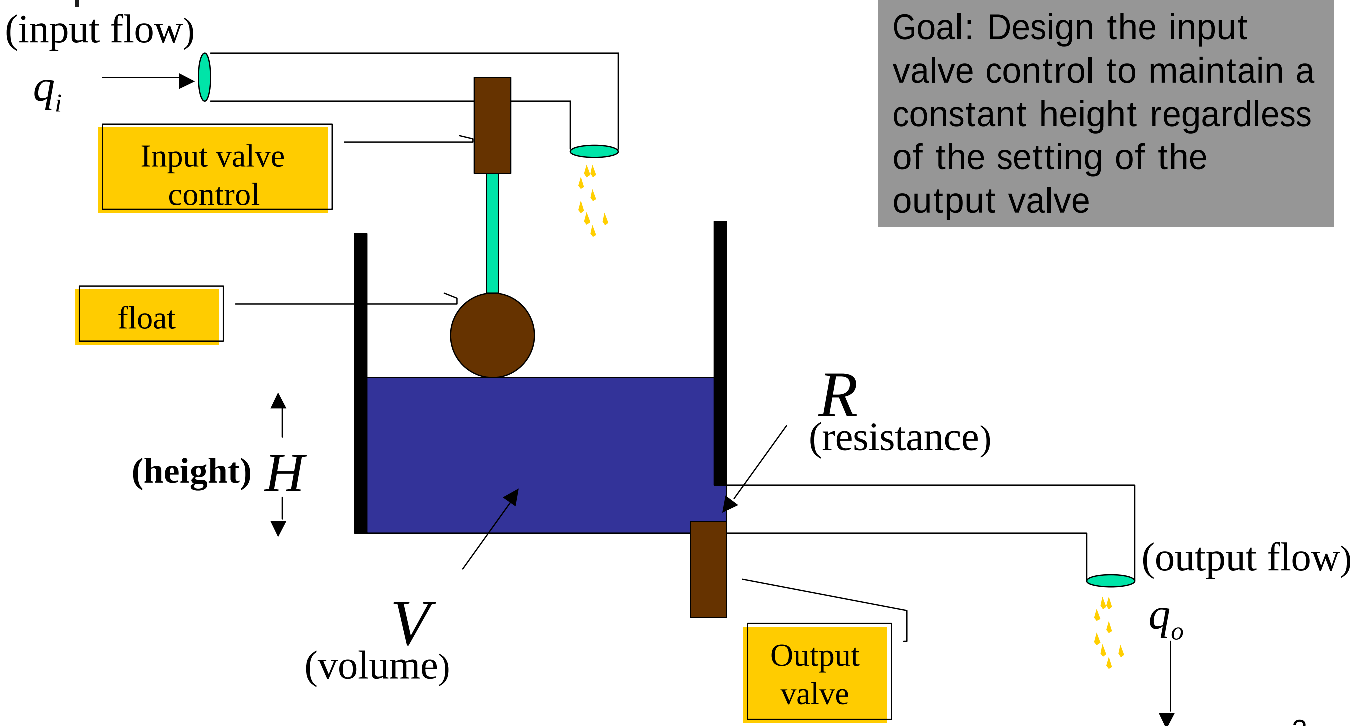
An Introduction to Control Theory With Applications to Computer Science

Joseph Hellerstein
And
Sujay Parekh
IBM T.J. Watson Research Center
{hellers,sujay}@us.ibm.com

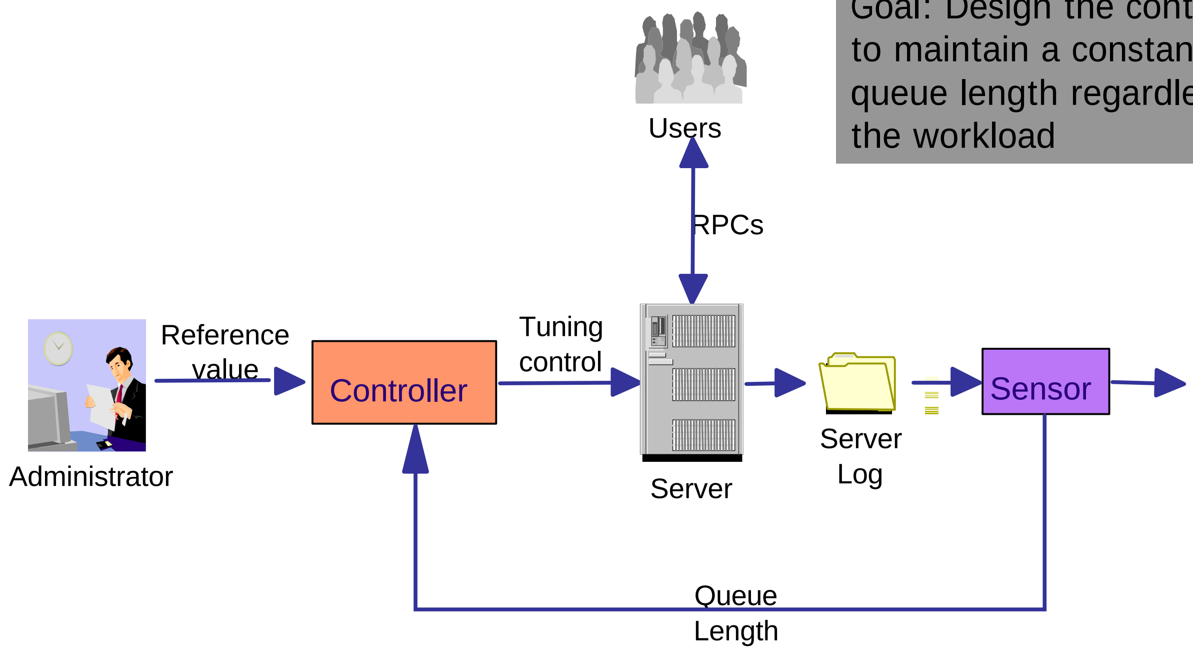
Example 1: Liquid Level System



Goal: Design the input valve control to maintain a constant height regardless of the setting of the output valve

Example 2: Admission Control

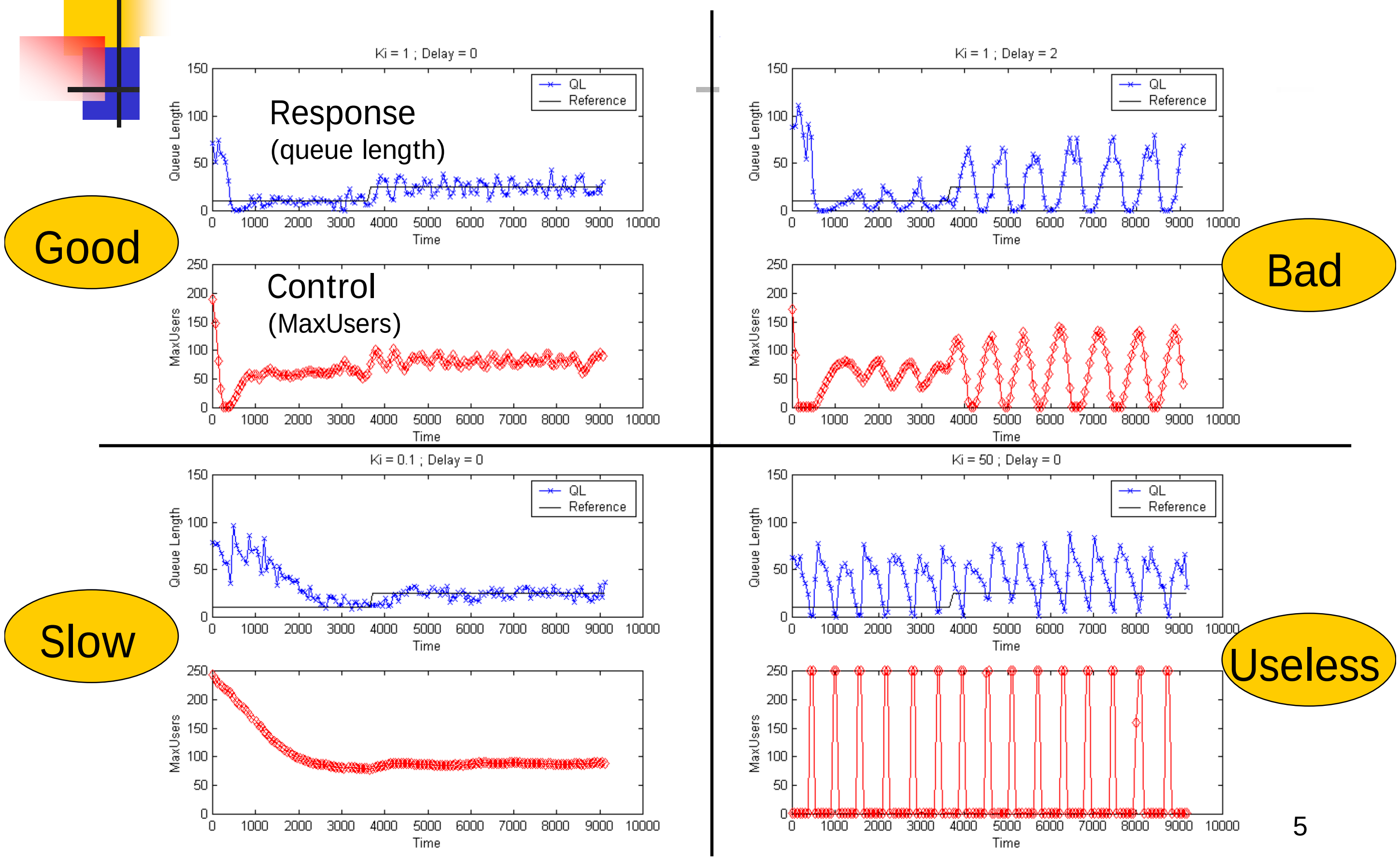
Goal: Design the controller to maintain a constant queue length regardless of the workload



Why Control Theory

- Systematic approach to analysis and design
  - Transient response
  - Consider sampling times, control frequency
  - Taxonomy of basic controls
  - Select controller based on desired characteristics
- Predict system response to some input
  - Speed of response (e.g., adjust to workload changes)
  - Oscillations (variability)
- Approaches to assessing stability and limit cycles

Example: Control & Response in an Email Server



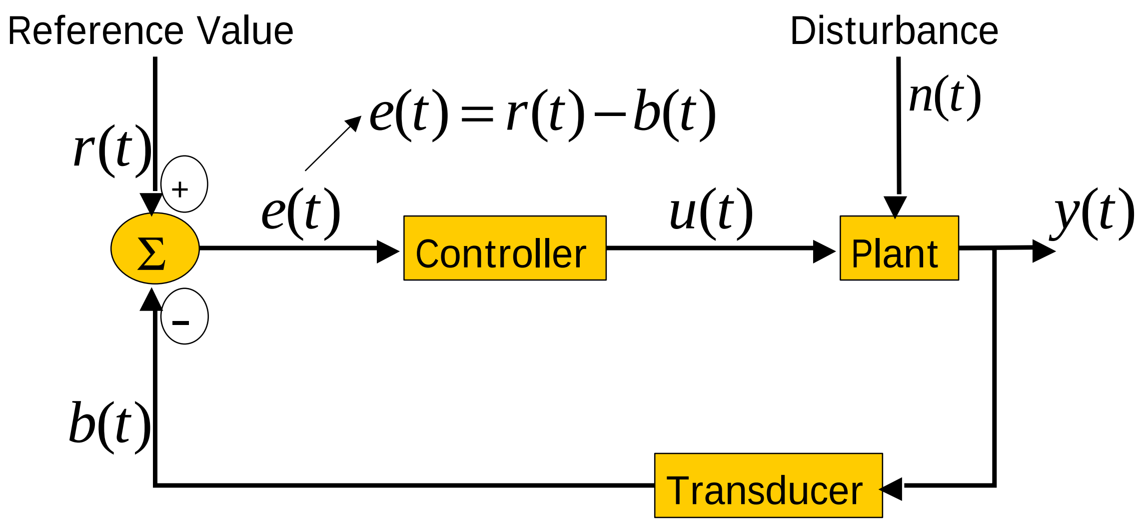
Examples of CT in CS

- Network flow controllers (TCP/IP – RED)
  - C. Hollot et al. (U.Mass)
- Lotus Notes admission control
  - S. Parekh et al. (IBM)
- QoS in Caching
  - Y. Lu et al. (U.Va)
- Apache QoS differentiation
  - C. Lu et al. (U.Va)

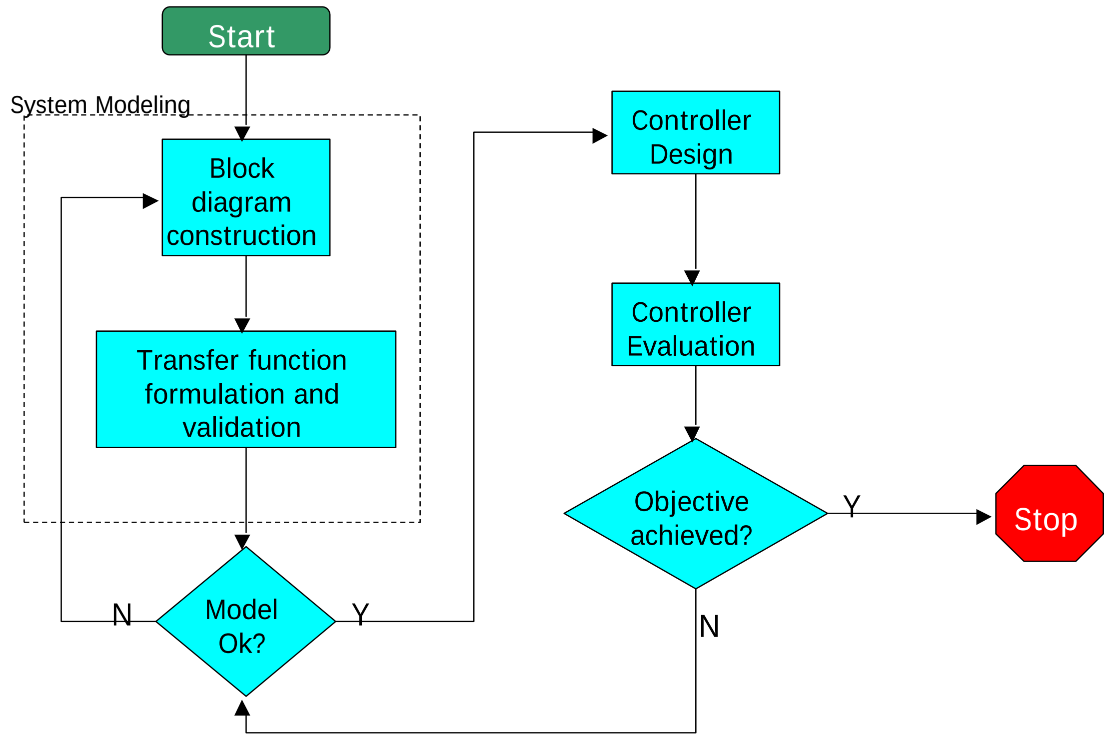
Outline

- Examples and Motivation
- Control Theory Vocabulary and Methodology
- Modeling Dynamic Systems
- Standard Control Actions
- Transient Behavior Analysis
- Advanced Topics
- Issues for Computer Systems
- Bibliography

Feedback Control System



Controller Design Methodology



Control System Goals

- Regulation
  - thermostat, target service levels
- Tracking
  - robot movement, adjust TCP window to network bandwidth
- Optimization
  - best mix of chemicals, minimize response times

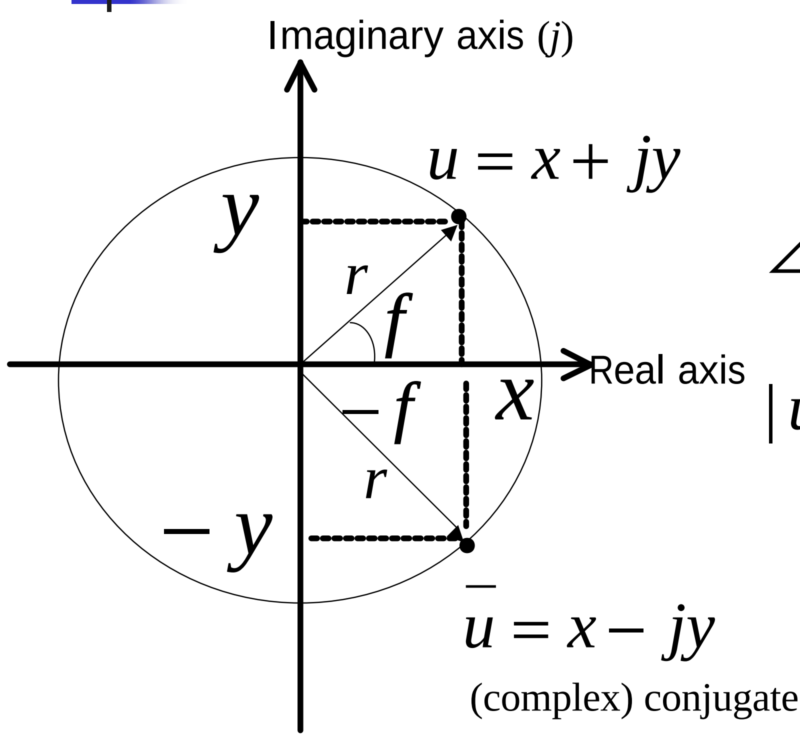
System Models

- **Linear** vs. non-linear (differential eqns)
  - eg, $a_1\dot{y} + a_0 y = b_2\ddot{x} + b_0 x$
  - Principle of superposition
- **Deterministic** vs. Stochastic
- **Time-invariant** vs. Time-varying
  - Are coefficients functions of time?
- **Continuous-time** vs. Discrete-time
  - $t \in R$ vs $k \in Z$

Approaches to System Modeling

- First Principles
  - Based on known laws
    - Physics, Queueing theory
  - Difficult to do for complex systems
- Experimental (System ID)
  - Statistical/data-driven models
  - Requires data
  - Is there a good “training set”?

The Complex Plane (review)



$$\angle u \equiv f = \tan^{-1}\frac{y}{x}$$

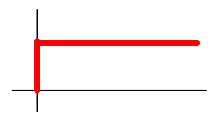
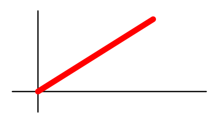
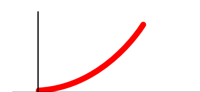
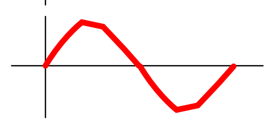
$$|u| \equiv r \equiv |\bar{u}| = \sqrt{x^2 + y^2}$$

Basic Tool For Continuous Time: Laplace Transform

$$\mathbf{L}[f(t)] = F(s) = \int_0^\infty f(t)e^{-st}dt$$

- Convert time-domain functions and operations into frequency-domain
  - $f(t) \rightarrow F(s) \quad (t \in \mathbb{R},\ s \in \mathbb{C})$
  - Linear differential equations (LDE) $\rightarrow$ algebraic expression in Complex plane
- Graphical solution for key LDE characteristics
- Discrete systems use the analogous z-transform

Laplace Transforms of Common Functions

Name	$f(t)$		$F(s)$
Impulse	$f(t)=\begin{cases}1 & t=0\\ 0 & t>0\end{cases}$		1
Step	$f(t)=1$		$\frac{1}{s}$
Ramp	$f(t)=t$		$\frac{1}{s^2}$
Exponential	$f(t)=e^{at}$		$\frac{1}{s-a}$
Sine	$f(t)=\sin(wt)$		$\frac{1}{w^2+s^2}$

Laplace Transform Properties

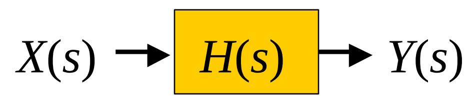
Addition/Scaling	$L[af_1(t) \pm bf_2(t)] = aF_1(s) \pm bF_2(s)$
Differentiation	$L\left[\frac{d}{dt} f(t)\right] = sF(s) - f(0\pm)$
Integration	$L\left[\int f(t)dt\right] = \frac{F(s)}{s} + \frac{1}{s}\left[\int f(t)dt\right]_{t=0\pm}$
Convolution	$\int_0^t f_1(t-t)f_2(t)dt = F_1(s)F_2(s)$
Initial-value theorem	$f(0+) = \lim_{s\to\infty} sF(s)$
Final-value theorem	$\lim_{t\to\infty} f(t) = \lim_{s\to 0} sF(s)$

Insights from Laplace Transforms

- What the Laplace Transform says about f(t)
  - Value of f(0)
    - Initial value theorem
  - Does f(t) converge to a finite value?
    - Poles of F(s)
  - Does f(t) oscillate?
    - Poles of F(s)
  - Value of f(t) at steady state (if it converges)
    - Limiting value of F(s) as s->0

Transfer Function

- Definition
  - H(s) = Y(s) / X(s)

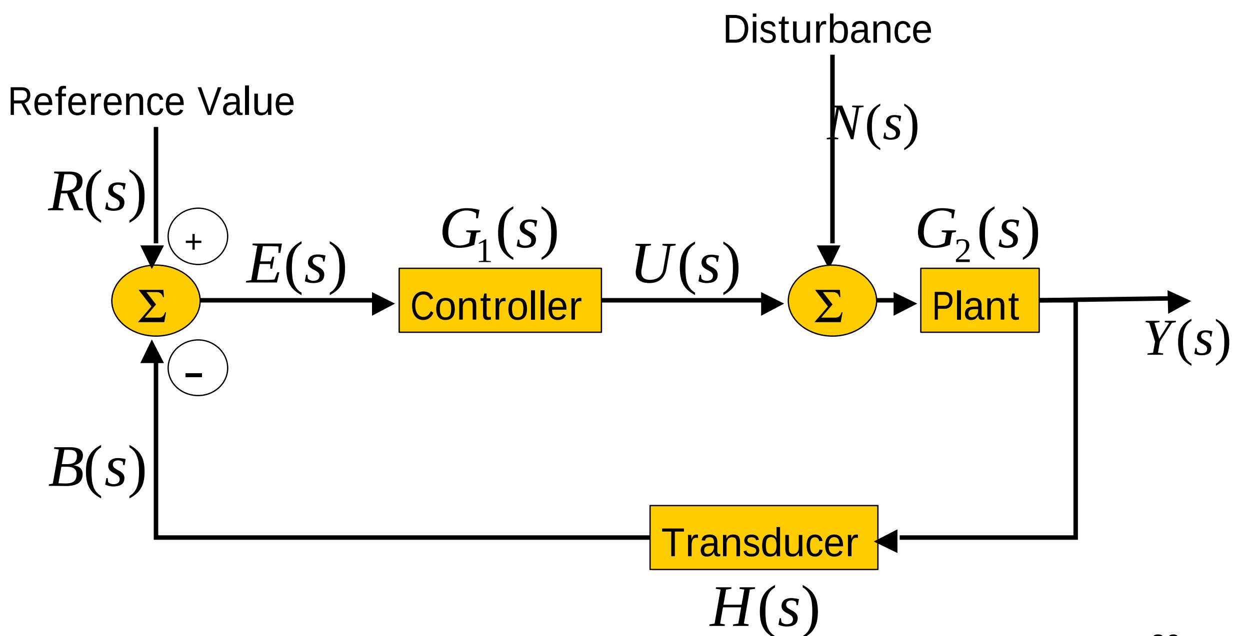


- Relates the output of a linear system (or component) to its input
- Describes how a linear system responds to an impulse
- All linear operations allowed
  - Scaling, addition, multiplication

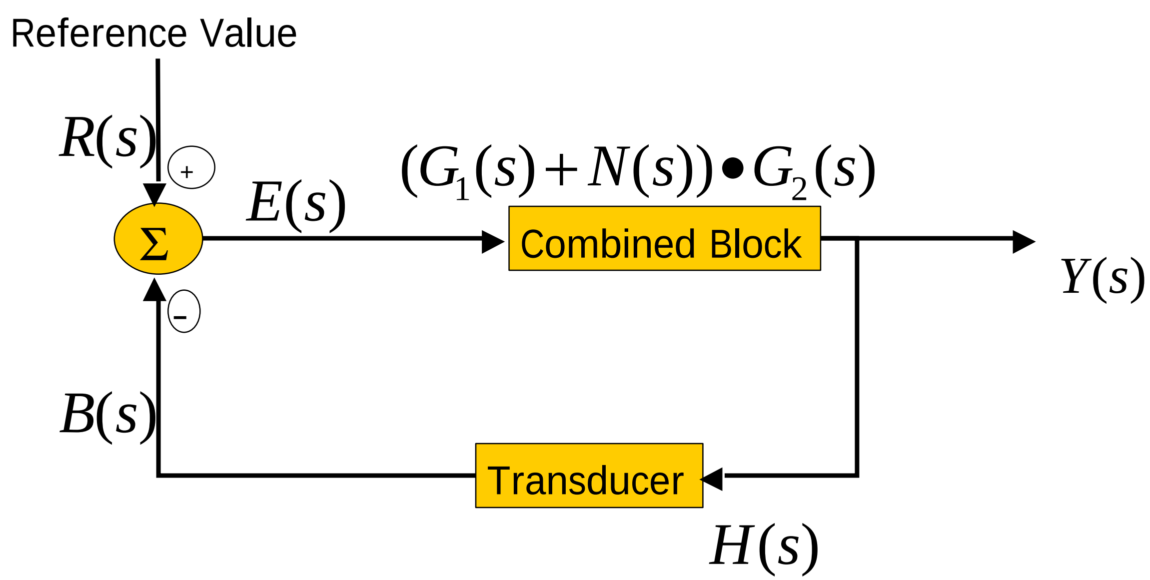
Block Diagrams

- Pictorially expresses flows and relationships between elements in system
- Blocks may recursively be systems
- Rules
  - Cascaded (non-loading) elements: convolution
  - Summation and difference elements
- Can simplify

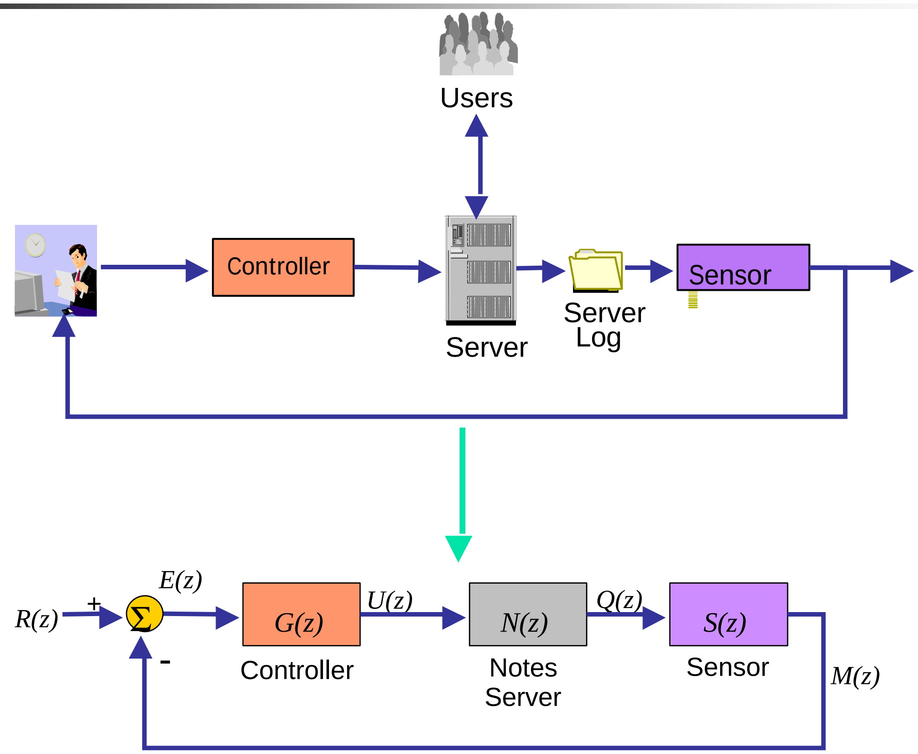
Block Diagram of System



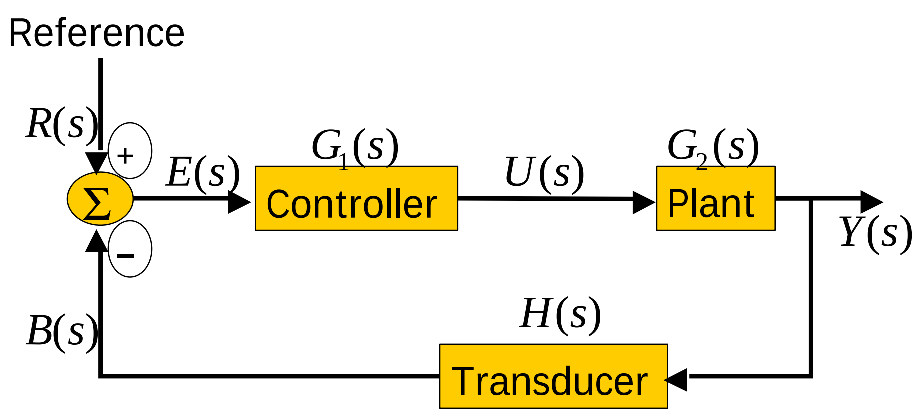
Combining Blocks



Block Diagram of Access Control



Key Transfer Functions



$$\text{Feedforward}: \frac{Y(s)}{E(s)} = \frac{Y(s)}{U(s)}\frac{U(s)}{E(s)} = G_1(s)G_2(s)$$

$$\text{Open-Loop}: \frac{B(s)}{E(s)} = G_1(s)G_2(s)H(s)$$

$$\text{Feedback}: \frac{Y(s)}{R(s)} = \frac{G_1(s)G_2(s)}{1+G_1(s)G_2(s)H(s)}$$

Rational Laplace Transforms

$$F(s) = \frac{A(s)}{B(s)}$$

$$A(s) = a_n s^n + \ldots + a_1 s + a_0$$

$$B(s) = b_m s^m + \ldots + b_1 s + b_0$$

Poles : $s^* \ni B(s^*) = 0$ (So, $F(s^*) = \infty$)

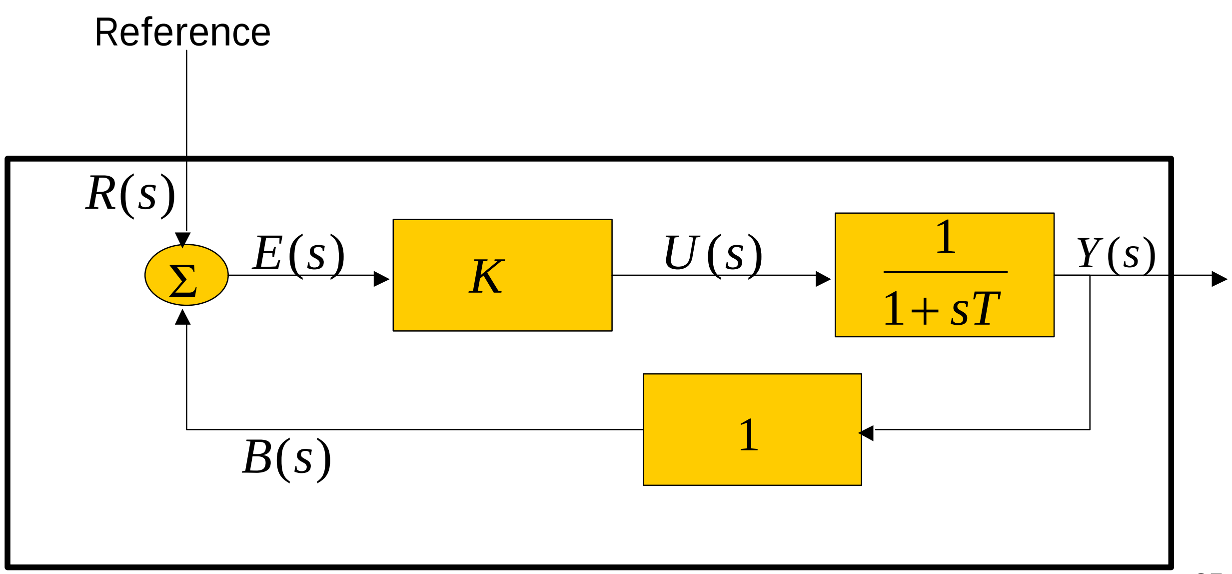
Zeroes : $s^* \ni A(s^*) = 0$ (So, $F(s^*) = 0$)

Poles and zeroes are complex

Order of system $=$ # poles $= m$

First Order System

$$\frac{Y(s)}{R(s)} = \frac{K}{1+K+sT} \approx \frac{K}{1+sT}$$



First Order System

Impulse response	$$\frac{K}{1+sT}$$	Exponential
Step response	$$\frac{K}{s} - \frac{K}{s+1/T}$$	Step, exponential
Ramp response	$$\frac{K}{s^2} - \frac{KT}{s} - \frac{KT}{s+1/T}$$	Ramp, step, exponential

No oscillations (as seen by poles)

Second Order System

Impulse response : $\frac{Y(s)}{R(s)} = \frac{K}{Js^2 + Bs + K} = \frac{w_N^2}{s^2 + 2xw_N s + w_N^2}$

Oscillates if poles have non - zero imaginary part (ie, $B^2 - 4JK < 0$)

Damping ratio : $x = \frac{B}{B_c}$ where $B_c = 2\sqrt{JK}$

Undamped natural frequency : $w_N = \sqrt{\frac{K}{J}}$

Second Order System: Parameters

Interpretation of damping ratio

$x = 0$: Undamped oscillation ($\mathrm{Re} = 0,\ \mathrm{Im} \neq 0$)

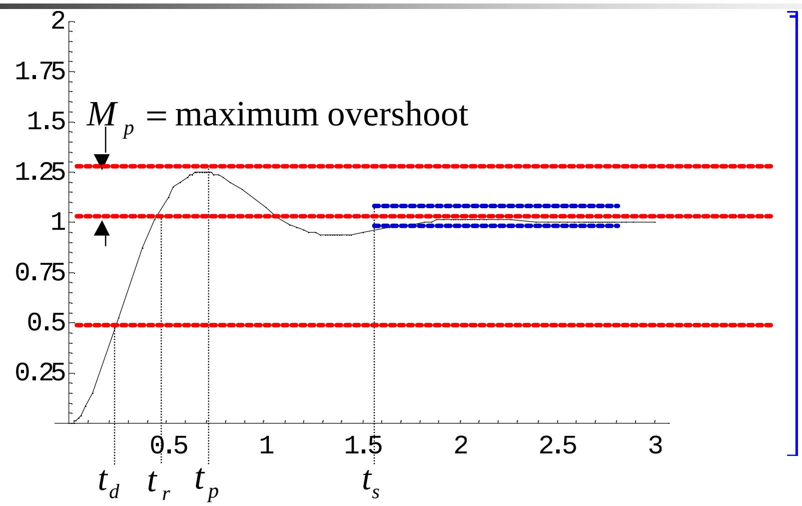
$0 < x < 1$: Underdamped ($\mathrm{Re} \neq 0 \neq \mathrm{Im}$)

$1 \leq x$: Overdamped ($\mathrm{Re} \neq 0, \mathrm{Im} = 0$)

Interpretation of undamped natural frequency

w_{N} gives the frequency of the oscillation

Transient Response Characteristics



t_d : Delay until reach 50% of steady state value

t_r : Rise time = delay until first reach steady state value

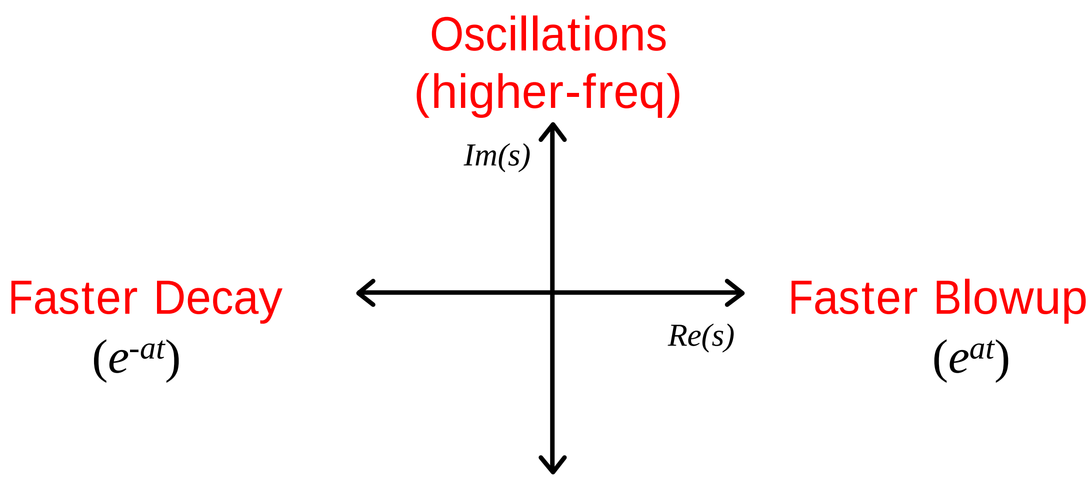
t_p : Time at which peak value is reached

t_s : Settling time = stays within specified % of steady state

Transient Response

- Estimates the shape of the curve based on the foregoing points on the x and y axis
- Typically applied to the following inputs
  - Impulse
  - Step
  - Ramp
  - Quadratic (Parabola)

Effect of pole locations



Basic Control Actions: u(t)

Proportional control: $u(t) = K_p e(t)$ $\qquad \dfrac{U(s)}{E(s)} = K_p$

Integral control: $u(t) = K_i \int_0^t e(t)dt$ $\qquad \dfrac{U(s)}{E(s)} = \dfrac{K_i}{s}$

Differential control: $u(t) = K_d \dfrac{d}{dt} e(t)$ $\qquad \dfrac{U(s)}{E(s)} = K_d s$

Effect of Control Actions

- Proportional Action
  - Adjustable gain (amplifier)
- Integral Action
  - Eliminates bias (steady-state error)
  - Can cause oscillations
- Derivative Action (“rate control”)
  - Effective in transient periods
  - Provides faster response (higher sensitivity)
  - Never used alone

Basic Controllers

- Proportional control is often used by itself
- Integral and differential control are typically used in combination with at least proportional control
  - eg, Proportional Integral (PI) controller:

$$G(s) = \frac{U(s)}{E(s)} = K_p + \frac{K_I}{s} = K_p\left(1 + \frac{1}{T_i s}\right)$$

Summary of Basic Control

- Proportional control
  - Multiply e(t) by a constant
- PI control
  - Multiply e(t) and its integral by separate constants
  - Avoids bias for step
- PD control
  - Multiply e(t) and its derivative by separate constants
  - Adjust more rapidly to changes
- PID control
  - Multiply e(t), its derivative and its integral by separate constants
  - Reduce bias and react quickly

Root-locus Analysis

- Based on characteristic eqn of closed-loop transfer function
- Plot location of **roots** of this eqn
  - Same as **poles** of closed-loop transfer function
  - Parameter (gain) varied from 0 to ∞
- Multiple parameters are ok
  - Vary one-by-one
  - Plot a root "contour" (usually for 2-3 params)
- Quickly get approximate results
  - Range of parameters that gives desired response

Digital/Discrete Control

- More useful for computer systems
- Time is discrete
  - denoted k instead of t
- Main tool is z-transform

$$\mathbf{Z}[f(k)] = F(z) = \sum_{k=0}^{\infty} f(k) z^{-k}$$

  - $f(k) \rightarrow F(z)$, where z is complex
  - Analogous to Laplace transform for s-domain
- Root-locus analysis has similar flavor
  - Insights are slightly different

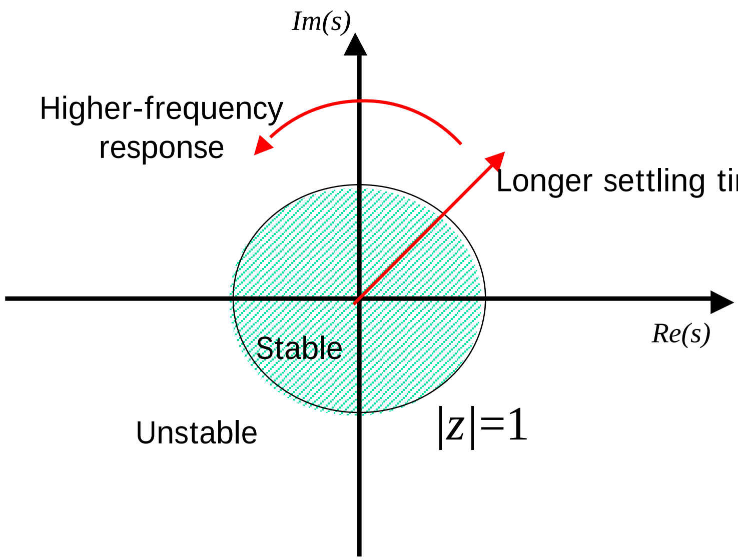
z-Transforms of Common Functions

Name	$f(t)$	$F(s)$	$F(z)$
Impulse	$f(t)=\begin{cases}1 & t=0\\ 0 & t>0\end{cases}$	1	1
Step	$f(t)=1$	$\frac{1}{s}$	$\frac{z}{z-1}$
Ramp	$f(t)=t$	$\frac{1}{s^2}$	$\frac{z}{(z-1)^2}$
Exponential	$f(t)=e^{at}$	$\frac{1}{s-a}$	$\frac{z}{z-e^a}$
Sine	$f(t)=\sin(wt)$	$\frac{1}{w^2+s^2}$	$\frac{z\,\mathrm{Sin}\,a}{z^2-2(\mathrm{Cos}\,a)z+1}$

Root Locus analysis of Discrete Systems

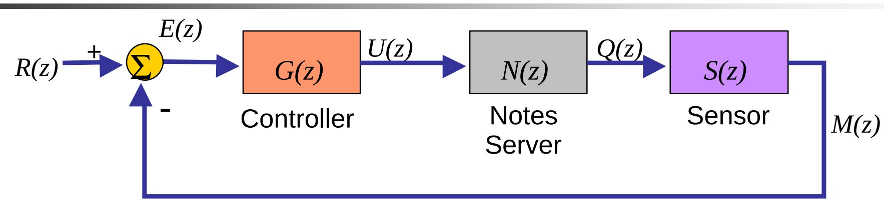
- Stability boundary: $|z|=1$ (Unit circle)
- Settling time = distance from Origin
- Speed = location relative to Im axis
  - Right half = slower
  - Left half = faster

Effect of discrete poles



Intuition: $z = e^{Ts}$

System ID for Admission Control



ARMA Models

$$q(t) = a_1 q(t-1) + b_0 u(t)$$

$$m(t) = c_1 m(t-1) + d_0 q(t) + d_1 q(t-1)$$

Control Law

$$u(t) = u(t-1) + K_i e(t)$$

Transfer Functions

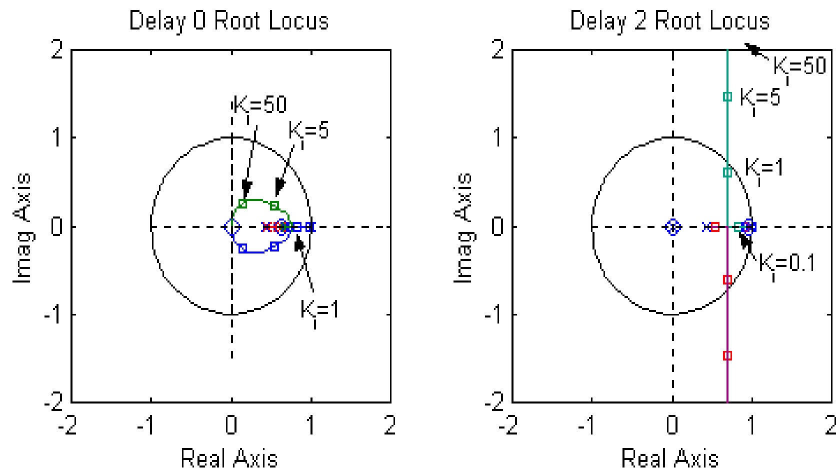
$$N(z) = \frac{b_0 z}{z - a_1}$$

$$S(z) = \frac{d_0 z + d_1}{z - c_1}$$

$$G(z) = \frac{K_i z}{z-1} \frac{1}{z^d}$$

Open-Loop: $$N(z)S(z)G(z) = \frac{b_0 z}{z - a_1} \frac{d_0 z + d_1}{z - c_1} \frac{K_i z}{z-1} \frac{1}{z^d}$$

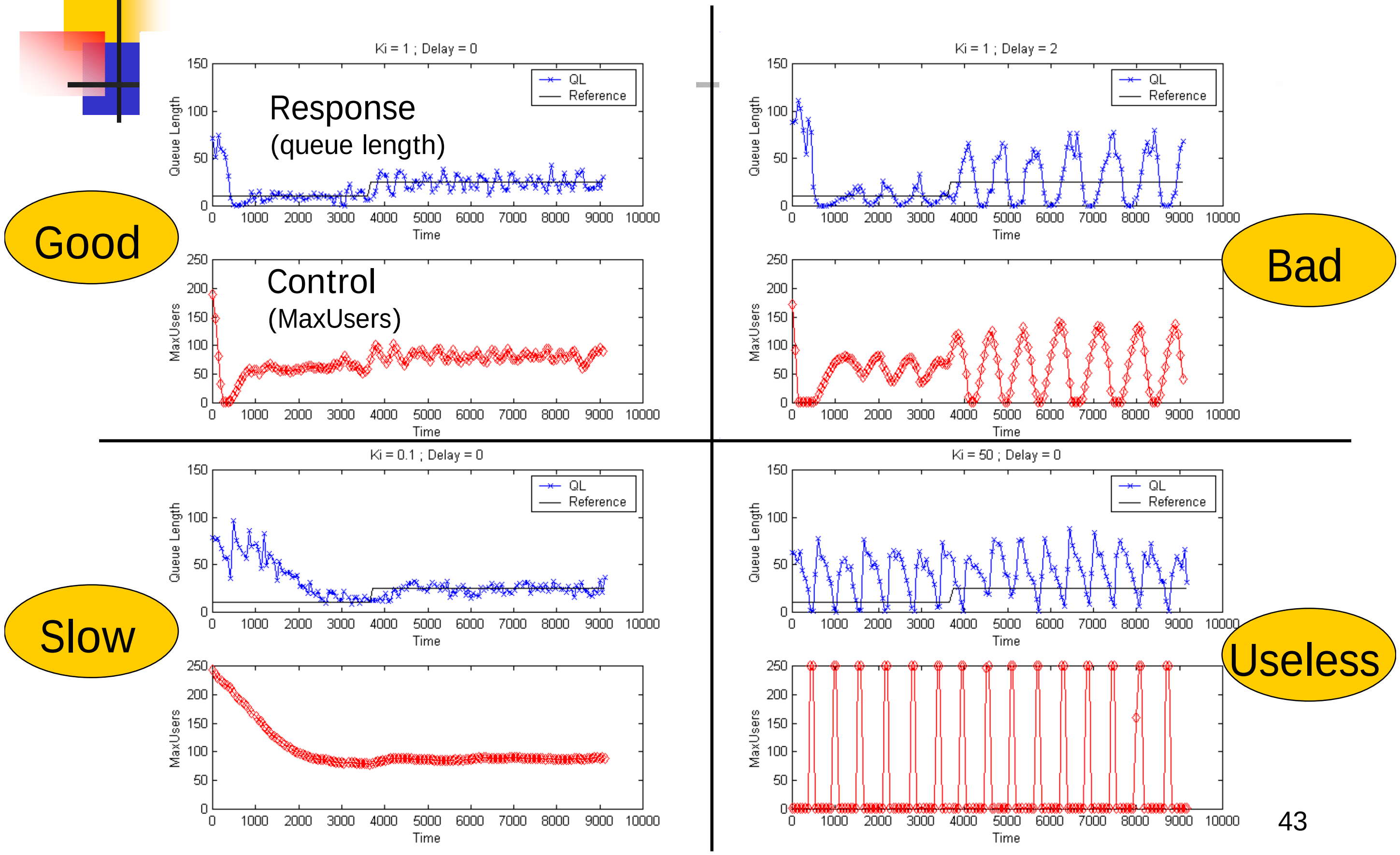
Root Locus Analysis of Admission Control



Predictions:

•K_i small => No controller-induced oscillations
•K_i large => Some oscillations
•K_i v. large => unstable system (d=2)
•Usable range of K_i for d=2 is small

Experimental Results



Advanced Control Topics

- Robust Control
  - Can the system tolerate noise?
- Adaptive Control
  - Controller changes over time (adapts)
- MIMO Control
  - Multiple inputs and/or outputs
- Stochastic Control
  - Controller minimizes variance
- Optimal Control
  - Controller minimizes a cost function of error and control energy
- Nonlinear systems
  - Neuro-fuzzy control
  - Challenging to derive analytic results

Issues for Computer Science

- Most systems are non-linear
  - But linear approximations may do
    - eg, fluid approximations
- First-principles modeling is difficult
  - Use empirical techniques
- Control objectives are different
  - Optimization rather than regulation
- Multiple Controls
  - State-space techniques
  - Advanced non-linear techniques (eg, NNs)

Selected Bibliography

- Control Theory Basics
  - G. Franklin, J. Powell and A. Emami-Naeini. *"Feedback Control of Dynamic Systems, 3rd ed"*. Addison-Wesley, 1994.
  - K. Ogata. *"Modern Control Engineering, 3rd ed"*. Prentice-Hall, 1997.
  - K. Ogata. *"Discrete-Time Control Systems, 2nd ed"*. Prentice-Hall, 1995.
- Applications in Computer Science
  - C. Hollot et al. *"Control-Theoretic Analysis of RED"*. IEEE Infocom 2001 (to appear).
  - C. Lu, et al. *"A Feedback Control Approach for Guaranteeing Relative Delays in Web Servers"*. IEEE Real-Time Technology and Applications Symposium, June 2001.
  - S. Parekh et al. *"Using Control Theory to Achieve Service-level Objectives in Performance Management"*. Int'l Symposium on Integrated Network Management, May 2001
  - Y. Lu et al. *"Differentiated Caching Services: A Control-Theoretic Approach"*. Int'l Conf on Distributed Computing Systems, Apr 2001
  - S. Mascolo. *"Classical Control Theory for Congestion Avoidance in High-speed Internet"*. Proc. 38th Conference on Decision & Control, Dec 1999
  - S. Keshav. *"A Control-Theoretic Approach to Flow Control"*. Proc. ACM SIGCOMM, Sep 1991
  - D. Chiu and R. Jain. *"Analysis of the Increase and Decrease Algorithms for Congestion Avoidance in Computer Networks"*. Computer Networks and ISDN Systems, 17(1), Jun 1989