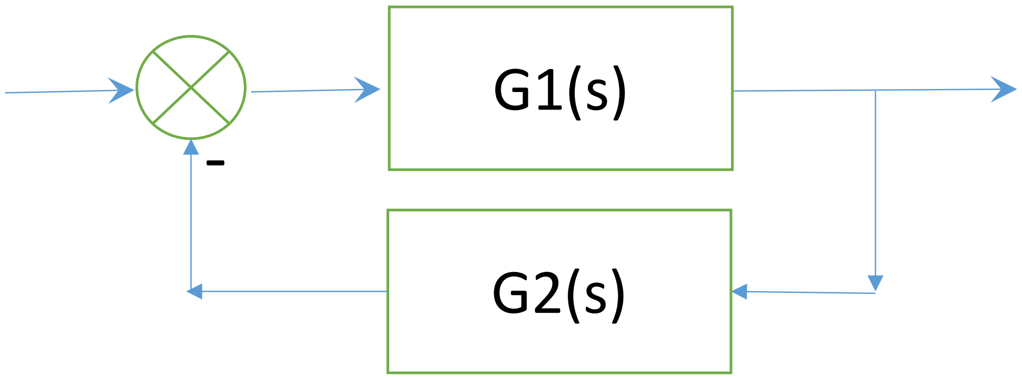

G(s)
R(s) = 1/s
(s + 2)/(s + 5)
C(s)
jω
s-plane
σ
−5
−2
(a)
(b)
Input pole
1/s
System zero
s + 2
System pole
1/(s + 5)
Output transform
C(s) = 2/5 / s + 3/5 / (s + 5)
Output time response
c(t) = 2/5 + 3/5 e^{−5t}
Forced response
Natural response

More Reviews

Transfer Functions



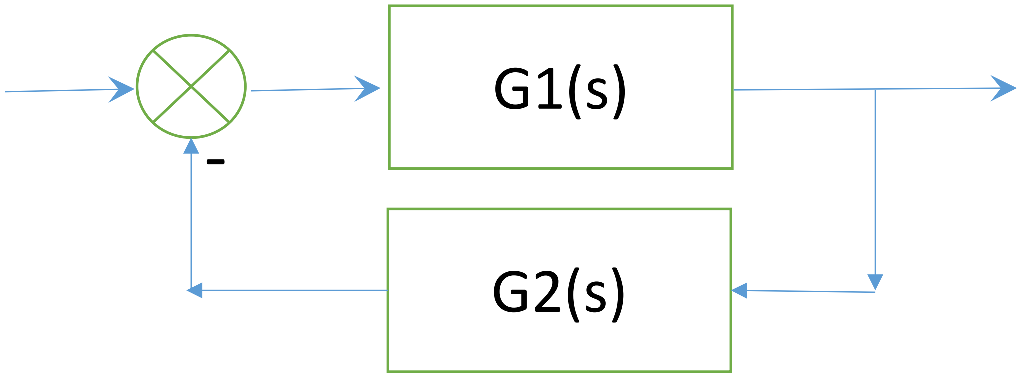




What are the system transfer functions?

Characteristic Equation of System

The **characteristic equation** is setting the denominator of the closed-loop transfer function to zero (0).

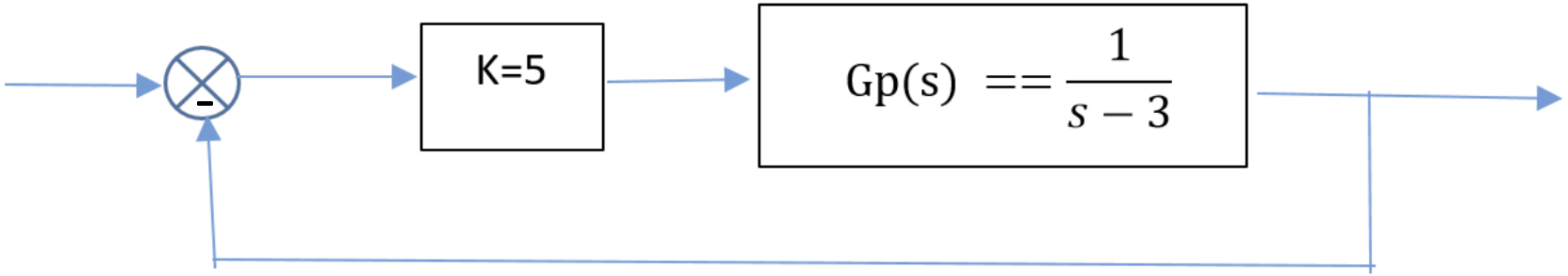


Characteristic equation of the system is

$1+G1(s)G2(s)=0$

Stability?

- Is this system stable?



Real Time Systems and Control Applications

Contents

Time Constant of First Order Systems

- First order system

$$X(s)=\frac{1}{s} \longrightarrow \overset{G(s)}{\boxed{\frac{a}{s+a}}} \longrightarrow Y(s)$$

- The output of a general first order system to a step input:

$$Y(s) = X(s)\,G(s) = \frac{a}{s\,(s+a)}$$

- This results in a time domain output given by: $\mathrm{y}(t) = 1 - e^{-at}$, where parameter a is the only parameter that affects the output.

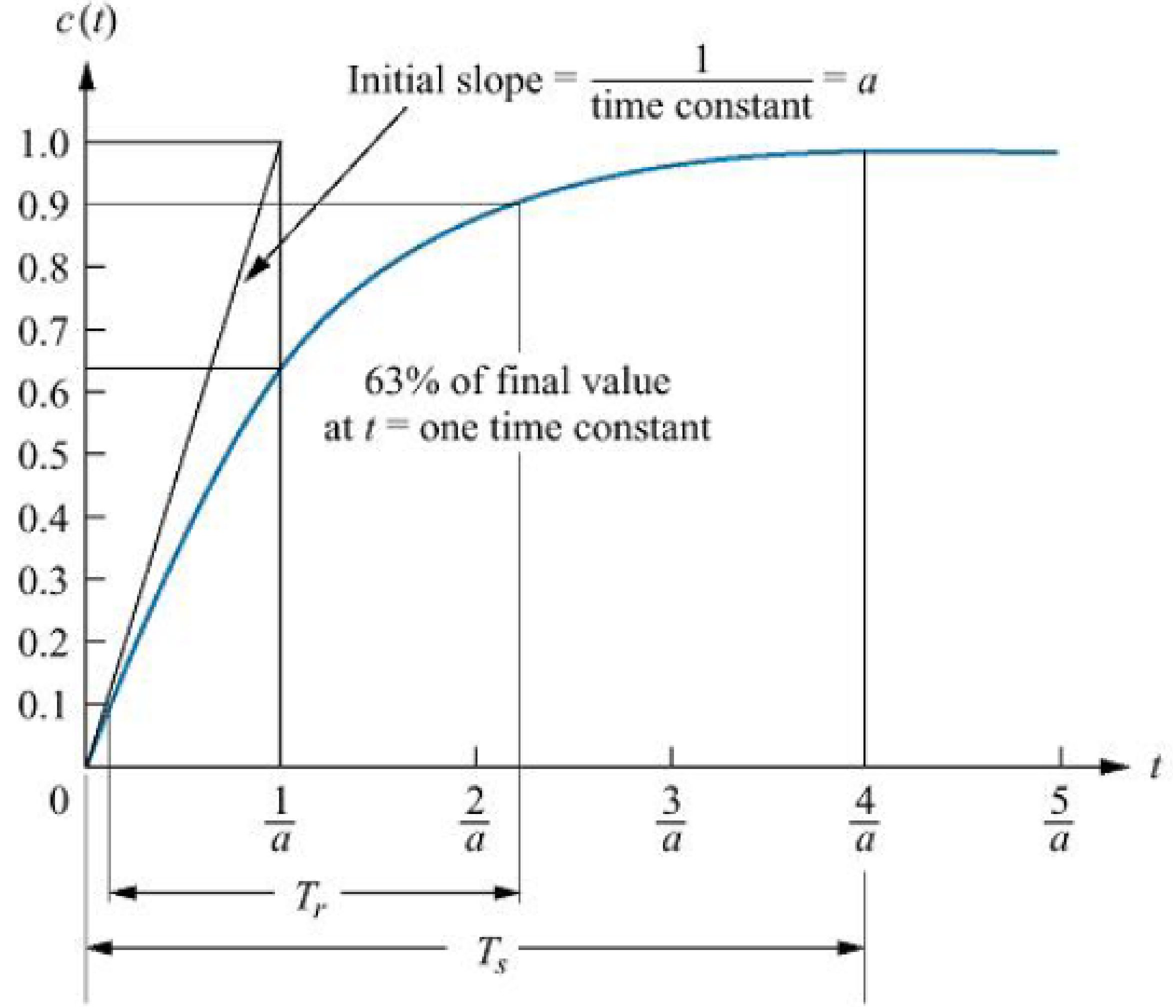
$$t = \frac{1}{a}, \mathrm{y}(t) = 0.63.$$

$\frac{1}{a}$ is the time constant of the response, and is the time it takes for the step response to rise to 63% of its final value.

Response in Time Domain

- Rise Time (T_r): time for the waveform to go from 0.1 to 0.9 of its final value. For first order systems: $T_r = \frac{2.2}{a}$

- Settling Time (T_s): time for the response to reach and stay within 2% of its final value. For first order systems: $T_s = \frac{4}{a}$



Another First Order Response Example

- Give $G(s) = \frac{1}{s+a}$, what is the time constant, rise time, and settling time?
  - $\mathrm{Y}(s) = \frac{1}{a}\left(\frac{1}{s} - \frac{1}{s+a}\right) \Rightarrow y(t) = \frac{1}{a}(1 - e^{-at})$

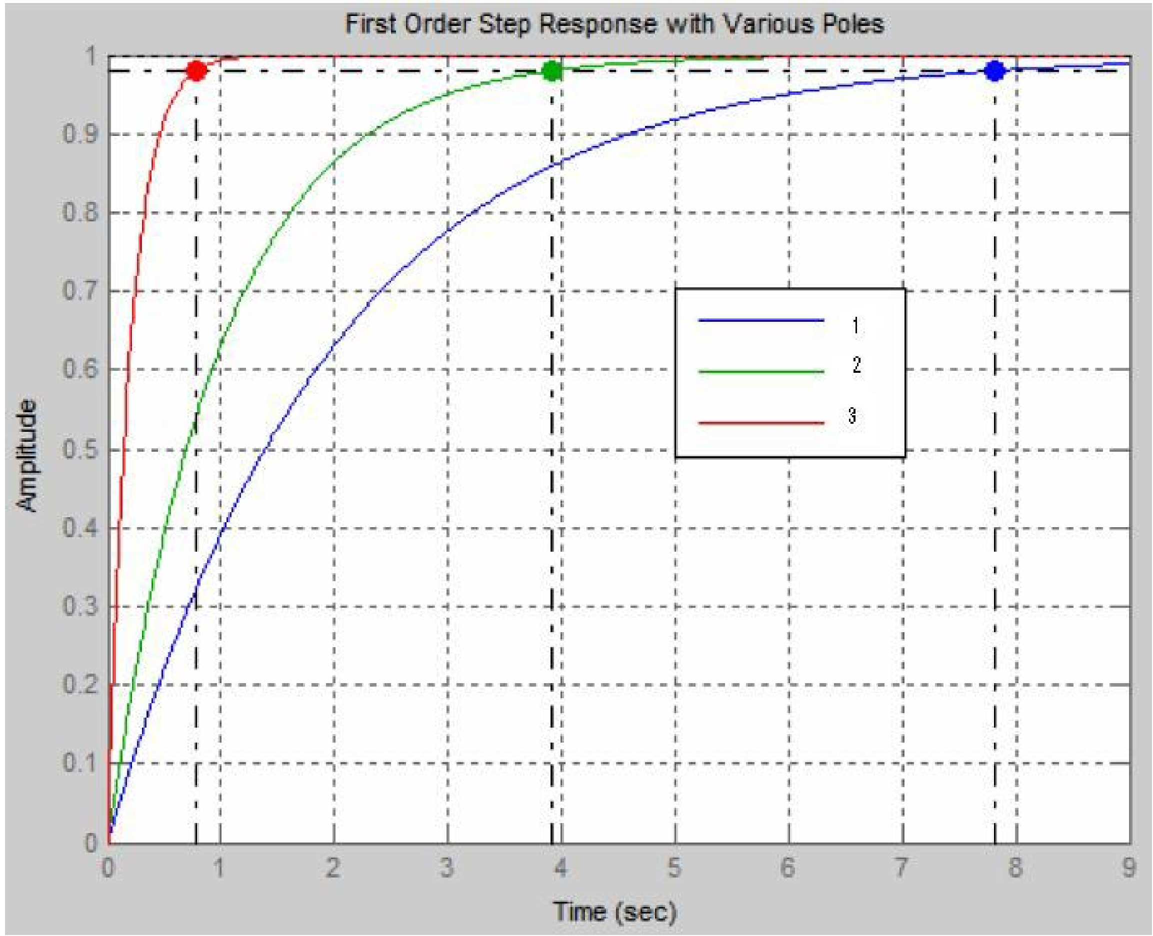
  Solution:
  - Time constant τ makes $y(\tau) = \frac{1}{a} * 0.63$, so $\tau = \frac{1}{a}$
  - Rise time $t_r = t_2 - t_1$, where t_2 makes $y(t_2) = \frac{1}{a} * 0.9$ and t_1 makes $y(t_1) = \frac{1}{a} * 0.1$, hence $t_r = \frac{2.3-0.1}{a} = 2.2 * \frac{1}{a}$
  - Settling time t_s makes $y(t_s)$=0.98$\frac{1}{a}$, hence $t_s \cong \frac{4}{a}$

More Example

The figure shows the response of three first order systems having transfer function $\frac{K}{s+a}$, where the values of K are different for the three systems.

Answer the following questions:

- 1. Which of the three curves (1, 2, 3) represents a system with the lowest time constant?
- 2. The big dots on the three graphs represent the time when the response settles within 2% of the final value. Find the transfer function for each of the three systems.



Solution

- The settling time for first order systems is given by $T_s=\frac{4}{a}$.
- From the figure, the values of T_s are 7.8, 3.9 and 0.8 respectively, so the value a for 3 systems are roughly 0.5, 1.0 and 5 respectively.

Since the steady state value of each system is 1, so K=a. Therefore the transfer function of the systems are $\frac{0.5}{s+0.5}$, $\frac{1}{s+1}$, and $\frac{5}{s+5}$.

Second Order Systems

- Most real-world systems are not first order systems. A general second order system defined by the transfer function:

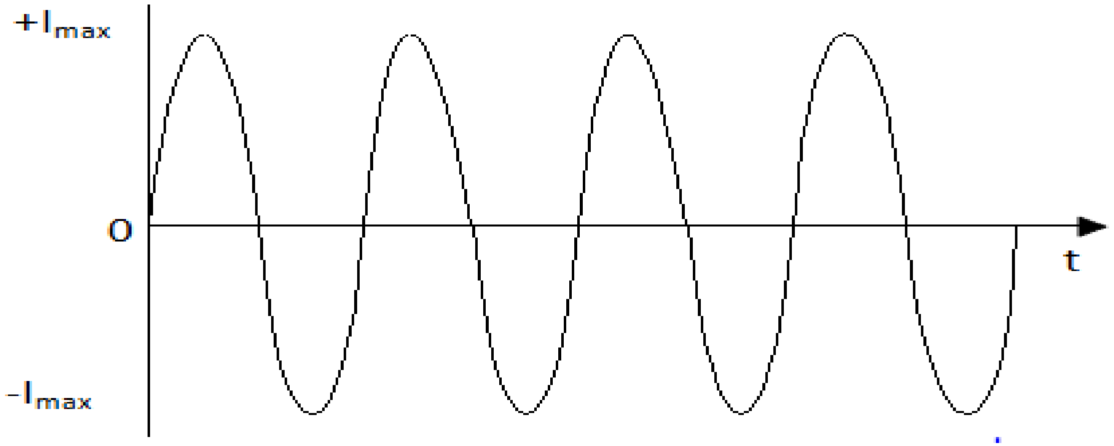
$$G(s) = \frac{b}{s^2 + as + b}$$

- Find the poles of this transfer function to examine the behaviour of the output response. Using quadratic formula:

$$s_1, s_2 = \frac{-a \pm \sqrt{a^2 - 4\mathrm{b}}}{2}$$

A Special Case (a=0)

- If a = 0, the transfer function is $G(s) = \frac{b}{s^2+b}$, and the poles will have only imaginary part $\pm jw$ and by definition the **natural frequency** $w_n = \sqrt{b}$ is the frequency of oscillation of this system.



Undamped Case

Damping Coefficient ζ, when a is not zero

- The complex poles have a real part $\sigma = \frac{-a}{2}$.
- The magnitude of σ is called the exponential decay frequency, and w_n the natural frequency. We define the *Damping Ratio* or *Damping Coefficient*, ζ as

$$\zeta = \frac{Exponential\ decay\ frequency}{Natural\ frequency}$$

$$\therefore\ \zeta = \frac{|\sigma|}{w_n} = \frac{\frac{a}{2}}{w_n} \quad so\ that\ \ a = 2\zeta\, w_n$$

General Second Order Transfer Function

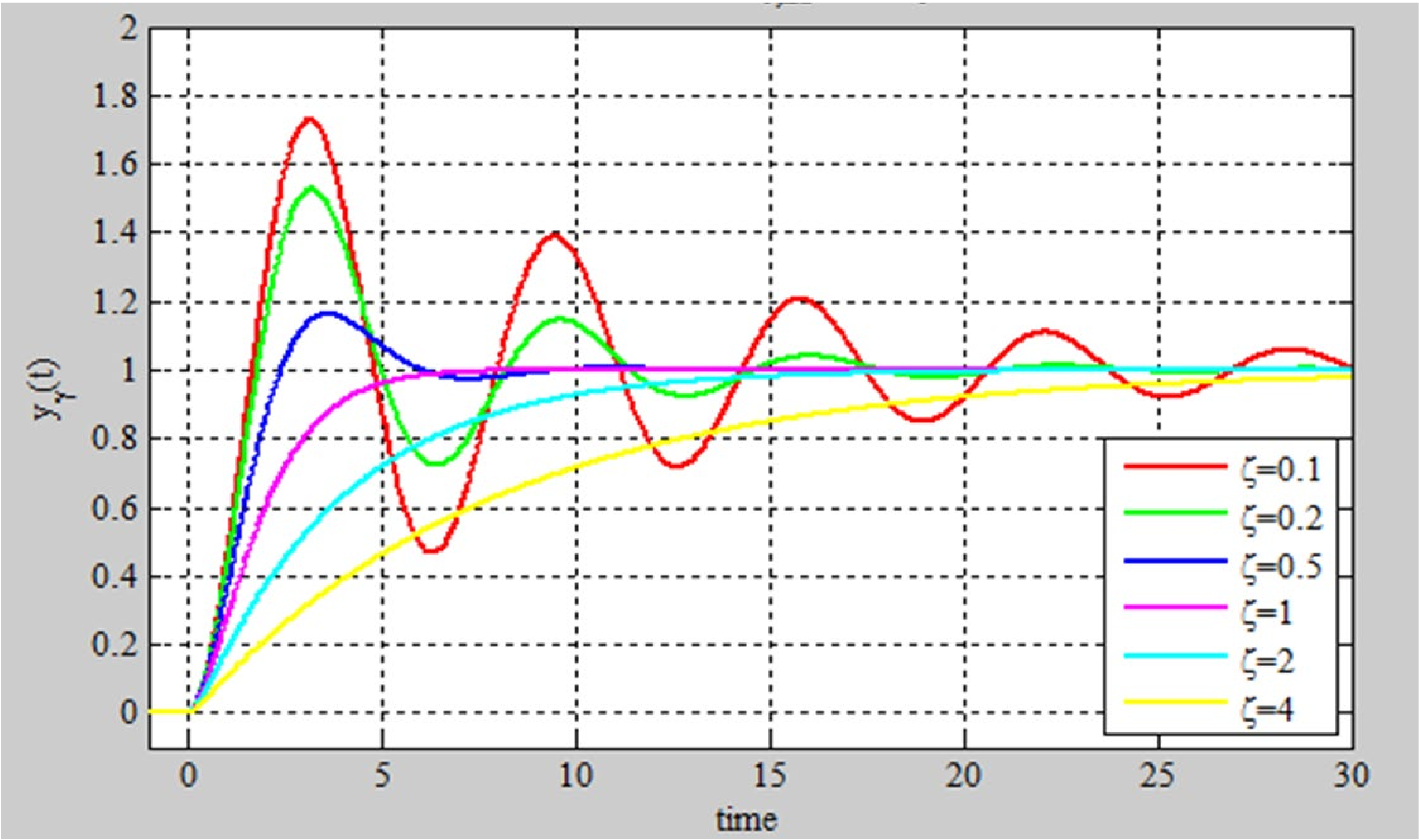
- The general second order transfer function can now be written as:

$$\mathrm{G(s)} = \frac{{w_n}^2}{s^2 + 2\zeta\, w_n\, s + \ {w_n}^2}$$

$$s_1, s_2 = -\zeta\, w_n \pm w_n\sqrt{\zeta^2 - 1}$$

- From above we can examine the effect of parameter ζ on the output of a second order system.

Effect Of Parameter ζ



Summary of Observations

- Two imaginary poles at $\pm j\omega_n$: ζ = 0 (**undamped**)

Natural response: undamped sinusoid of frequency ω_n equal to the imaginary part of the poles. Or $c(t) = A\cos(\omega_n t - \varphi)$

- Two complex poles at $\sigma_d \pm j\omega_d$: 0 < ζ < 1 (**underdamped**)

Natural response: Underdamped response in the form of sinusoid with an exponential envelope whose time constant is equal to the reciprocal of the pole's real part. Or $c(t) = A\, e^{(-\sigma_d)t}\cos(\omega_d t - \varphi)$, where $w_d = w_n\sqrt{1-\zeta^2}$.

Continued…

- Two real poles at σ_1: ζ = 1 (**critically damped**)

Natural response: critically damped system has the time domain response as:

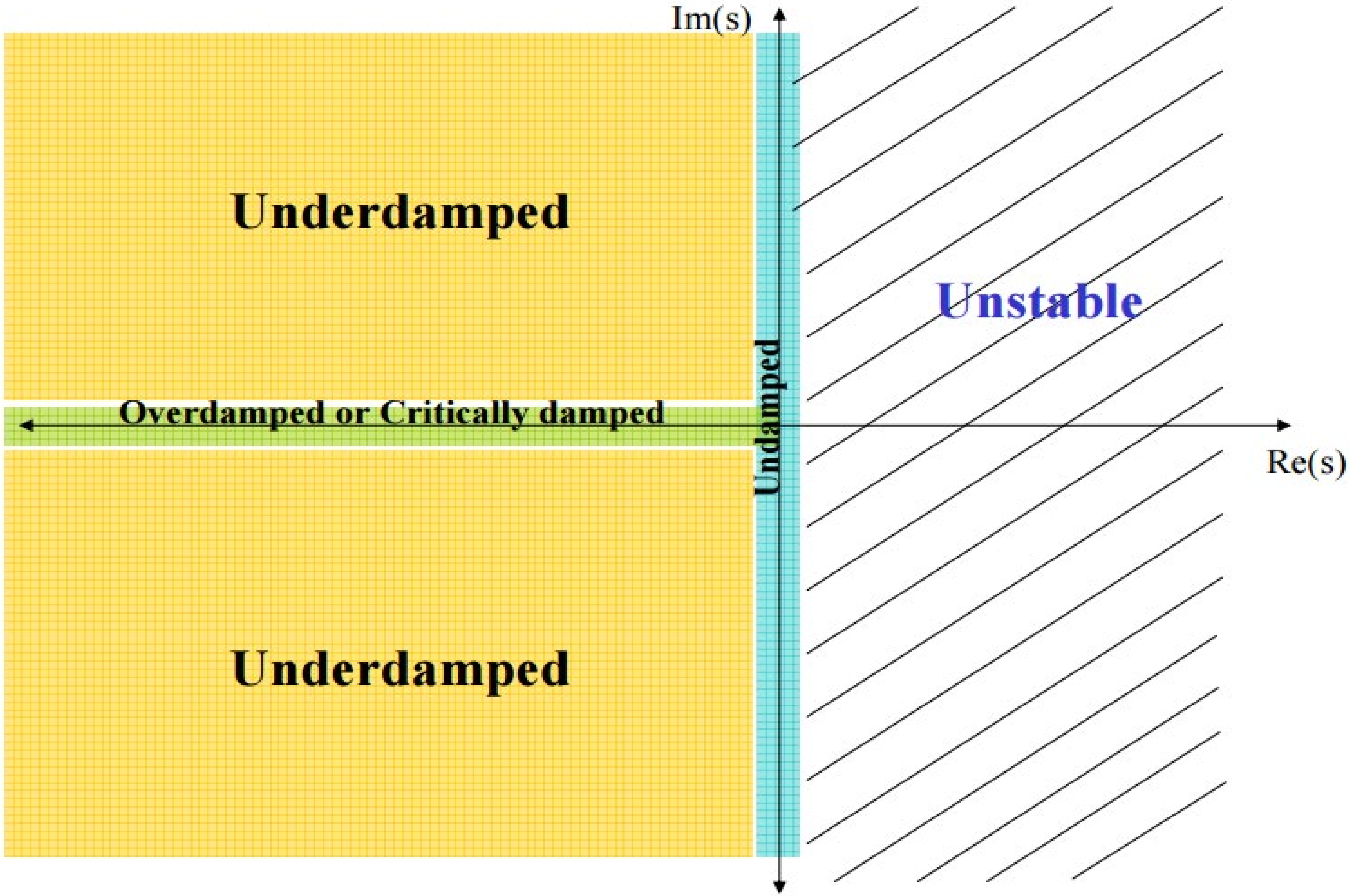
$$c(t) = Kte^{\sigma_1 t}$$

- Two real poles at σ_1 and σ_2: ζ > 1 (**overdamped**)

Natural response: overdamped with two exponentials having time constants equal to the reciprocal of the pole locations. Or

$$c(t) = K(e^{\sigma_1 t} + e^{\sigma_2 t})$$

Second Order Impulse Response



Underdamped Second Order Step Response

- The general Transfer Function of a second order system is:

$$\mathrm{G(s)} = \frac{{w_n}^2}{s^2 + 2\zeta\, w_n\, s + \ {w_n}^2}$$

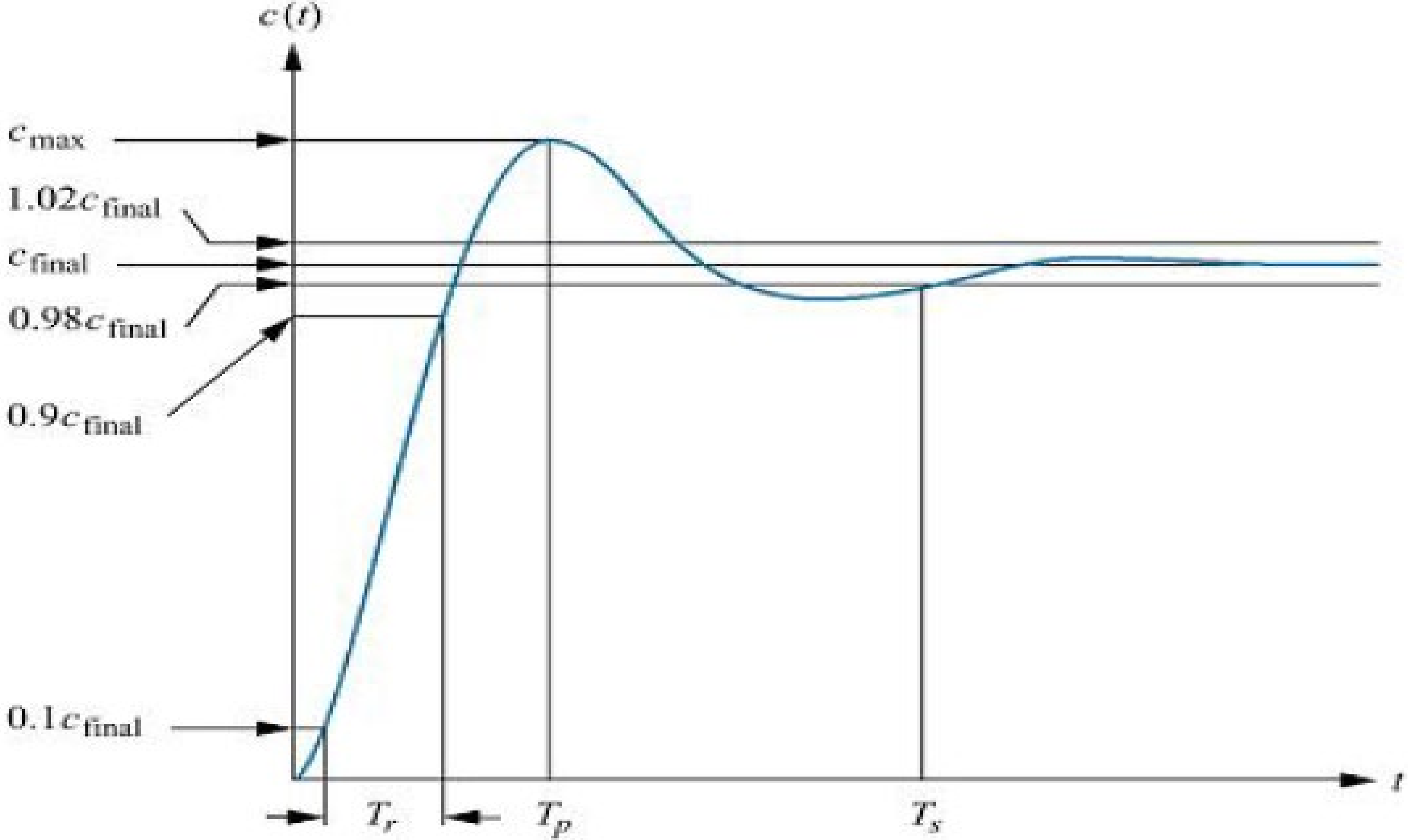
- Consider response for a step input. The transfer function of response C(s) is given by:

$$\mathrm{C(s)} = \frac{{w_n}^2}{s(s^2 + 2\zeta\, w_n\, s + \ {w_n}^2)}$$

- Taking the inverse LT to get response in time domain results in:

$$c(t) = 1 - \frac{1}{\sqrt{1-\zeta^2}}\, e^{-\zeta\, w_n t} \cos(\sqrt{1-\zeta^2}\omega_n t + \varphi)$$

where $\varphi = tan^{-1}\,(\frac{\zeta}{\sqrt{1-\zeta^2}})$.


$c(t)$
c_{max}
$1.02c_{final}$
c_{final}
$0.98c_{final}$
$0.9c_{final}$
$0.1c_{final}$
t
T_r
T_p
T_s

$$c(t) = 1 - \frac{1}{\sqrt{1-\zeta^2}} e^{-\zeta w_n t} \cos(\sqrt{1-\zeta^2}\,\omega_n t + \varphi)$$

$$\varphi = tan^{-1}\left(\frac{\zeta}{\sqrt{1-\zeta^2}}\right)$$

Peak Time, Tp

The time required to reach the first or maximum peak. This can be found by differentiating c(t), and equating to zero which gives:

$$T_p = \frac{\pi}{\omega_n \sqrt{1-\zeta^2}}$$

Because c'(t)= $-\frac{1}{\sqrt{1-\zeta^2}} e^{-\zeta w_n t} \sin\left(\sqrt{1-\zeta^2}\,\omega_n t + \varphi\right)\sqrt{1-\zeta^2}\,\omega_n -$

$$\frac{1}{\sqrt{1-\zeta^2}}(-\zeta\, w_n) e^{-\zeta w_n t} \cos\left(\sqrt{1-\zeta^2}\,\omega_n t + \varphi\right) = 0$$

$$\therefore tan\left(\sqrt{1-\zeta^2}\,\omega_n t + \varphi\right) = \frac{\zeta}{\sqrt{1-\zeta^2}}$$

Therefore, $\sqrt{1-\zeta^2}\,\omega_n t = \pi$ ➔ $T_p = \frac{\pi}{\omega_n \sqrt{1-\zeta^2}}$

Percent Overshoot, % OS

The amount that the waveform overshoots the steady state of final value at peak time, expressed as percentage of steady state value: %OS $= \frac{c_{max} - c_{final}}{c_{final}} \times 100$, where $c_{final} = 1$ and $c_{max} = c(T_p)$. Substituting the expression for $c(T_p) = 1 + e^{-\frac{\pi\zeta}{\sqrt{1-\zeta^2}}}$ in previous subsection and some manipulation results in: %OS $= e^{-\frac{\pi\zeta}{\sqrt{1-\zeta^2}}} \times 100.$

Note that %OS is a function of ζ, the damping ratio only. The above expression gives an expression for ζ in terms of %OS.

$$\zeta = \frac{-\ln(\frac{\%OS}{100})}{\sqrt{\pi^2 + ln^2(\frac{\%OS}{100})}}$$

Finding T_p and %OS From Transfer Function

- Given a transfer function $G(s) = \frac{100}{s^2+15s+100}$, find T_p and %OS.

Solution: $\omega_n = \sqrt{100} = 10$ and ζ = $\frac{\frac{a}{2}}{\omega_n}$ =15/20 = 0.75.

$$\therefore T_p = \frac{\pi}{\omega_n\sqrt{1-\zeta^2}} = 0.475\ s$$

And finally, $\%\text{OS} = e^{-\frac{\pi\zeta}{\sqrt{1-\zeta^2}}} \times 100 = e^{-\frac{0.75\pi}{\sqrt{1-0.75^2}}} \times 100 = 2.838\%.$

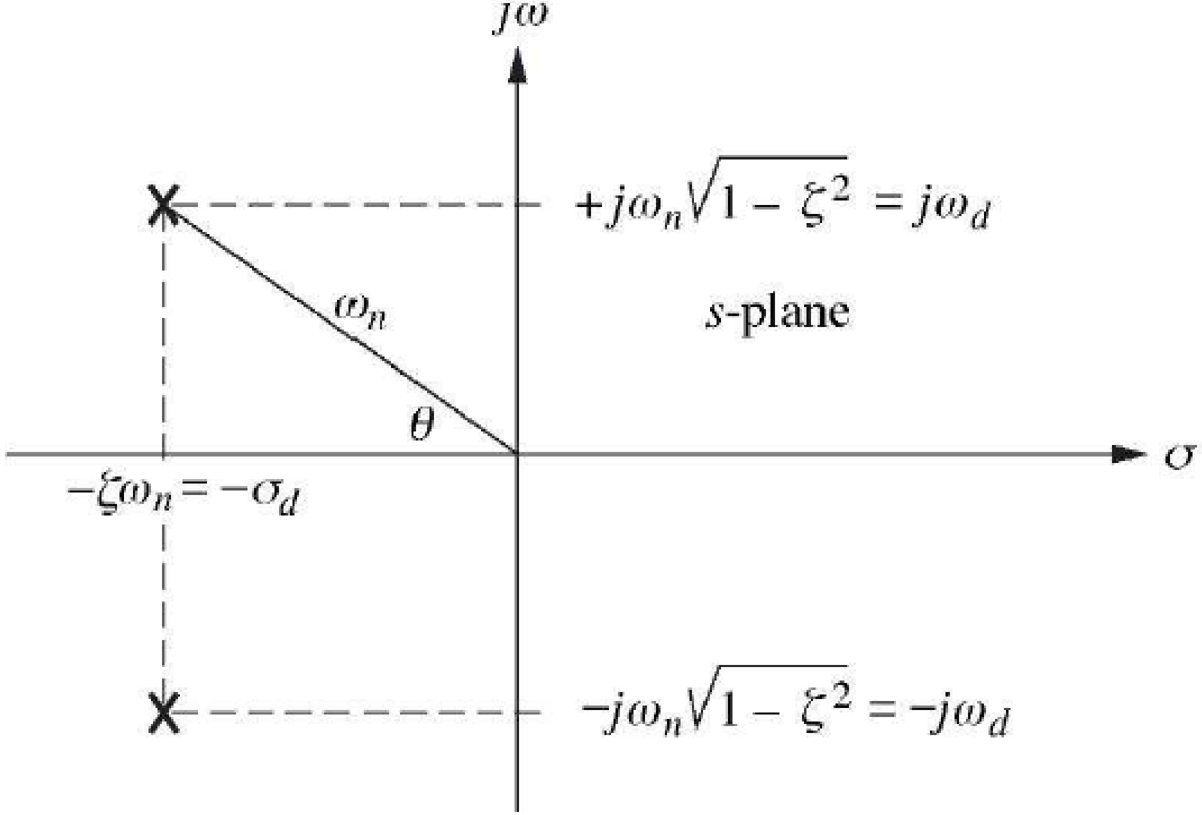
How are T_p And T_s Related To Location of Poles

$$G(s) = \frac{{w_n}^2}{s^2 + 2\zeta w_n s + {w_n}^2}$$

$$s_1, s_2 = -\zeta w_n \pm j w_n \sqrt{\zeta^2 - 1}$$

$$T_p = \frac{\pi}{\omega_n \sqrt{1 - \zeta^2}}$$

$$T_s \cong \frac{4}{\zeta \omega_n}$$

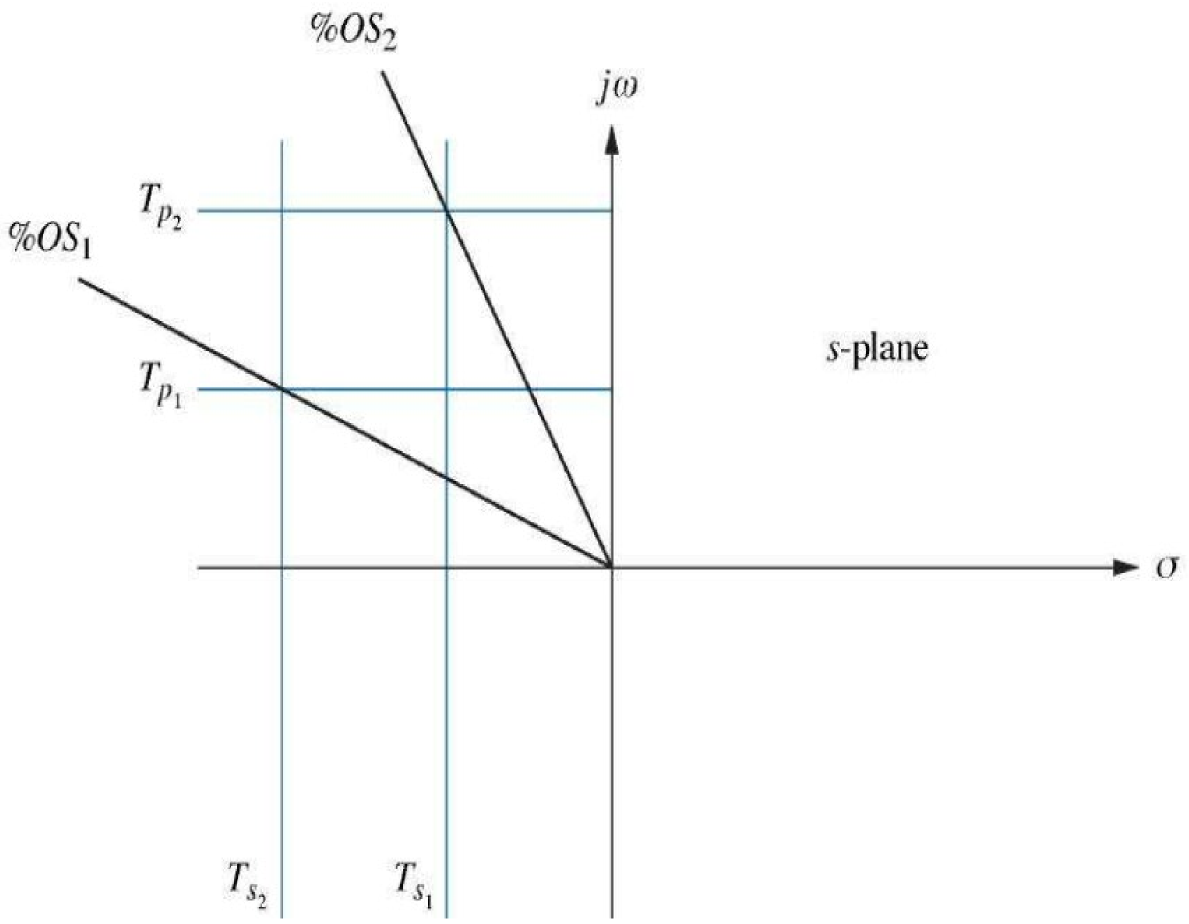


Poles of a second order underdamped system

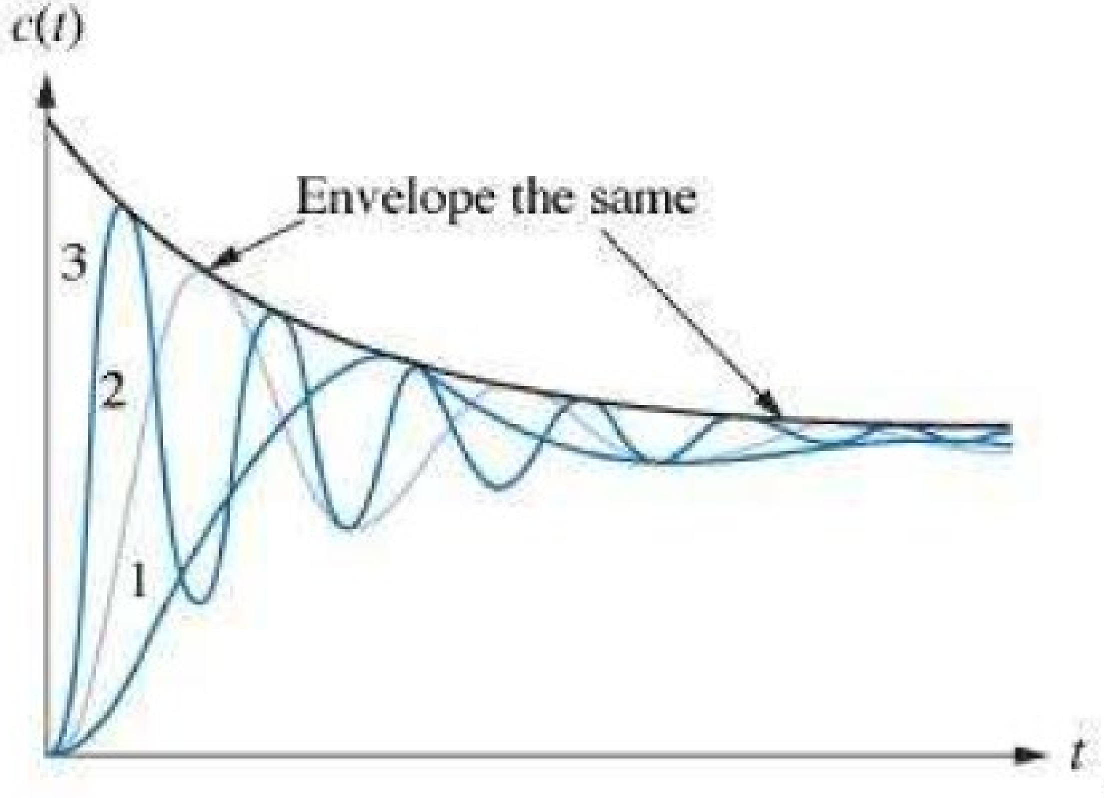
Lines of Constants T_p, T_s and %OS on s-Plane

- Note that horizontal lines on s-plane are lines of constant ω_d consequently they represent lines of constant Tp. Also vertical lines represent constant values of and are therefore lines of constant Ts.
- Finally, since $\zeta = cos\theta$, radial lines represent lines of constant damping ratio. But %Overshoot depends only on ζ .

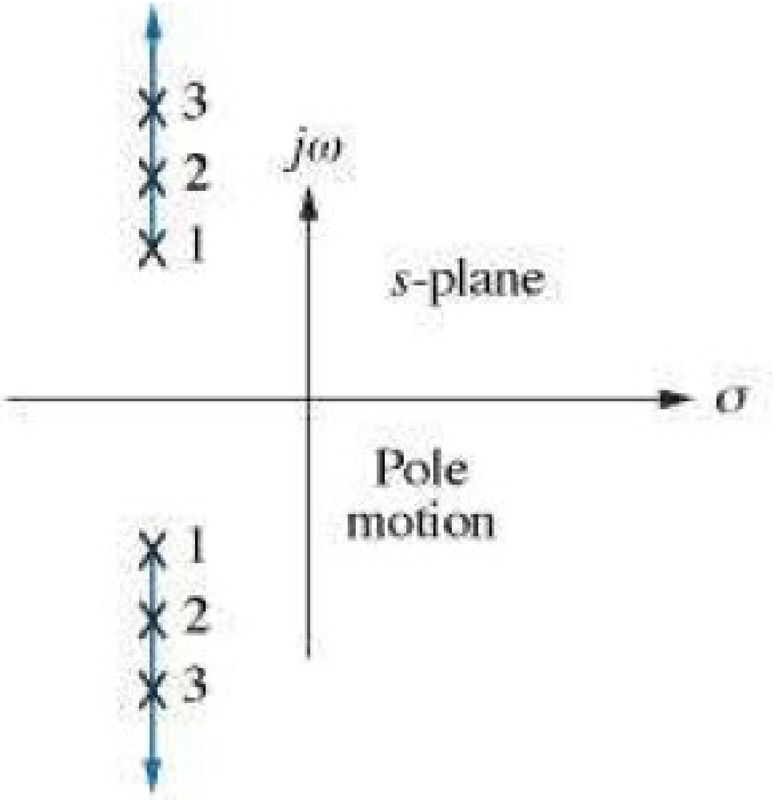
$$\%\text{OS} = e^{-\frac{\pi\zeta}{\sqrt{1-\zeta^2}}} \times 100$$



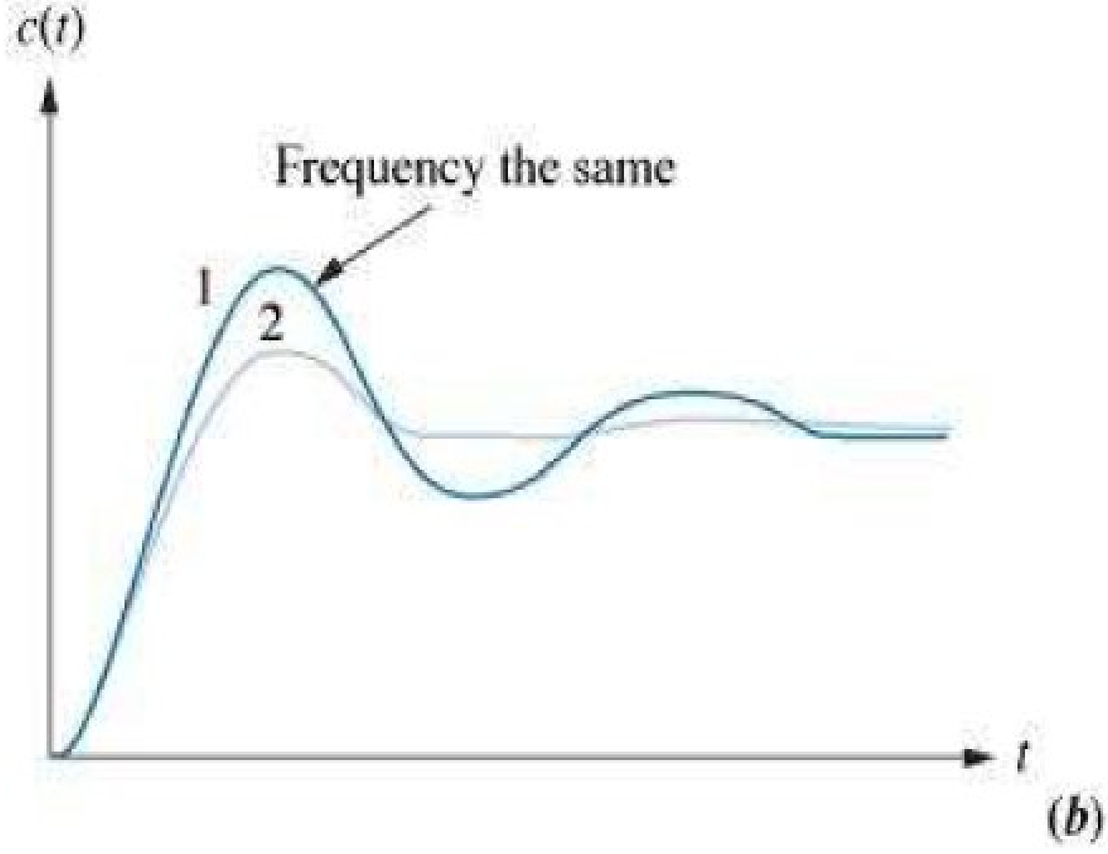
Location Of Poles vs Response (1)

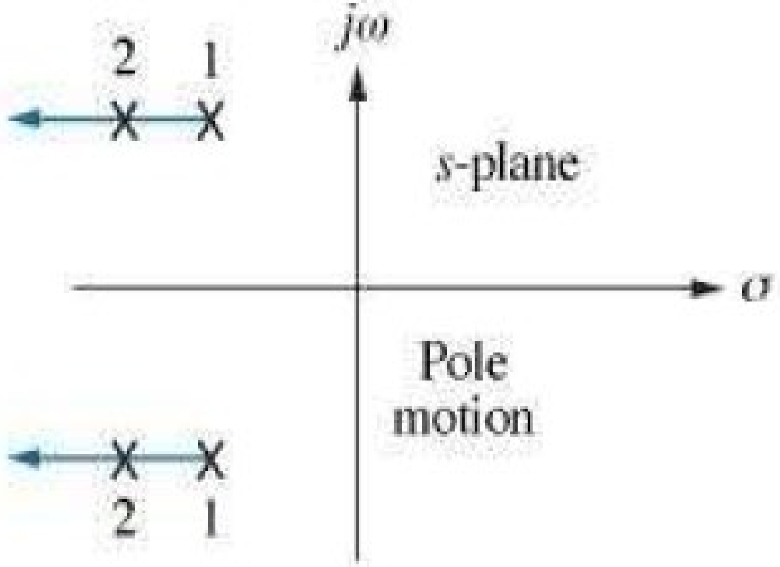


(a)

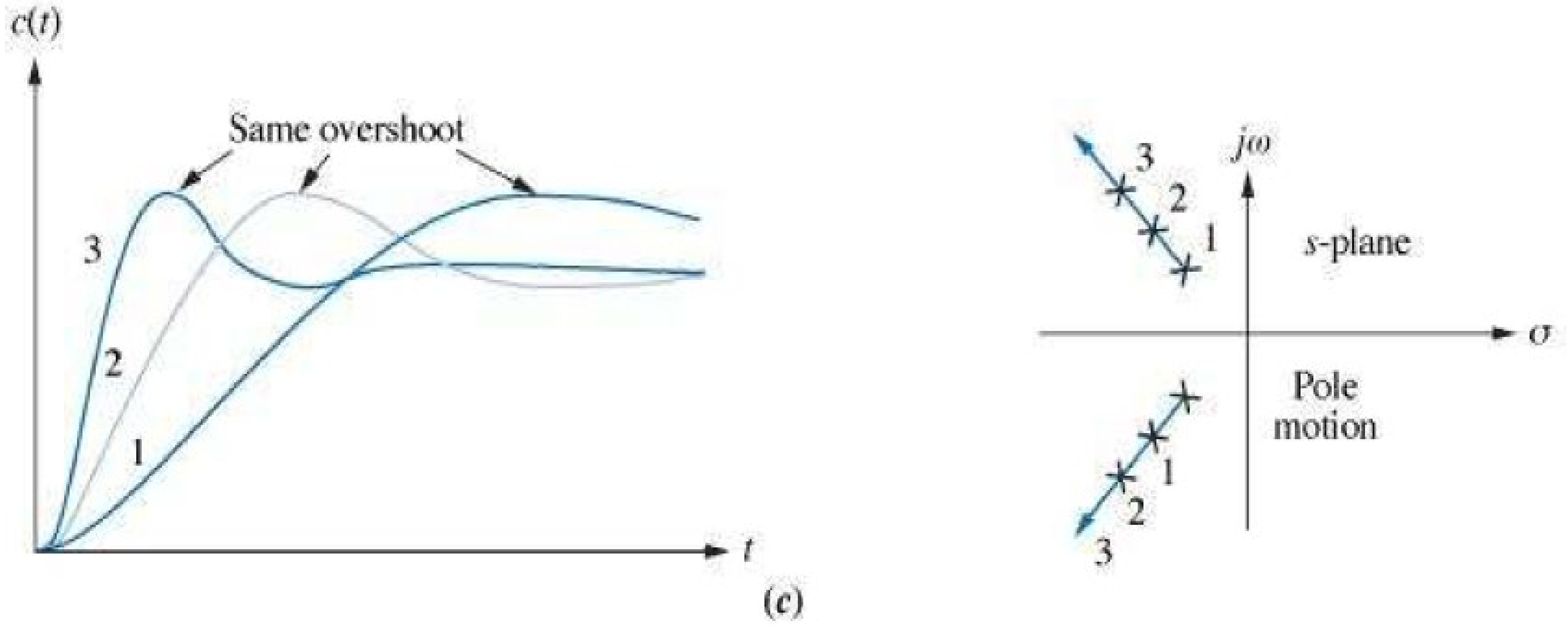


Location Of Poles vs Response (2)



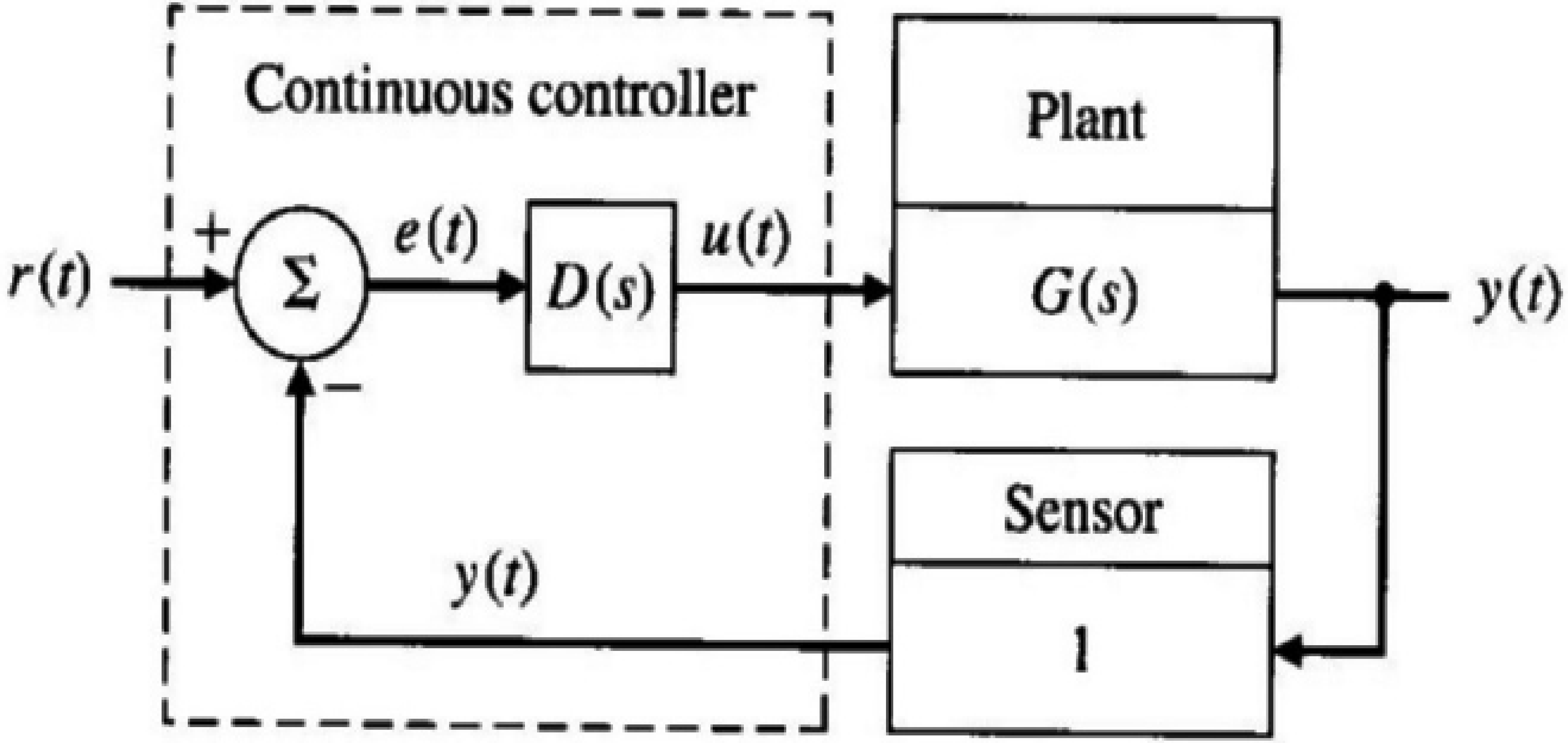


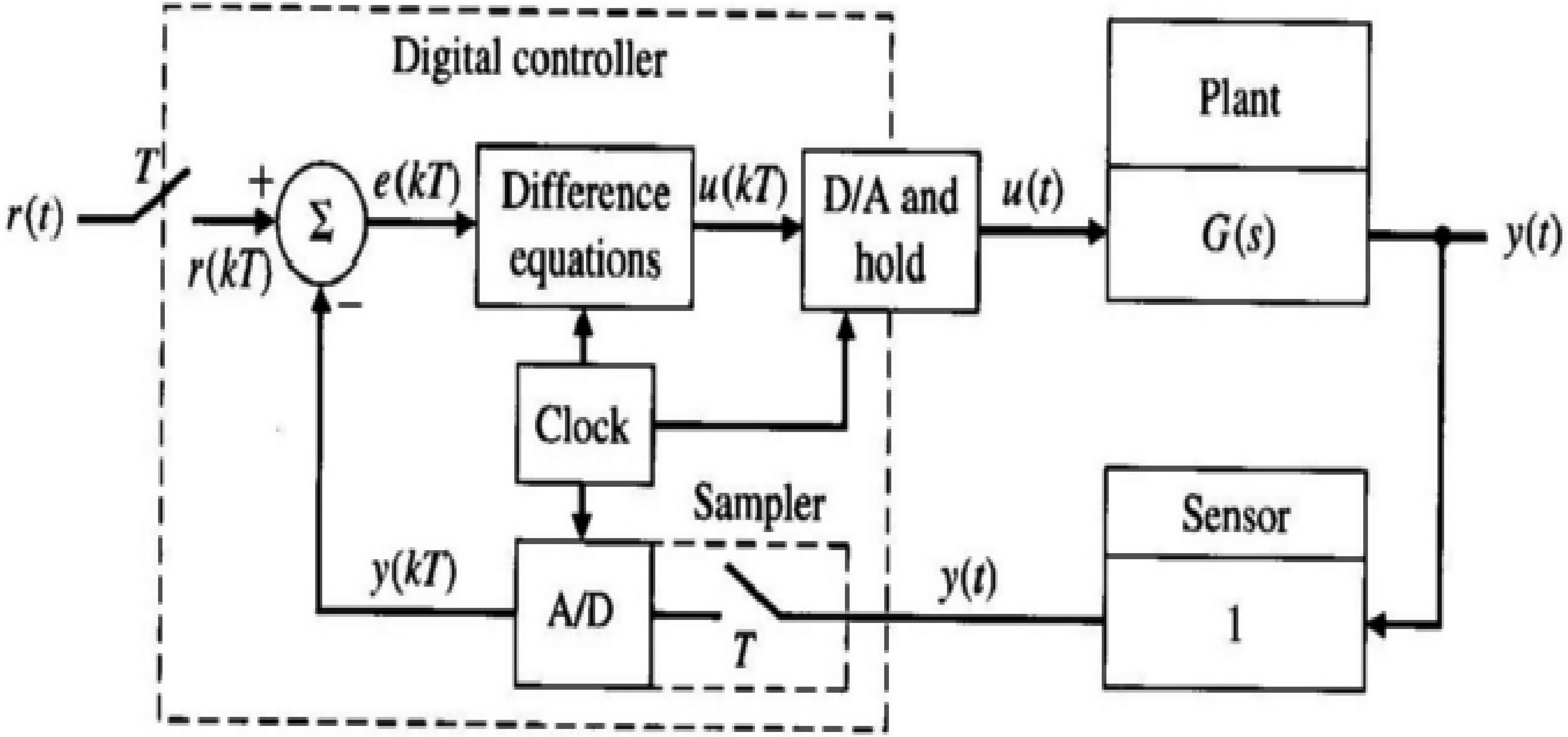
Location Of Poles vs Response (3)



Real Time Systems and Control Applications

Contents


Continuous controller
Plant
Sensor
$r(t)$
+
Σ
$e(t)$
$D(s)$
$u(t)$
$G(s)$
$y(t)$
−
$y(t)$
1



Digital controller
$r(t)$
T
$+$
Σ
$-$
$r(kT)$
$e(kT)$
Difference equations
$u(kT)$
D/A and hold
$u(t)$
Plant
$G(s)$
$y(t)$
Clock
Sampler
$y(kT)$
A/D
T
$y(t)$
Sensor
1

Goals

- Introduce suitable mathematical models for analysis on discrete systems, so that we can build on the existing knowledge of continuous control systems (CCS).

- Show how we convert CCS transfer functions into equivalent transfer functions for DCS.

- If a stable CCS is transformed into a DCS, determine if the DCS remains stable?

Sampling (Analog to Digital Converter)

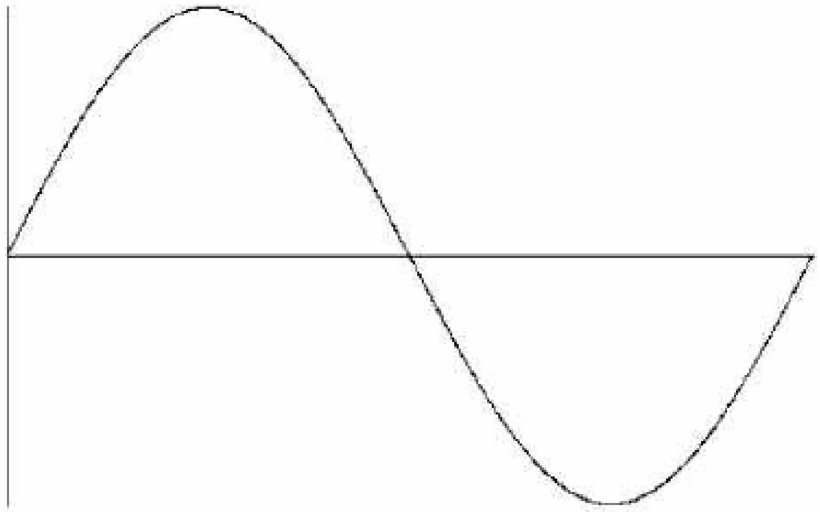
- The analog signal is thus represented by a sequence of sampled values.

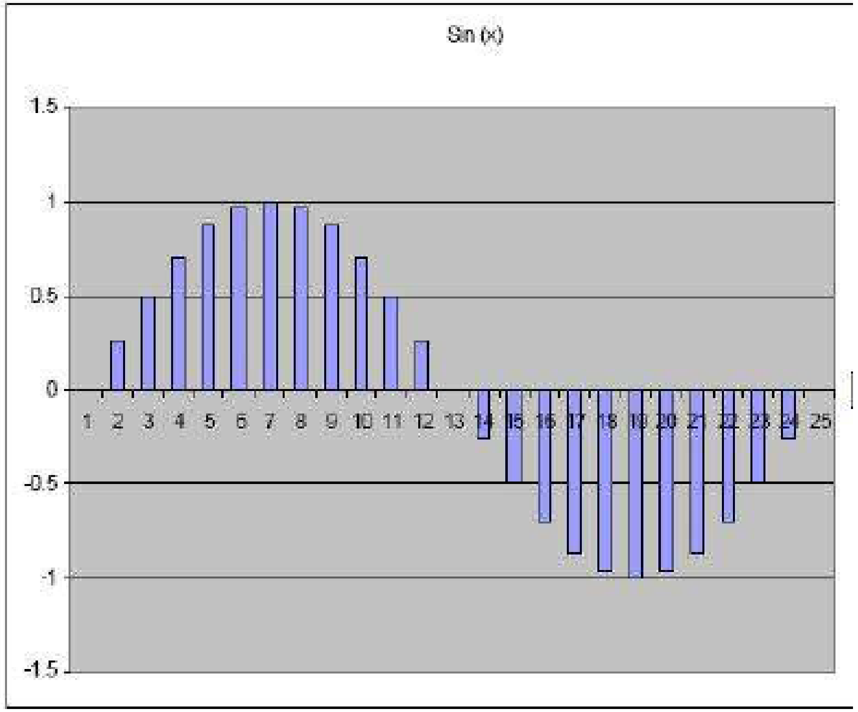
- The conversion of analog signal takes place repetitively at instants of time that are T seconds apart, T is called the sampling period, and 1/T is the sampling rate in cycles per second.

- The accuracy of the digital signal depends on the number of bits used to represent the samples.

Analog to Digital Converter

- Sampling: the process of taking values at discrete time intervals
- Amplitude values are represented using binary numbers, which have finite resolution



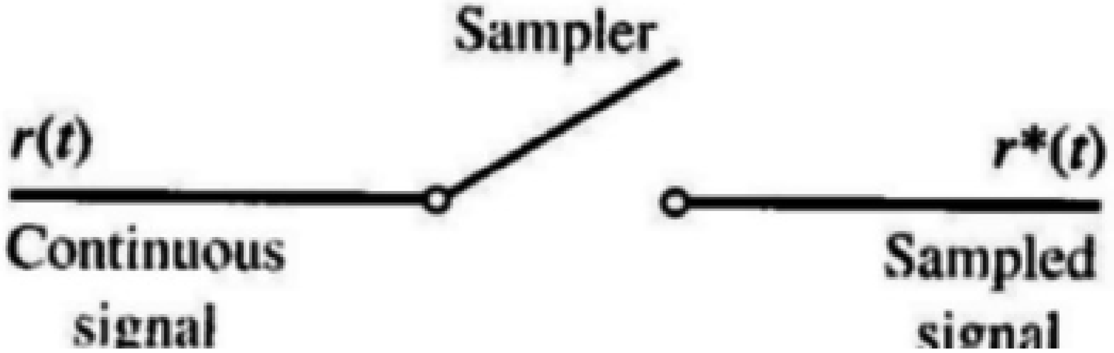


Quantization Error

- An example: Assume that three bits are used to represent a sampled value as a binary number. Let M be the maximum analog voltage that is divided into 8 levels of M/8 volts.
- A 3-bit number can represent all of the eight levels. Values that fall in between these levels must be approximated to next higher/lower binary values.
- This may result in a maximum error of M/14 volts in a sampled digital output. In general **quantization error** is equal to $\frac{1}{2} \times \frac{M}{2^n - 1} = \frac{M}{2^{n+1} - 2}$, where n is the number of bits used for digitization.
- The **resolution** of A/D converter is the minimum value of the output that can be represented as a binary number, i.e., $\frac{M}{2^n}$.

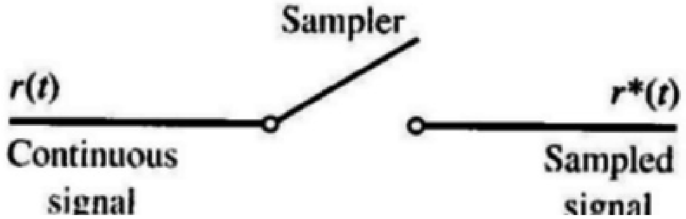
Sampled Data System

- Input signals are available only at sample intervals of time (A2D conversion)
- Thus the reference input r is a sequence of sample values r(kT), instead of r(t)
- A sampler is basically a switch that closes every T seconds for one instant



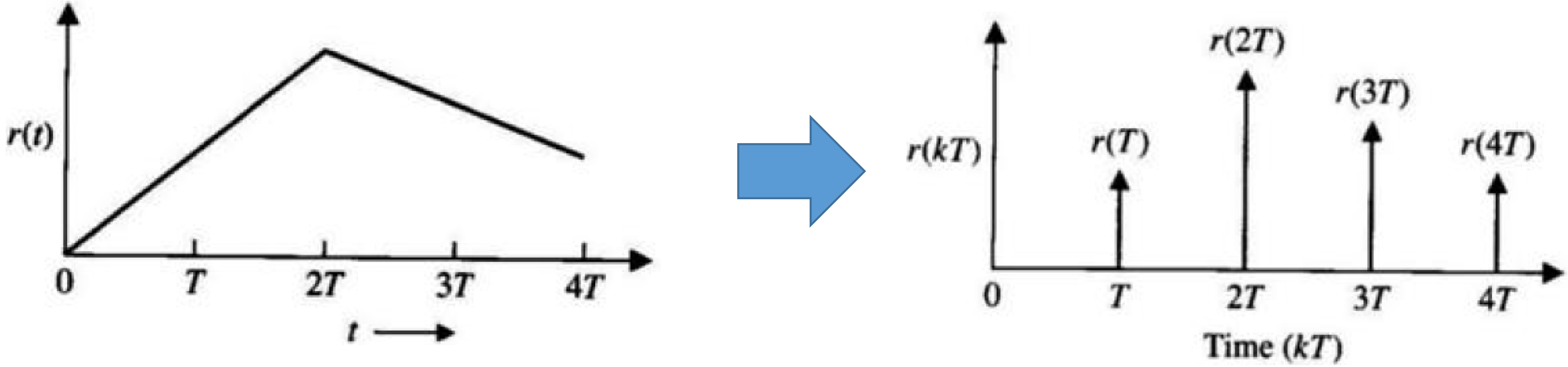
$$r(kT)\,\delta(t-kT)$$

Example of Sampled Data System



$r(kT)\,\delta(t-kT)$

- Assume that we sample a signal r(t) and obtain r*(t).



- We can write

$$r^*(t)=\sum_{k=0}^{\infty} r(kT)\,\delta(t-kT) \qquad (t>0)$$

Transfer Function of Sampled Data System

- Consider $r^*(t) = \sum_{k=0}^{\infty} r(kT)\,\delta(t-kT)$, using Laplace Transform we have:

$$R^*(s) = \mathcal{L}\big(r^*(t)\big) = \sum_{k=0}^{\infty} r(kT)\,e^{-ksT}$$

- We can transform the above series to a more manageable expression

by defining **z-Transform**.

Z-Transform (Laplace Transform of Sampled Data)

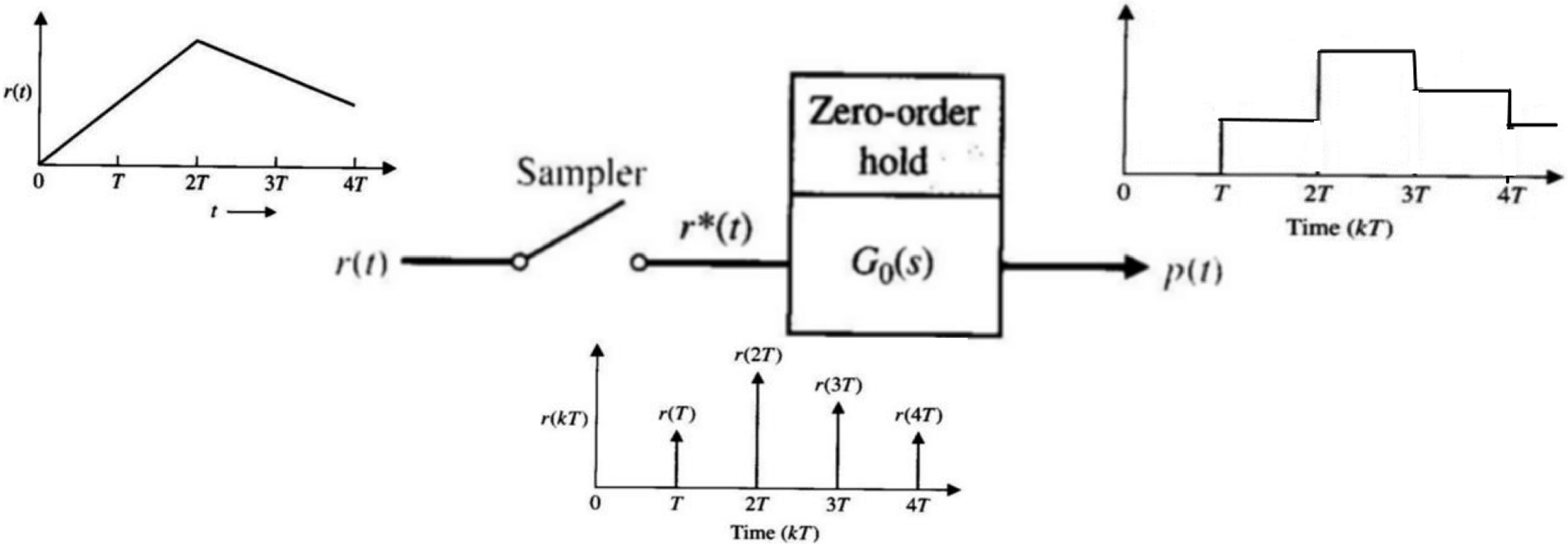
- In $\mathcal{L}\left(r^*(t)\right) = \sum_{k=0}^{\infty} r(kT)\, e^{-ksT}$, if we define $z = e^{sT}$, now the z-Transform is defined as:

$$\mathrm{Z}\{r(t)\} = Z\left(r^*(t)\right) = \sum_{k=0}^{\infty} r(kT)\, z^{-k}$$

- In general, the z-Transform of a function $f(t)$ is defined as:

$$\boldsymbol{Z\{f(t)\} = F(z) = \sum_{k=0}^{\infty} f(kT)\, z^{-k}}$$

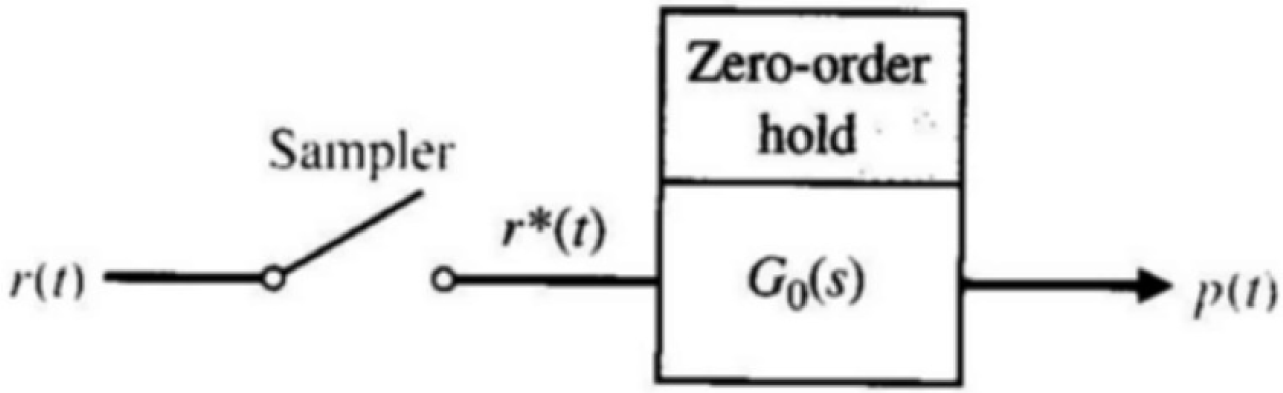
A2D and D2A

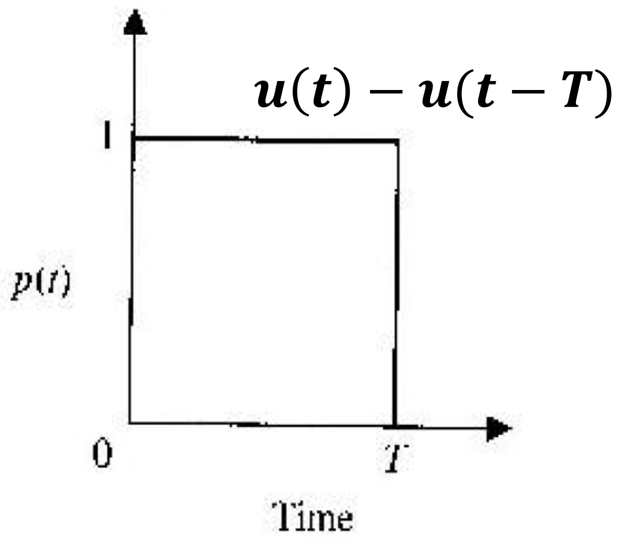


- r(t) and p(t) are not same!
- If T is small enough, p(t) is close to r(t).

Zero Order Hold (Digital to Analog Converter)

- A device that holds the sampled signal r(t) to a constant value for duration of the sampling period. (i.e., The zoh takes the value r(kT) and holds it constant for kT < t < (k+1)T.)





Transfer Function of Zero Order Hold $\mathcal{L}\left(\boldsymbol{u(t) - u(t-T)}\right) = \frac{1}{s} - \frac{e^{sT}}{s}$

Example 1

- Let $f(t) = e^t$ for t $\geq$ 0, what is $Z\big(f(t)\big)$?

$$Z\big(f(t)\big) = \sum_{k=0}^{\infty} e^{kT} z^{-k} = \sum_{k=0}^{\infty} (\frac{e^T}{z})^k = \frac{1}{1 - \frac{e^T}{z}} = \frac{z}{z - e^T}$$

Sum of geometric series

We also have $Z(e^{at}) = \frac{z}{z-e^{aT}}$

Example 2

$$Z(e^{at}) = \frac{z}{z - e^{aT}}$$

- Let $f(t) = \sin(\omega t)$ for t ≥ 0. What is $Z\big(f(t)\big)$?

- We write $\sin(\omega t) = \frac{e^{j\omega t} - e^{-j\omega t}}{2j}$ by Euler's Formula.

$$Z(f(t)) = Z(\frac{e^{j\omega t}}{2j} - \frac{e^{-j\omega t}}{2j}) = \frac{1}{2j}(\frac{z}{z - e^{j\omega T}} - \frac{z}{z - e^{-j\omega T}})$$

$$= \frac{1}{2j}\ \frac{z(e^{j\omega T} - e^{-j\omega T})}{z^2 - z(e^{j\omega T} + e^{-j\omega T}) + 1} = \frac{z\sin(\omega T)}{z^2 - 2z\ \cos(\omega T) + 1}$$

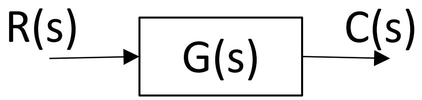
Partial Table of z-Transforms

TABLE 2-3 z-Transforms

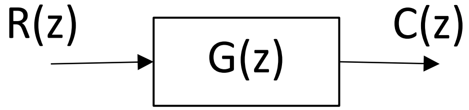
Sequence	Transform
$\delta(k-n)$	z^{-n}
1	$\dfrac{z}{z-1}$
k	$\dfrac{z}{(z-1)^2}$
k^2	$\dfrac{z(z+1)}{(z-1)^3}$
a^k	$\dfrac{z}{z-a}$
ka^k	$\dfrac{az}{(z-a)^2}$
$\sin ak$	$\dfrac{z\sin a}{z^2-2z\cos a+1}$
$\cos ak$	$\dfrac{z(z-\cos a)}{z^2-2z\cos a+1}$
$a^k \sin bk$	$\dfrac{az\sin b}{z^2-2az\cos b+a^2}$
$a^k \cos bk$	$\dfrac{z^2-az\cos b}{z^2-2az\cos b+a^2}$

Transfer Function

- Transfer function of a continuous system is $G(s) = \frac{C(s)}{R(s)}$, where $\mathrm{R}(\mathrm{s}) = \mathcal{L}(r(t))$ and $\mathrm{C}(\mathrm{s}) = \mathcal{L}(c(t))$.



- The transfer function of a sampled system with discrete input and output is given by $G(z) = \frac{C(z)}{R(z)}$, where $\mathrm{R}(z)$ and $\mathrm{C}(z)$ are z-transform of sampled input and output.



Finding the Discrete Transfer Function G(z)

Steps:

(1) Start with the frequency domain transfer function G(s).

(2) Find g(t) using inverse Laplace Transform Tables.

(3) Then use z-transform tables to find G(z).

Note that the discrete transfer function depends on sampling period T.

Example

- Find the transfer function in z-domain of $G(s) = \frac{s^2+4s+3}{s^3+6s^2+8s}$

$$G(s) = \frac{s^2 + 4s + 3}{s^3 + 6s^2 + 8s} = \frac{0.375}{s} + \frac{0.25}{s+2} + \frac{0.375}{s+4}$$

$$\mathrm{G}(t) = \mathcal{L}^{-1}\left[\frac{0.375}{s} + \frac{0.25}{s+2} + \frac{0.375}{s+4}\right] = 0.375 + 0.25e^{-2t} + 0.375e^{-4t}$$

$$G(z) = Z(0.375 + 0.25e^{-2t} + 0.375e^{-4t}) = 0.375\frac{z}{z-1} + 0.25\frac{z}{z-e^{-2T}} + 0.375\frac{z}{z-e^{-4T}}$$

For a given sampling period T, (i.e., T=0.1), we have

$$G(z) = \frac{z^3 - 1.658z^2 + 0.68z}{z^3 - 2.489z^2 + 2.038z - 0.5488}$$

Inverse z-Transform

- $G(z) \rightarrow x(k)$

- Methods:

(1) Power Series Method

(2) Partial-Fraction Expression Method

(3) Inversion-Formula Method

Power Series Method

- If $G(z)$ is expressed as the ratio of two polynomials in z, we can write $G(z)$ in a power series:

$$G(z) = a_0 + a_1 z^{-1} + a_2 z^{-2} + \cdots$$

- Example:

$$G(z) = \frac{z}{z^2 - 3z + 2}$$

Using long division, we have $G(z) = z^{-1} + 3z^{-2} + 7z^{-3} + 15z^{-4} + \cdots$

Hence, $g(0) = 0, g(T) = 1, g(2T) = 3, g(3T) = 7$ …

Partial-Fraction Expression Method

- Let's consider

$$\frac{z}{z-a} = 1 + az^{-1} + a^2z^{-2} + a^3z^{-3} + a^4z^{-4} + \cdots$$

- Write $G(z)$ in partial fraction expression, and use the above equation to find $g(k)$.

- Example: $G(z) = \frac{z}{(z-1)(z-2)} = \frac{-z}{z-1} + \frac{z}{z-2} = \sum_{k=0}^{\infty}(-1+2^k)z^{-k}$
- Therefore, $g(kT) = 2^k - 1$

How to get partial fraction expression of $G(z) = \frac{z}{z^2-3z+2}$?

- $G(z) = \frac{z}{z^2-3z+2} \Rightarrow G(z) = \frac{z}{(z-1)(z-2)}$
- Write $G(z) = \frac{A}{(z-1)} + \frac{B}{(z-2)} = \frac{z}{(z-1)(z-2)}$, so we can solve A and B.

$$\frac{z}{(z-1)(z-2)} = \frac{A}{(z-1)} + \frac{B}{(z-2)} = \frac{A(z-2)+B(z-1)}{(z-1)(z-2)} = \frac{(A+B)z-(2A+B)}{(z-1)(z-2)}$$

So $(A+B) = 1$ and $(2A+B) = 0$

We got A=-1 and B=2

- We know $G(z) = \frac{-1}{(z-1)} + \frac{2}{(z-2)}$

Another Form of Z Transfer Function

- $G(z) = \frac{z}{(z-1)(z-2)}$
- We first get the partial fraction expression of $\frac{G(z)}{z} = \frac{1}{(z-1)(z-2)}$
- We get $\frac{G(z)}{z} = \frac{-1}{z-1} + \frac{1}{z-2}$ ➔ $G(z) = \frac{-z}{z-1} + \frac{z}{z-2}$

On the previous page, we got $G(z) = \frac{-1}{(z-1)} + \frac{2}{(z-2)}$.

Are these G(z) we got both correct?

$$\frac{z}{z-2} = 1 + 2z^{-1} + 2^2 z^{-2} + 2^3 z^{-3} + \cdots = \sum_{k=0}^{\infty} 2^k z^{-k}$$

Are they the same?

$$\frac{z}{z-1} = 1 + z^{-1} + z^{-2} + z^{-3} + \cdots = \sum_{k=0}^{\infty} z^{-k}$$

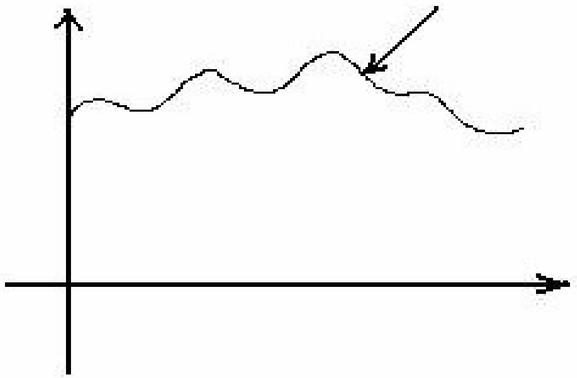
- $G(z) = \frac{-1}{(z-1)} + \frac{2}{(z-2)} = \left[\frac{-z}{(z-1)} + \frac{2z}{(z-2)}\right] z^{-1}$

$$g(kT) = -1 + 2 \times 2^{k-1} = 2^k - 1$$

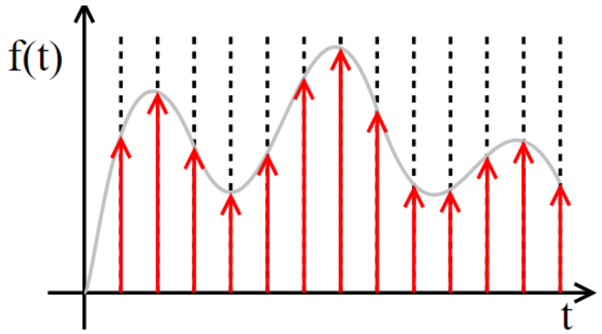
- $(2)\ G(z) = \frac{-z}{z-1} + \frac{z}{z-2}$

$$g(kT) = 2^k - 1$$

A/D, D/A, and Z-Transform



sampling



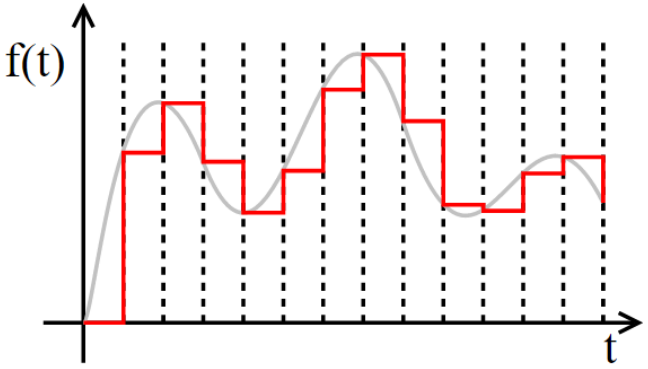
Z-Transform

G(z) represents difference equation relationship between input and output

$$\longrightarrow \boxed{G(z)} \longrightarrow$$

Inverse Z-Transform

z.o,h



End