ACTUATORS & SENSORS IN MECHATRONICS
MEAE 6960   Summer 2002

Instructor: Dr. Kevin Craig, Associate Professor of Mechanical Engineering
Office: JEC 3018, Phone: 276-6626 Laboratory: JEC 1022, Phone: 276-8978
E-mail: craigk@rpi.edu, FAX: 518-276-4860

Class Hours: Wednesday 5:30 – 9 PM (13 sessions: May 15, 29; June 5, 12, 26; July 3, 10, 17, 24, 31; August 7, 14, 21)

Course Description:
Mechatronics, as an engineering discipline, is the synergistic combination of mechanical engineering, electronics, control engineering, and computer science, all integrated through the design process. This course studies actuators and sensors commonly used in the design of mechatronic systems. A balance between the modeling / analysis and the hardware implementation of these various devices is emphasized. Actual mechatronic systems from industry and academia are used to demonstrate actuator and sensor use. Actuators studied include: brushed dc motors, stepper motors, brushless dc motors, solenoids, and hydraulic and pneumatic actuators. Analog and digital sensors studied include: optical encoders, Hall-effect sensors, potentiometers, gyroscopes, variable-inductance transducers, permanent-magnet transducers, eddy-current transducers, variable-capacitance transducers, and piezoelectric transducers. Smart sensors and actuators are discussed.

Course Objectives:

- Understand the fundamental concepts of electromechanics and fluid mechanics (hydraulics and pneumatics).
- Apply these fundamental concepts to the modeling, analysis, and control of brushed dc motors, stepper motors, brushless dc motors, solenoids, and hydraulic and pneumatic actuators.
- Understand the key elements of a measurement system
- Understand the basic performance specifications of analog and digital sensors and actuators
- Become familiar with the operating principles and physical/mathematical models of a variety of analog and digital motion sensors
- Understand the key issues in hardware implementation of analog and digital actuators and sensors
- Become proficient in the use of MatLab/Simulink to model and analyze actuators and sensors for use in mechatronic systems
- Understand what comprises a mechatronic motion control system and the key elements in its design

Prerequisites: Senior or Graduate-Student standing in engineering

References: (not to be purchased)

1. *Electromechanical Motion Devices*, P. Krause and O. Wasynczuk, McGraw Hill, 1989.
2. *Electric Machinery Fundamentals*, 3rd Edition, S. Chapman, McGraw Hill, 1999.
3. *Driving Force, The Natural Magic of Magnets*, J. Livingston, Harvard University Press, 1996.
4. *Applied Electromagnetics*, M. Plonus, McGraw Hill, 1978.
5. *Control Sensors and Actuators*, C. deSilva, Prentice Hall, 1989.
6. *Electromechanical Dynamics,* H. Woodson and J. Melcher, Wiley, 1968.
7. *Electromechanics and Electric Machines,* S. Nasar and L. Unnewehr, Wiley, 1979.
8. *Electric Motor Drives: Modeling, Analysis, and Control,* R. Krishnan, Prentice Hall, 2001.
9. *Stepping Motors: A Guide to Modern Theory and Practice*, 3rd Edition, P. Acarnley, IEE, 1992.
10. *Step Motor System Design Handbook*, 2nd Edition, A. Leenhouts, Litchfield Engineering Company, 1997.
11. *Stepping Motors and Their Microprocessor Controls*, 2nd Edition, T. Kenjo, Oxford, 1994.
12. *Measurement Systems, Application and Design*, 4th Edition, E. Doebelin, McGraw Hill, 1990.
13. *DC Motors, Speed Controls, Servo Systems*, The Electro-Craft Engineering Handbook, Reliance Motion Control, Inc.
14. *Electric Motors and Their Controls,* T. Kenjo, Oxford 1991.
15. *Basics of Design Engineering*, Machine Design, 1994.
16. *Control of Fluid Power, Analysis and Design*, 2nd Edition, D. McCloy and H. Martin, Ellis Horwood Limited, 1980.
17. *Hydraulic Control Systems,* H. Merritt, Wiley, 1967.
18. *The Analysis and Design of Pneumatic Systems,* B. Andersen, Wiley, 1967.

Lecture Topics: (13 periods, each 3.5 hours duration)

1. Fundamentals of Electromechanical Motion Devices (2 periods)
2. Direct-Current Motors (1 period)
3. Stepper Motors (1 period)
4. Brushless DC Motors (1 period)
5. Fundamentals of Fluid Power: Modeling, Analysis, and Control (1 period)
6. Hydraulic Actuators (1 period)
7. Pneumatic Actuators (1 period)
8. Sensors and Measurement Systems (1 period)
9. Analog & Digital Motion Sensors (3 periods)
10. Motion Control Systems (1 period)

Computer Software:

MatLab, Simulink, and the Control System Toolbox are all required for the course. These are available at RPI or can be purchased directly from the MathWorks for individual use. Student versions of the software are acceptable.

Assessment:

***Homework Assignments*:**
10 assignments will be given during the course. Students will have 2 weeks to complete each assignment. They are to be *individually* and professionally done and handed in. Collaborative discussion (Not Copying!) on homework assignments is permitted and encouraged. MatLab / Simulink (and some Toolboxes) will be used in the solution of some of these assignments. The assignments will be graded and will count 75% (10 @ 7.5% each) of the final grade. Regarding late assignments, all reasonable excuses will be accepted provided they are discussed with the instructor *prior* to the due date for the assignment.

Final Examination:
There will be a closed-book, closed-note, final exam covering the entire course. It is worth 25% of the final grade. It will focus on the fundamental concepts studied in the course.

Grade Summary:

Final Exam	25%
Homework Problems (10 @ 7.5%)	75%
Total	100%

Class Attendance and Participation:
Attendance at *all* classes is mandatory and participation in class is strongly encouraged. An effort is being made to make all classes more interactive and thus greatly enhance the learning process.

***Course Conduct*:**
Students are expected to conduct themselves in a professional manner at all times. Any cheating on the final exam will result in a grade of zero for the exam. Appeals should be directed to the Dean of Students.

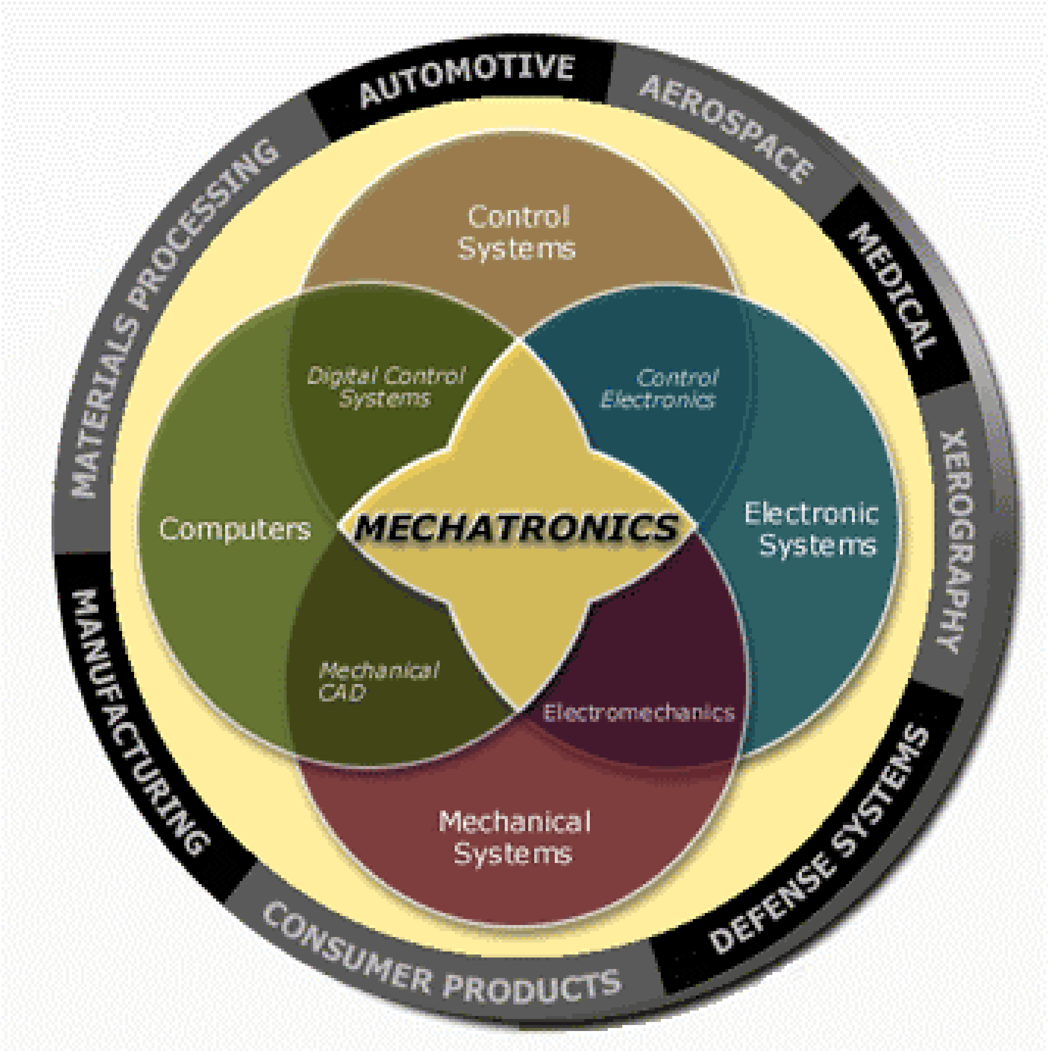
Sensors & Actuators In Mechatronics

Dr. Kevin Craig

Associate Professor of Mechanical Engineering
Department of Mechanical, Aerospace, and Nuclear Engineering
Rensselaer Polytechnic Institute
Troy, NY 12180
Office: JEC 3018 Phone: 518-276-6626
E-Mail: craigk@rpi.edu Fax: 518-276-4860
Mechatronics Laboratory: JEC 1022 Phone: 518-276-8978

What is Mechatronics ?

Mechatronics is the ***synergistic*** combination of mechanical engineering, electronics, controls engineering, and computers, all ***integrated*** through the design process.



The Design Challenge

The cost-effective incorporation of electronics, computers, and control elements in mechanical systems requires a new approach to design.

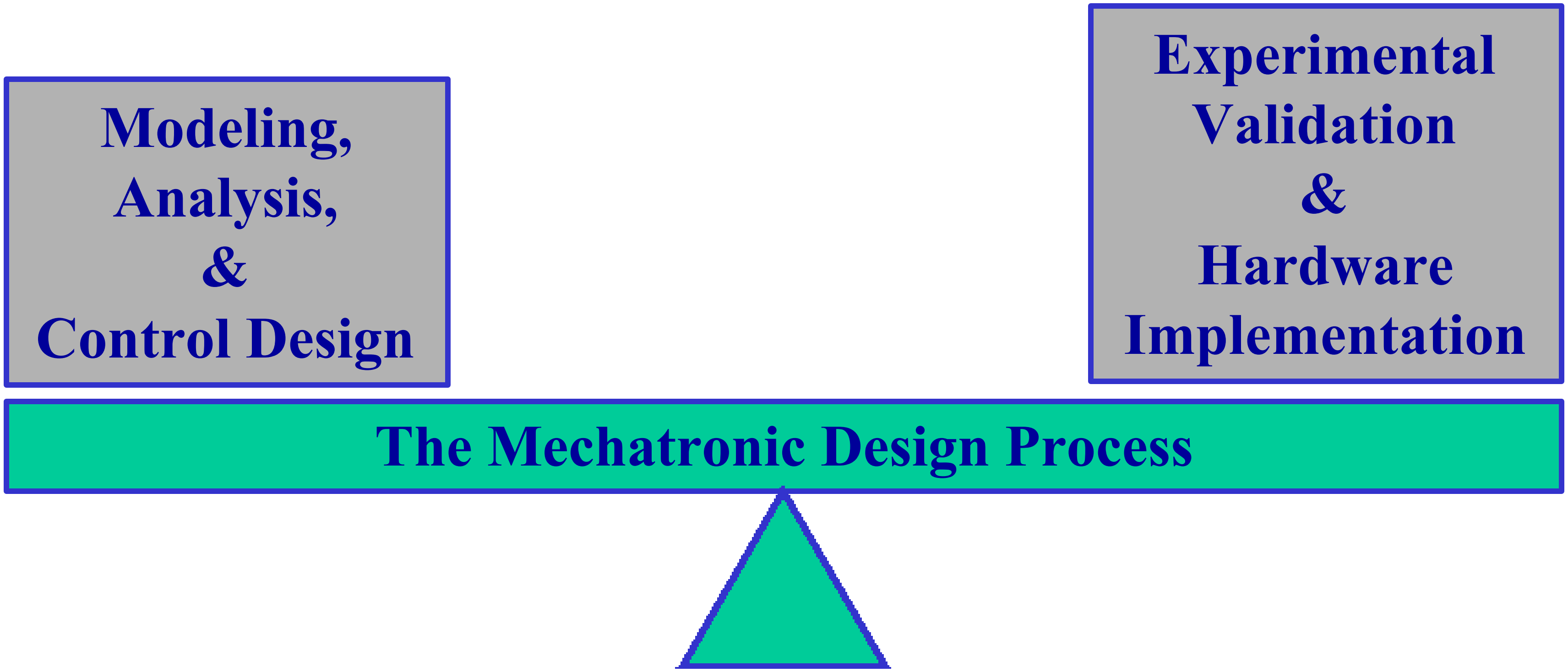
The modern engineer must draw on the synergy of

Mechatronics

Difficulties in Mechatronic Design

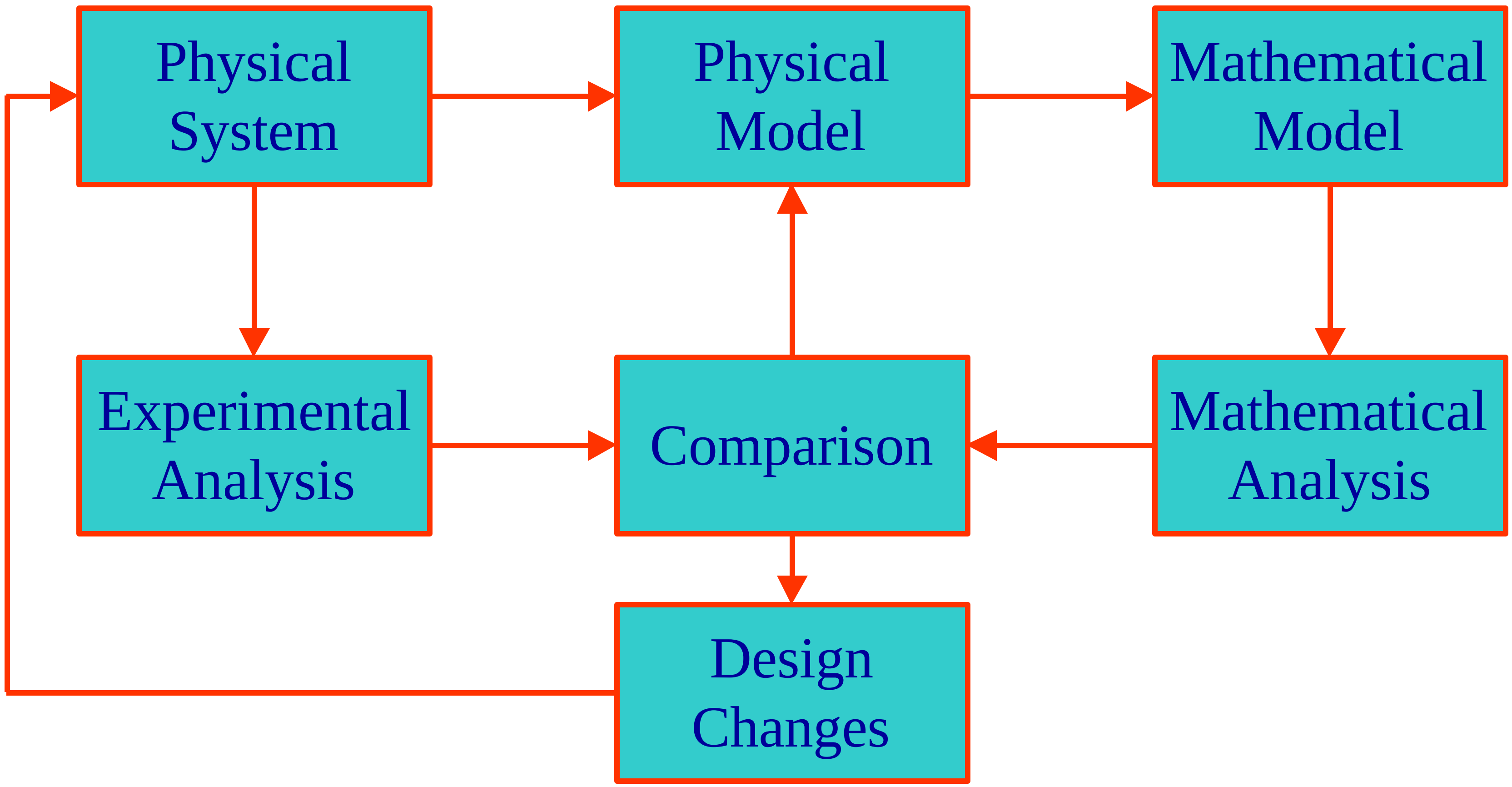
- Requires System Perspective
- System Interactions Are Important
- Requires System Modeling
- Control Systems Go Unstable

Balance: The Key To Success



Computer Simulation Without Experimental Verification Is At Best Questionable, And At Worst Useless!

Dynamic System Investigation



Realm of Mechatronics

- High Speed
- High Precision
- High Efficiency
- Highly Robust
- Micro-Miniature

Mechatronic Design Concepts

- Direct Drive Mechanisms
- Simple Mechanics
- System Complexity
- Accuracy and Speed from Controls
- Efficiency and Reliability from Electronics
- Functionality from Microcomputers

Think System !

Mechatronic Areas of Study

- Mechatronic system design principles
- Modeling, analysis, and control of dynamic physical systems
- Selection and interfacing of sensors, actuators, and microcontrollers
- Analog and digital control electronics
- Real-time programming for control
- Advanced topics, e.g.,
  - fuzzy logic control
  - smart materials as sensors and actuators
  - magnetic bearings

Challenge To Industry

- Control Design and Implementation is still the domain of the specialist.
- Controls and Electronics are still viewed as afterthought add-ons.
- Electronics and Computers are considered costly additions to mechanical designs.
- Few engineers perform any kind of modeling.
- Mathematics is a subject not viewed as enhancing one's engineering skills but as an obstacle to avoid.
- Few engineers can balance the modeling\analysis and hardware implementation essential for Mechatronics.

Course Topics

- Fundamentals of Electromechanical Motion Devices (2 periods)
- Direct-Current Motors (1 period)
- Stepper Motors (1 period)
- Brushless DC Motors (1 period)
- Fundamentals of Fluid Power: Modeling, Analysis, and Control (1 period)
- Hydraulic Actuators (1 period)
- Pneumatic Actuators (1 period)
- Sensors & Measurement Systems (1 period)
- Analog & Digital Motion Sensors (3 periods)
- Motion Control Systems (1 period)

Course Objectives

- Understand the fundamental concepts of electromechanics and fluid mechanics (hydraulics and pneumatics).
- Apply these fundamental concepts to the modeling, analysis, and control of brushed dc motors, stepper motors, brushless dc motors, solenoids, and hydraulic and pneumatic actuators.
- Understand the key elements of a measurement system.
- Understand the basic performance specifications of analog and digital sensors and actuators.
- Become familiar with the operating principles and physical/mathematical models of a variety of analog and digital motion sensors.

- Understand the key issues in hardware implementation of analog and digital actuators and sensors.
- Become proficient in the use of MatLab/Simulink to model and analyze actuators and sensors for use in mechatronic systems.
- Understand what comprises a mechatronic motion control system and the key elements in its design.

Assessment

- Homework Assignments
  - 10 assignments will be given during the course.
  - Students will have 2 weeks to complete each assignment.
  - They are to be **individually** and **professionally** done and handed in. Collaborative discussion (Not Copying!) on homework assignments is permitted and encouraged.
  - MatLab / Simulink (and some Toolboxes) will be used in the solution of some of these assignments.
  - The assignments will be graded and will count 75% (10 @ 7.5% each) of the final grade.

- Regarding late assignments, all reasonable excuses will be accepted provided they are discussed with the instructor **prior** to the due date for the assignment.

- Final Examination
  - There will be a closed-book, closed-note, final exam covering the entire course. It is worth 25% of the final grade. It will focus on the fundamental concepts studied in the course.

- Grade Summary
  - Homework Assignments (10 @ 7.5%) 75%
  - Final Exam 25%

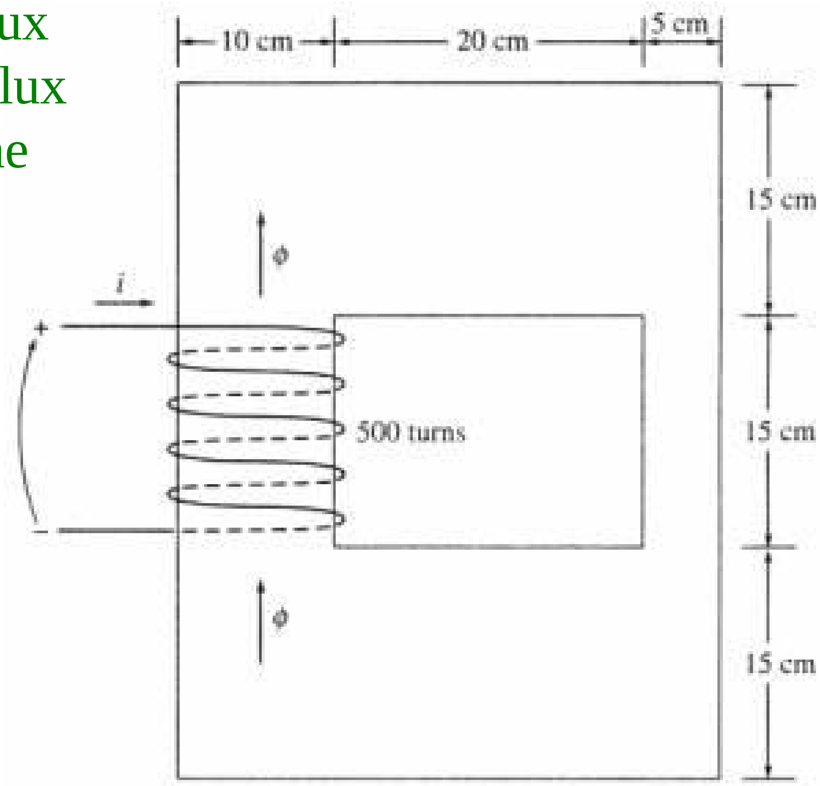
Sensors & Actuators in Mechatronics

MEAE 6960
Summer 2002

Assignment # 1

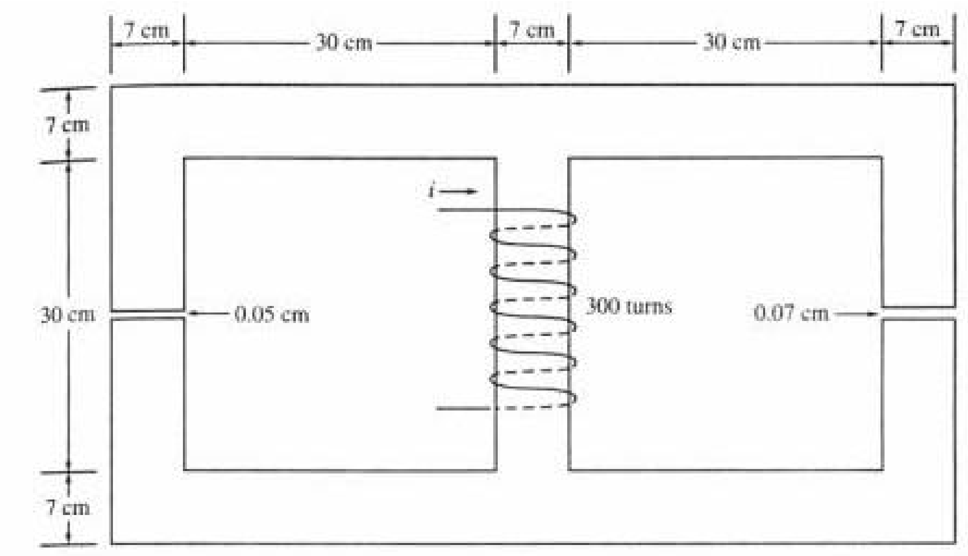
Problem # 1

A ferromagnetic core is shown. The depth of the core is 5 cm. The other dimensions of the core are as shown in the figure. Find the value of the current that will produce a flux of 0.003 Wb. With this current, what is the flux density at the top of the core? What is the flux density at the right side of the core? Assume that the relative permeability of the core is 1000.



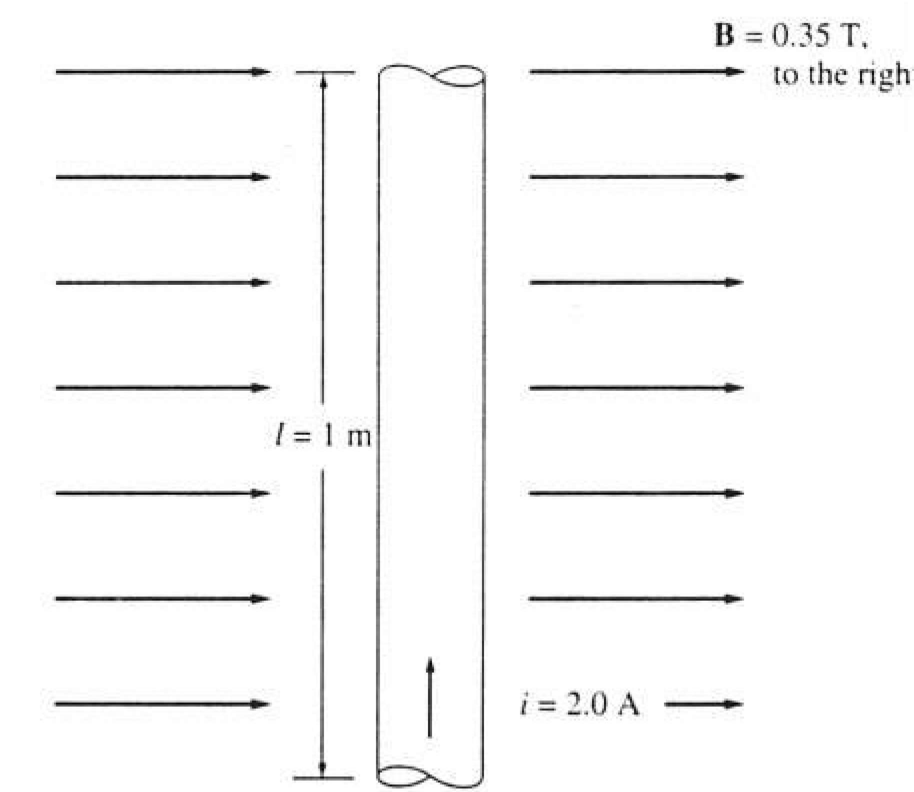
Problem # 2

A ferromagnetic core with a relative permeability of 2000 is shown. The dimensions are as shown in the figure, and the depth of the core is 7 cm. The air gaps on the left and right sides of the core are 0.050 and 0.070 cm, respectively. Because of fringing effects, the effective area of the air gaps is 5 % larger than their physical size. If there are 300 turns in the coil wrapped around the center leg of the core and if the current in the coil is 1.0 A, what is the flux in each of the left, center, and right legs of the core? What is the flux density in each air gap?



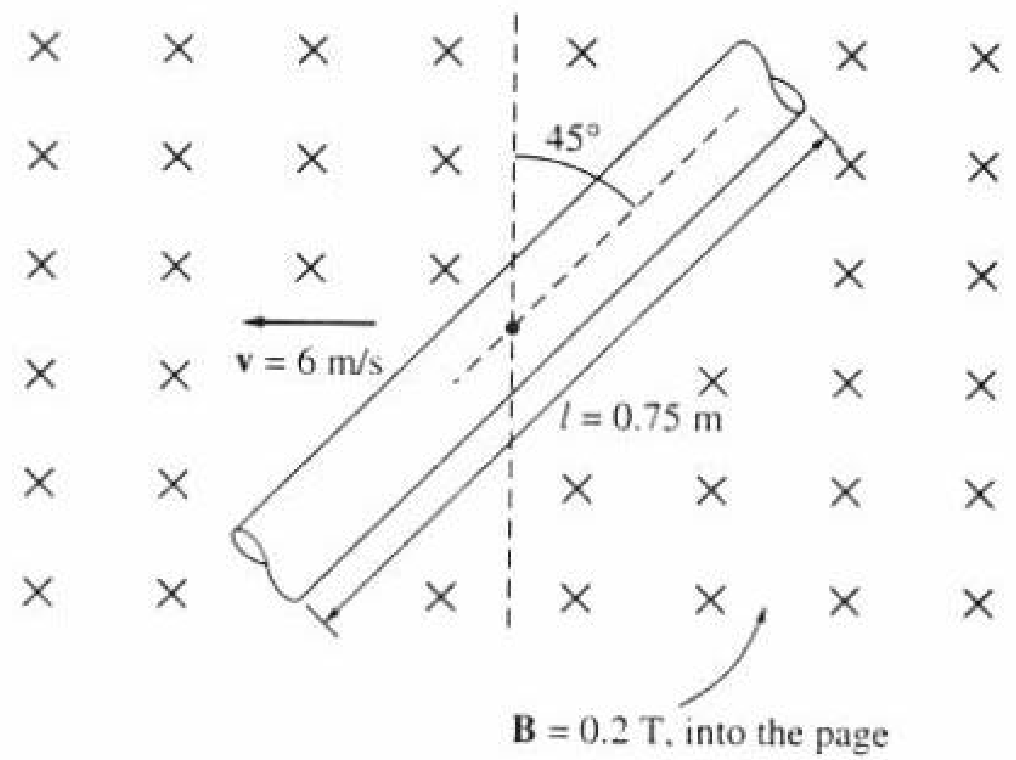
<u>Problem # 3</u>

A wire is shown in the figure which is carrying 2.0 A in the presence of a magnetic field. Calculate the magnitude and direction of the force induced on the wire.



Problem # 4

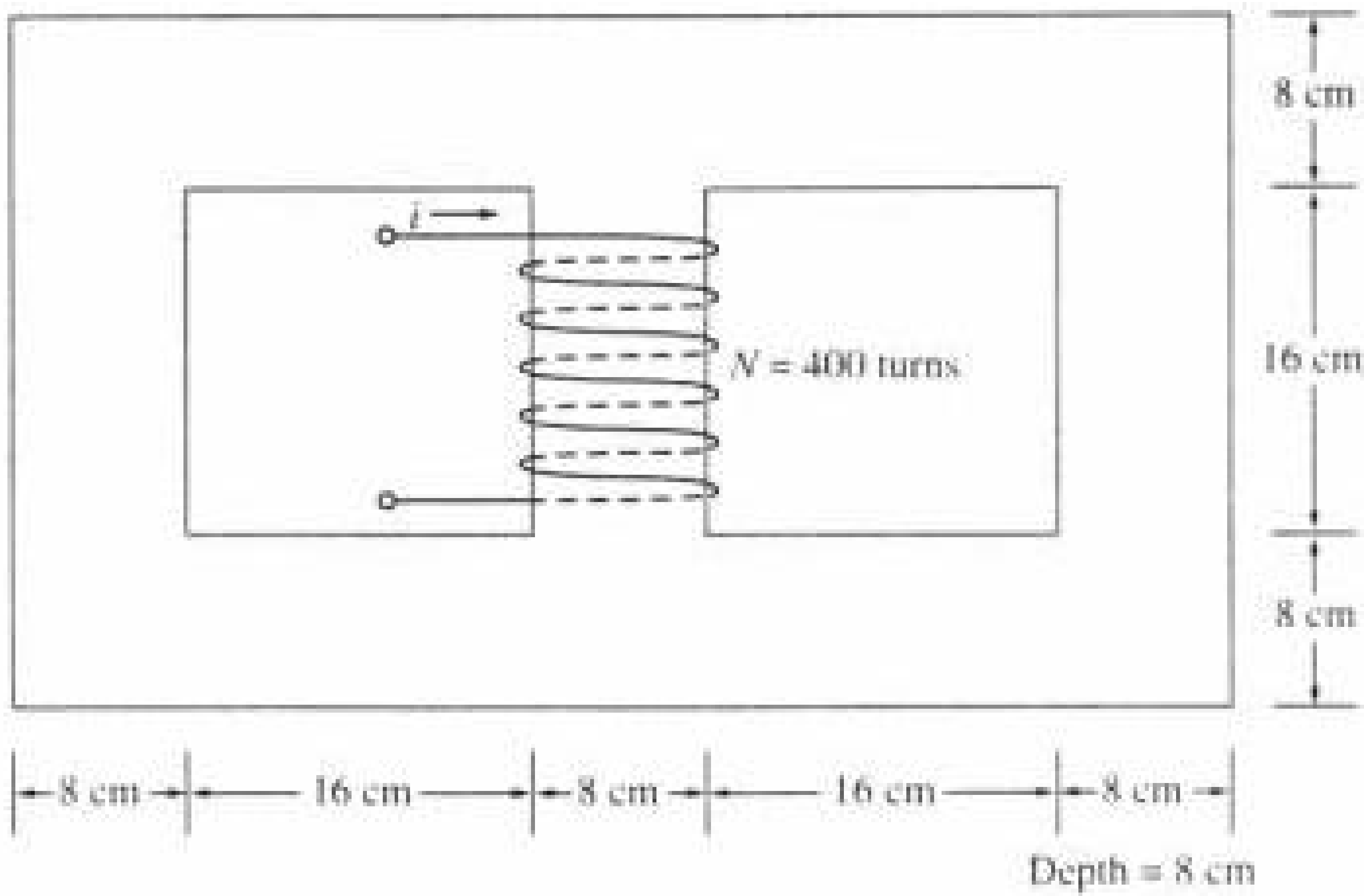
A wire is shown in the figure which is moving in the presence of a magnetic field. With the information given in the figure, determine the magnitude and direction of the induced voltage in the wire.

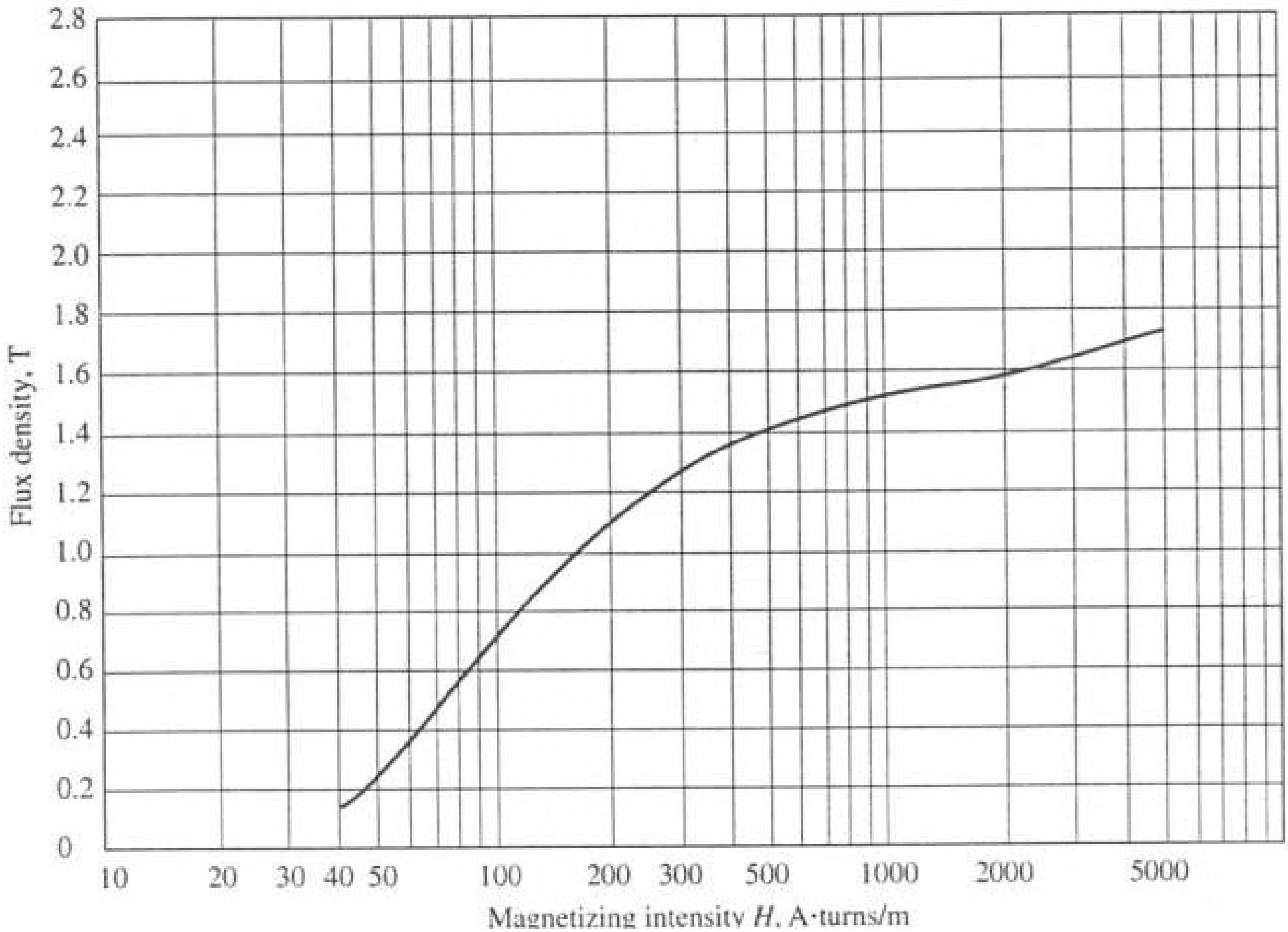


Problem # 5

A core with three legs is shown. Its depth is 8 cm, and there are 400 turns on the center leg. The remaining dimensions are shown in the figure. The core is composed of a steel having the magnetization curve shown. Answer the following questions about this core:

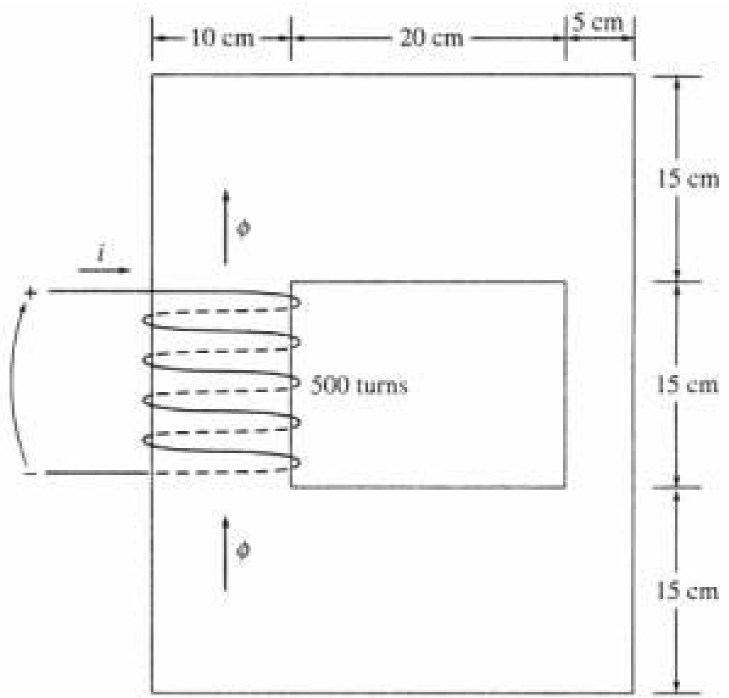
a) What current is required to produce a flux density of 0.5 T in the central leg of the core?
b) What current is required to produce a flux density of 1.0 T in the central leg of the core? Is it twice the current in part (a)?
c) What are the reluctances of the central and right legs of the core under the conditions of part (a)?
d) What are the reluctances of the central and right legs of the core under the conditions in part (b)?
e) What conclusion can you make about reluctances in real magnetic cores?

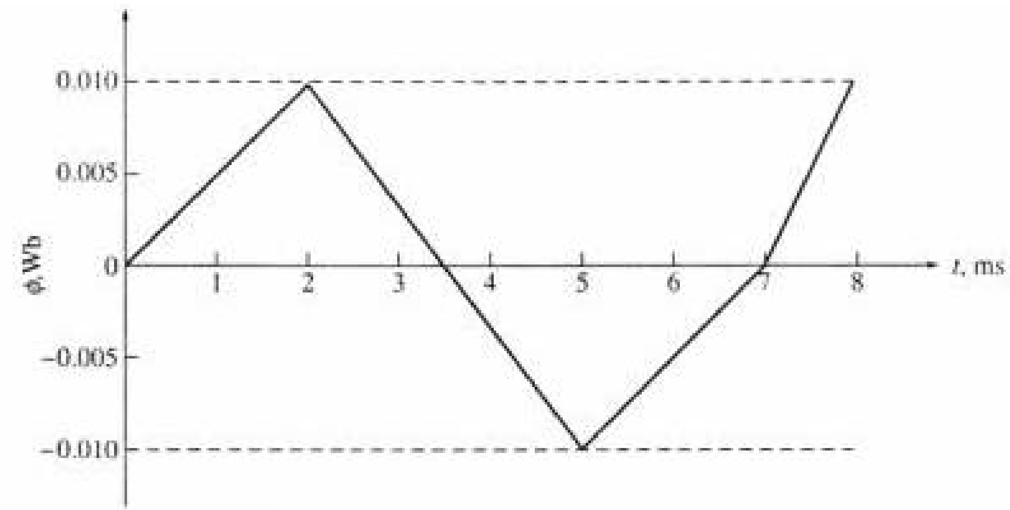

8 cm
16 cm
8 cm
i
N = 400 turns
8 cm
16 cm
8 cm
16 cm
8 cm
Depth = 8 cm


2.8
2.6
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0
Flux density, T
10
20
30
40
50
100
200
300
500
1000
2000
5000
Magnetizing intensity H, A·turns/m

Problem # 6

The core shown has the flux ϕ shown. Sketch the voltage present at the terminals of the coil.





Problem # 7

A linear machine has the following characteristics:

$$\vec{B} = 0.5 \text{ T into page} \qquad R = 0.25\ \Omega$$

$$\ell = 0.5 \text{ m} \qquad V_B = 120 \text{ V}$$

(a) If this bar has a load of 20 N attached to it opposite to the direction of motion, what is the steady-state speed of the bar?

(b) If the bar runs off into a region where the flux density falls to 0.45 T, what happens to the bar? What is its final steady-state speed?

(c) Suppose V_B is now decreased to 100 V with everything else remaining as in part (b). What is the new steady-state speed of the bar?

(d) From the results for parts (b) and (c), what are two methods of controlling the speed of a linear machine (or a real dc motor)?

Electromechanical Motion Fundamentals

- Electric Machine – device that can convert either mechanical energy to electrical energy or electrical energy to mechanical energy
  - mechanical to electrical: generator
  - electrical to mechanical: motor
  - all practical motors and generators convert energy from one form to another through the action of a magnetic field
- Transformer – device that converts *ac* electric energy at one voltage level to *ac* electric energy at another voltage level

- It operates on the same principles as generators and motors, i.e., it depends on the action of a magnetic field to accomplish the change in voltage level

- Motors, Generators, and Transformers are ubiquitous in modern daily life. Why?
  - Electric power is:
    - Clean
    - Efficient
    - Easy to transmit over long distances
    - Easy to control
    - Environmental benefits

- Purpose of this Study
  - provide basic knowledge of electromechanical motion devices for mechatronic engineers
  - focus on electromechanical rotational devices commonly used in low-power mechatronic systems
    - permanent magnet dc motor
    - brushless dc motor
    - stepper motor
- Topics Covered:
  - Magnetic and Magnetically-Coupled Circuits
  - Principles of Electromechanical Energy Conversion

References

- Electromechanical Motion Devices, P. Krause and O. Wasynczuk, McGraw Hill, 1989.
- Electromechanical Dynamics, H. Woodson and J. Melcher, Wiley, 1968.
- Electric Machinery Fundamentals, 3rd Edition, S. Chapman, McGraw Hill, 1999.
- Driving Force, The Natural Magic of Magnets, J. Livingston, Harvard University Press, 1996.
- Applied Electromagnetics, M. Plonus, McGraw Hill, 1978.
- Electromechanics and Electric Machines, S. Nasar and L. Unnewehr, Wiley, 1979.

Magnetic & Magnetically-Coupled Circuits

- Introduction
- Magnetic Field
- Magnetic Circuits
- Properties of Magnetic Materials
- Farady's Law and Lenz's Law
- Production of an Induced Force on a Wire
- Induced Voltage on a Conductor Moving in a Magnetic Field

- Linear DC Machine – A Simple Example
- Stationary Magnetically-Coupled Circuits
- Magnetic Systems with Mechanical Motion
  - Elementary Electromagnet
  - Elementary Reluctance Machine
  - Windings in Relative Motion

Introduction

- Review concepts and terms for use in the study of electromechanical motion devices.
- In all electromechanical devices, mechanical motion must occur, either translational or rotational, and this motion is reflected into the electrical system either as a change of flux linkages (electromagnetic system) or as a change of charge (electrostatic system).
- Focus is primarily on electromagnetic systems.

- If the magnetic system is linear, then the change in flux linkages results owing to a change in the inductance, i.e., inductances of electric circuits associated with electromechanical motion devices are functions of the mechanical motion.
- Learn to express self- and mutual-inductances for simple translational and rotational electromechanical devices, and to handle these changing inductances in the voltage equations describing the electrical circuits associated with the electromechanical system.

Magnetic Field

- 10 facts about The Force
  - Known for Hundreds of Years
    - If free to rotate, permanent magnets point approximately north-south.
    - Like poles repel, unlike poles attract.
    - Permanent magnets attract some things (like iron and steel), but not others (like wood and glass). Magnetic forces attract only magnetic materials.
    - Magnetic forces act at a distance, and they can act through nonmagnetic barriers.
    - Things attracted to a permanent magnet become temporary magnets themselves.

– Known only since the 19th Century

- A coil of wire with an electric current running through it becomes a magnet.
- Putting iron inside a current-carrying coil greatly increases the strength of the electromagnet.
- A changing magnetic field induces an electric current in a conductor (like copper).
- A charged particle experiences no magnetic force when moving parallel to a magnetic field, but when it is moving perpendicular to the field it experiences a force perpendicular to both the field and the direction of motion.
- A current-carrying wire in a perpendicular magnetic field experiences a force perpendicular to both the wire and the field.

- Magnetic Fields are the fundamental mechanism by which energy is converted from one form to another in motors, generators, and transformers.
- Four Basic Principles describe how magnetic fields are used in these devices:
  - A current-carrying wire produces a magnetic field in the area around it.
  - A time-changing magnetic field induces a voltage in a coil of wire if it passes through that coil (basis of transformer action).
  - A current-carrying wire in the presence of a magnetic field has a force induced on it (basis of motor action).
  - A moving wire in the presence of a magnetic field has a voltage induced in it (basis of generator action).

- In the study of electricity, one learns that stationary charges produce an electric field.
- If the charges move with uniform velocity, a secondary effect takes place: magnetism
- If we accelerate charges, there is an additional effect; the accelerated charges now produce a radiating electromagnetic field, i.e., a field that can transport energy.
- Magnetism and electromagnetic fields are special cases of electricity!

- Since motion is relative, a given physical experiment which is purely electrostatic in one coordinate system can appear as electromagnetic in another coordinate system that is moving with respect to the first. Magnetic fields seem to appear and vanish merely by a change in the motion of the observer!
- A magnetic field is thus associated with moving charges. The sources of magnetic field are currents.

$$v_q = 0 \quad \Rightarrow \quad E \neq 0, \;\; B = 0$$

$$v_q \neq 0 \quad \Rightarrow \quad E \neq 0, \;\; B \neq 0$$

$$\frac{dv_q}{dt} \neq 0 \quad \Rightarrow \quad E \neq 0, \;\; B \neq 0, \;\; \text{Radiation Fields}$$

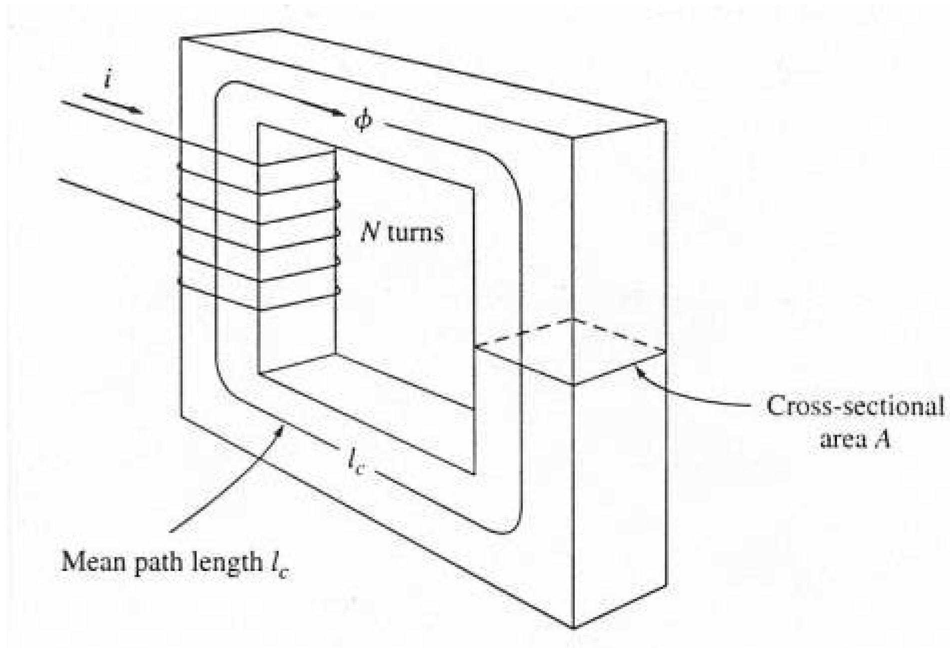
v_q = velocity of charge q

- Units of the Magnetic Field (SI and CGS)
  - Magnetic Flux Density B
    - Also called magnetic field and magnetic induction
    - 1 tesla (T) = 1 weber/meter^2 (1 Wb/m^2)
    - 1 T = 10^4 G (gauss)
    - Earth magnetic field is about 0.5 G
    - Small permanent magnet is about 100 G
    - Large electromagnet is about 20,000 G
  - Magnetic Field Intensity (or Strength) H
    - 1 ampere-turn/meter = $4\pi \times 10^{-3}$ oersted (Oe)
  - Magnetic Flux $\Phi = BA$
    - 1 weber (Wb) = 10^8 maxwell (Mx)

Magnetic Circuits

- Ampere's Law $\oint \vec{H} \bullet \vec{dL} = i_n$
  - The line integral of the magnetic field intensity (or magnetic field strength) about a closed path is equal to the net current enclosed within this closed path of integration.

Consider the elementary magnetic circuit shown.



- Rectangular ferromagnetic core with N turns of wire wrapped about one leg of the core.
- The net current passing within the path of integration is Ni.
- <u>Assumptions</u>
  - All the magnetic field produced by the current remains inside the core. Therefore the path of integration is the mean path length of the core.
  - The magnetic field intensity exists only in the direction of the given path of integration or, in other words, perpendicular to a cross section of the magnetic material (valid except in the vicinity of the corners).

- Carrying out the integration: $H\ell_c = Ni$
  - The right-hand side is referred to as ampere-turns (At) or magnetomotive force (mmf), analogous to electromotive force (emf) in electric circuits.
- The magnetic field intensity is a measure of the "effort" that a current is putting into the establishment of a magnetic field.
- The strength of the magnetic field flux produced in the core also depends on the material of the core. For linear, isotropic magnetic materials the magnetic flux density is related to the magnetic field intensity as:

$$\vec{B} = \mu \vec{H}$$

$$\left\{\begin{array}{l}\vec{H} = \text{magnetic field intensity (At/m; 1 At/m = 0.0126 Oe)} \\ \mu = \text{magnetic permeability of the material (Wb/A}\cdot\text{m or H/m)} \\ \vec{B} = \text{magnetic flux density (Wb/m}^2\text{ or T; 1 Wb/m}^2 = 10^4\text{ G)}\end{array}\right.$$

- μ , the permeability of the medium, represents the relative ease of establishing a magnetic field in a given material.
- The permeability of any other material compared to the permeability of free space or air (μ_0) is called relative permeability.

$$\mu_r = \frac{\mu}{\mu_0} \qquad \text{where } \mu_0 = 4\pi\times10^{-7}\ \text{H/m}$$

- Relative permeability is a convenient way to compare the magnetizability of materials.

- The surface integral of the flux density is equal to the total flux Φ (Wb) in a given area:

$$\Phi = \int_A \vec{B} \bullet \overrightarrow{dS}$$

- If the flux density vector is assumed perpendicular to a plane of area, and if the flux density is constant throughout the area, then: $\Phi = BA$
- In the elementary magnetic circuit: $\Phi = BA = \dfrac{\mu NiA}{\ell_c}$
- Electrical / Magnetic Circuit Analogy

$$V = iR \qquad \Im = \Phi \Re$$

$$\left\{ \begin{array}{l} V \Leftrightarrow \Im \\ i \Leftrightarrow \Phi \\ R \Leftrightarrow \Re \end{array} \right.$$

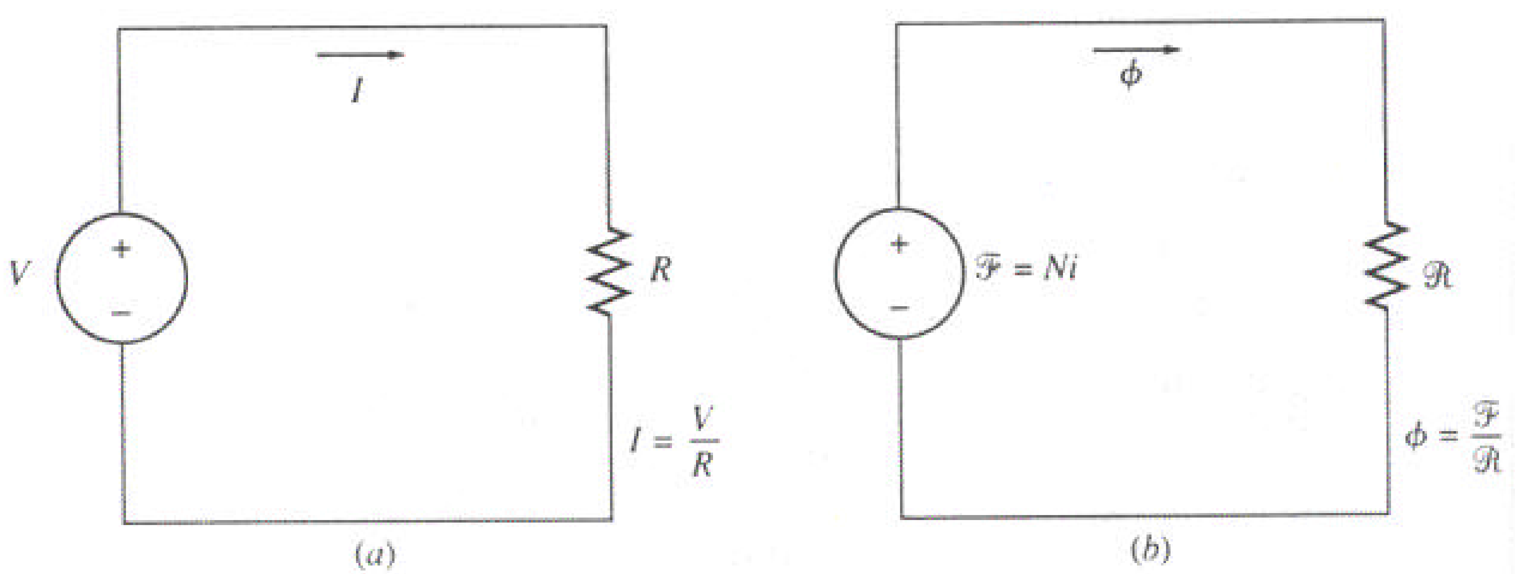
$\Im = Ni$ magnetomotive force (At)

$\Re = \dfrac{\ell_c}{\mu A}$ reluctance (At/Wb)

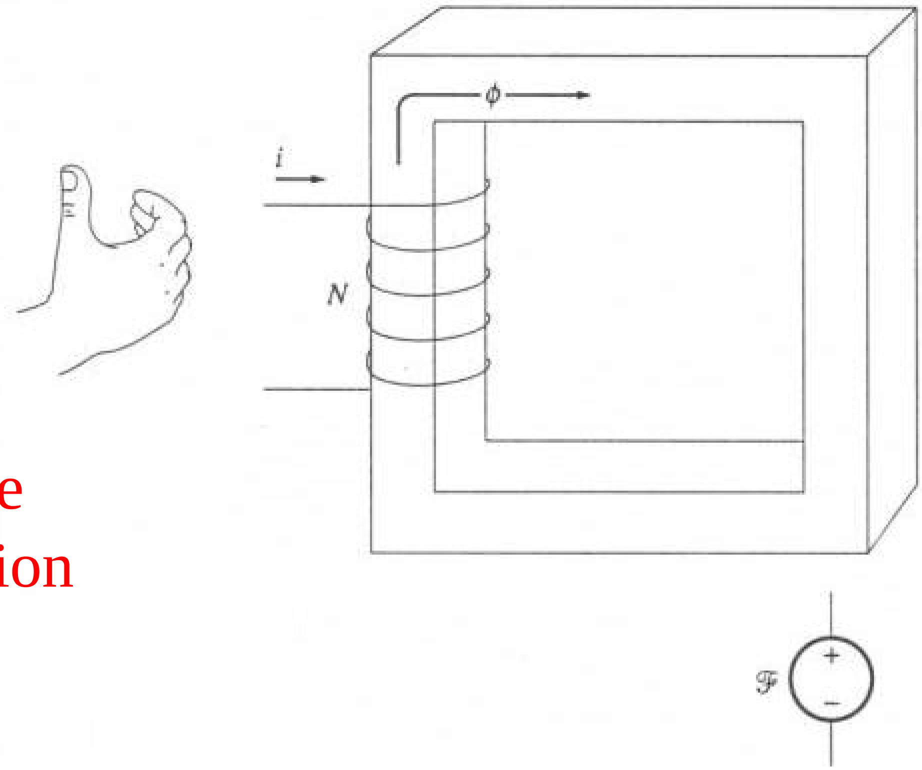
$\dfrac{1}{\Re} =$ permeance

$$R = \frac{\ell}{\sigma A}$$

- The magnetic circuit model of magnetic behavior is often used in the design of electric machines and transformers to simplify the complex design process.
- (a) Electric Circuit and (b) Magnetic Circuit



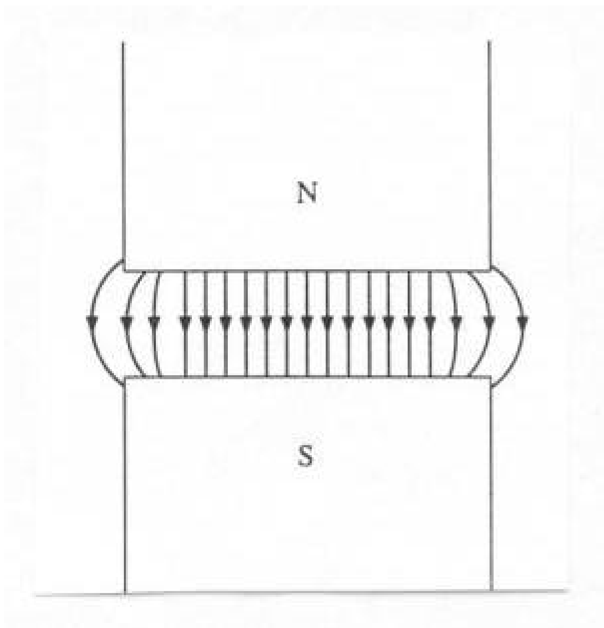
- The magnetomotive force, like voltage, has a polarity associated with it.



Modified right-hand rule for determining the direction of the positive mmf.

- Reluctances in a magnetic circuit obey the same rules for parallel and series combinations as resistances in an electric circuit.

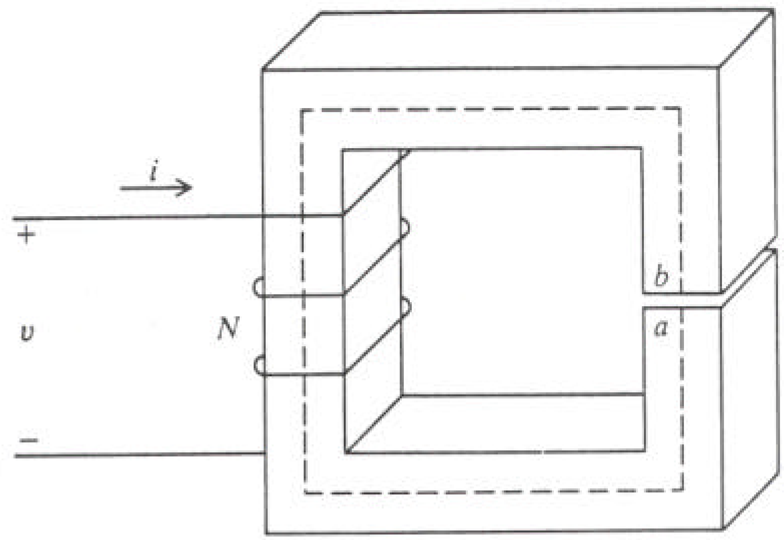
- Calculations of the flux in a core using magnetic circuit concepts are always approximations – accurate to within 5% at best! Why?
  - It is not true that all the flux is confined within the magnetic core. Flux outside the core is called *leakage flux*.
  - Calculation of reluctance assumes a certain mean path length and cross-sectional area for the core. These assumptions are not very good, especially at corners.
  - In ferromagnetic materials, the permeability varies with the amount of flux already in the material. This is a nonlinear effect. Reluctances depend on the permeability of the material.
  - "Fringing Effect" of the magnetic field at an air gap causes an increased effective cross-sectional area of the air gap.



Fringing Effect
of the magnetic field
at an air gap

- "Corrected" mean path lengths and cross-sectional areas can be used to offset these inherent sources of error.
- Magnetic circuit concept is still the easiest design tool available for calculation of fluxes.

- Consider the magnetic system shown.



- Assume that the magnetic system (circuit) consists only of the magnetic member and the air gap.
- Apply Ampere's Law to the elementary magnetic system:

$$\int_a^b H_i dL + \int_b^a H_g dL = Ni$$

- Assume that the field intensity exists only in the direction of the given path of integration or, in other words, perpendicular to a cross section of the magnetic material taken in the same sense as the air gap is cut through the material (valid except in the vicinity of the corners where the field intensity changes gradually rather than abruptly)
- Path of integration is taken as the mean length about the magnetic member ("mean length approximation")

- Carrying out the integration: $H_i \ell_i + H_g \ell_g = Ni$
- Assume that the flux density is uniformly distributed over the cross-sectional area and also perpendicular to the cross-sectional area:

$$\Phi_i = B_i A_i$$

$$\Phi_g = B_g A_g$$

- Streamlines of flux density are closed, hence the flux in the air gap is equal to the flux in the core:

$$\Delta \bullet \vec{B} = 0 \quad \text{Maxwell's } 4^{\text{th}} \text{ Equation} \qquad \Phi_i = \Phi_g \equiv \Phi$$

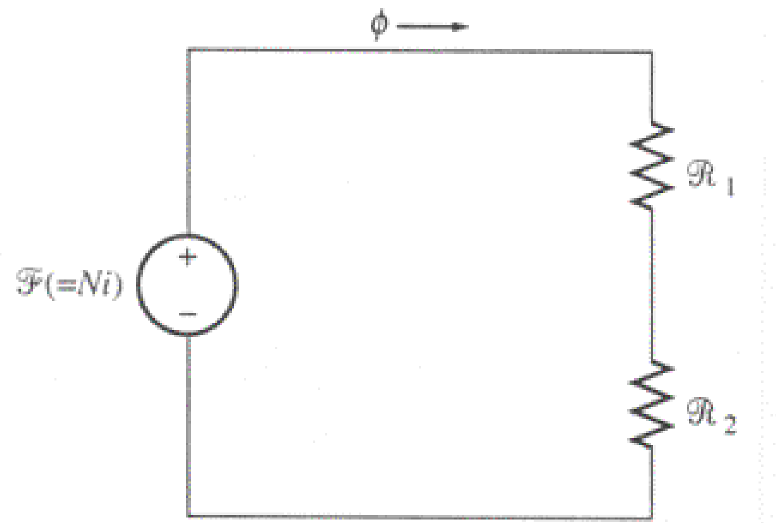
- Assume $A_g = A_i$, even though we know $A_g = kA_i$ where k > 1 due to the fringing effect.
- Results:

$$\frac{\ell_i}{\mu_i A_i}\Phi + \frac{\ell_g}{\mu_g A_g}\Phi = Ni$$

$$\mu_i = \mu_{ri}\mu_0 = (500 \rightarrow 4000)\mu_0$$

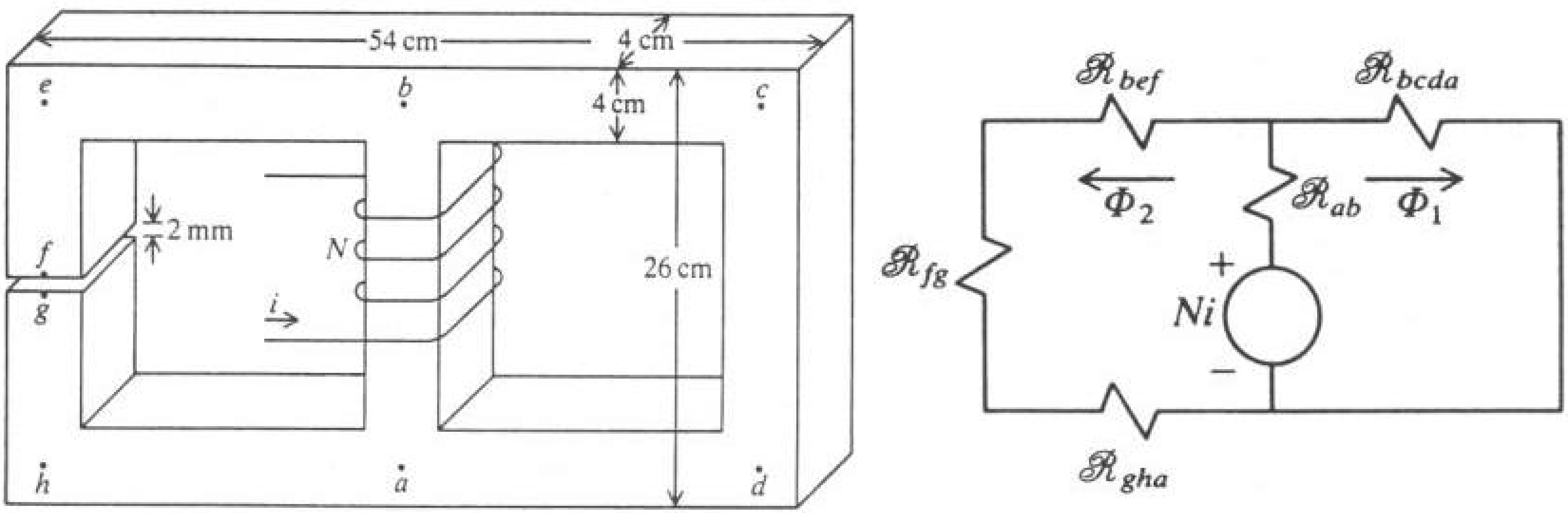
$$\mu_g = \mu_{rg}\mu_0 = (1)\mu_o$$

$$(\Re_i + \Re_g)\Phi = Ni$$



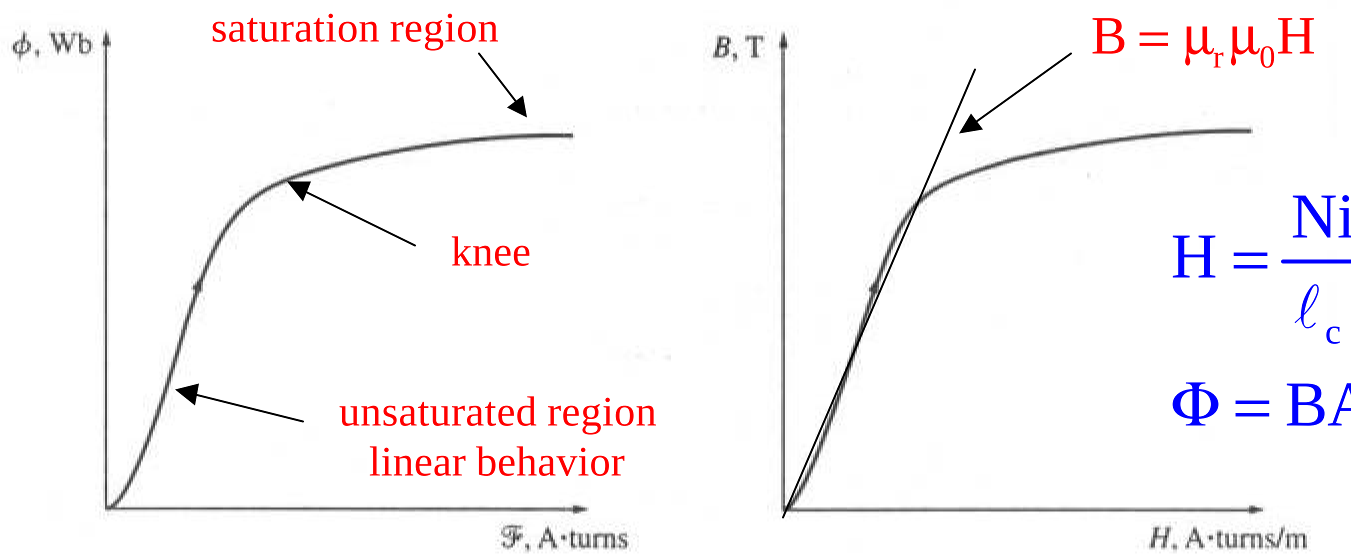
Example Problem

- In the magnetic system shown, the total number of turns is 100, the relative permeability of the iron is 1000, and the current is 10A. Calculate the total flux in the center leg.



Properties of Magnetic Materials

- The permeability of free space μ_0 is constant.
- The permeability of ferromagnetic materials (e.g., iron, nickel, cobalt) is very high (500 to 4000 times that of free space) but it is not constant. It depends on the mmf applied to the material.
- Experiment:
  - Apply a direct current to the elementary magnetic circuit previously discussed, starting with 0 A and slowly working up to the maximum permissible current. Assume that B and H are initially zero.
  - Plot flux produced in the core vs. mmf

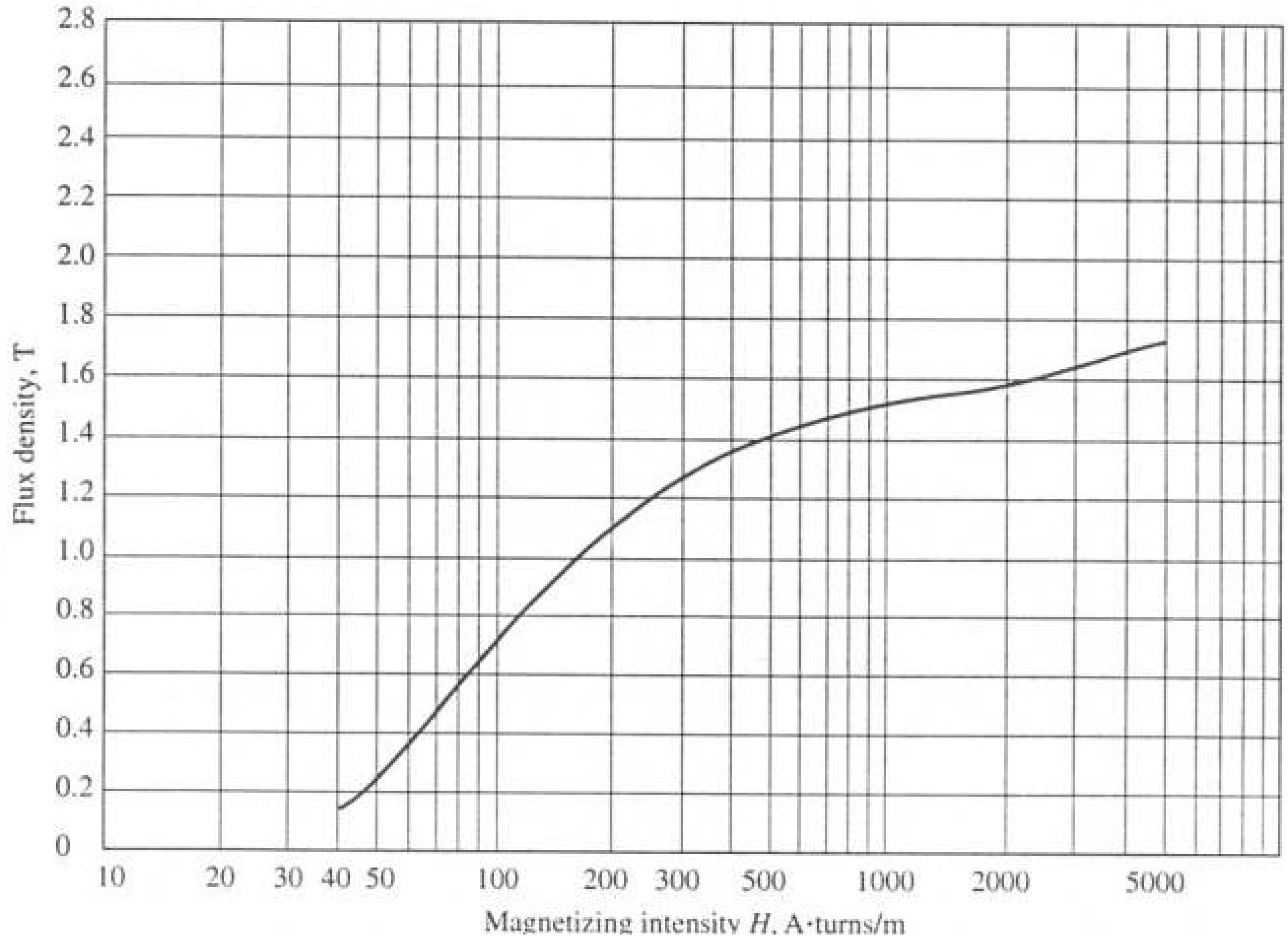


$$H = \frac{Ni}{\ell_c}$$

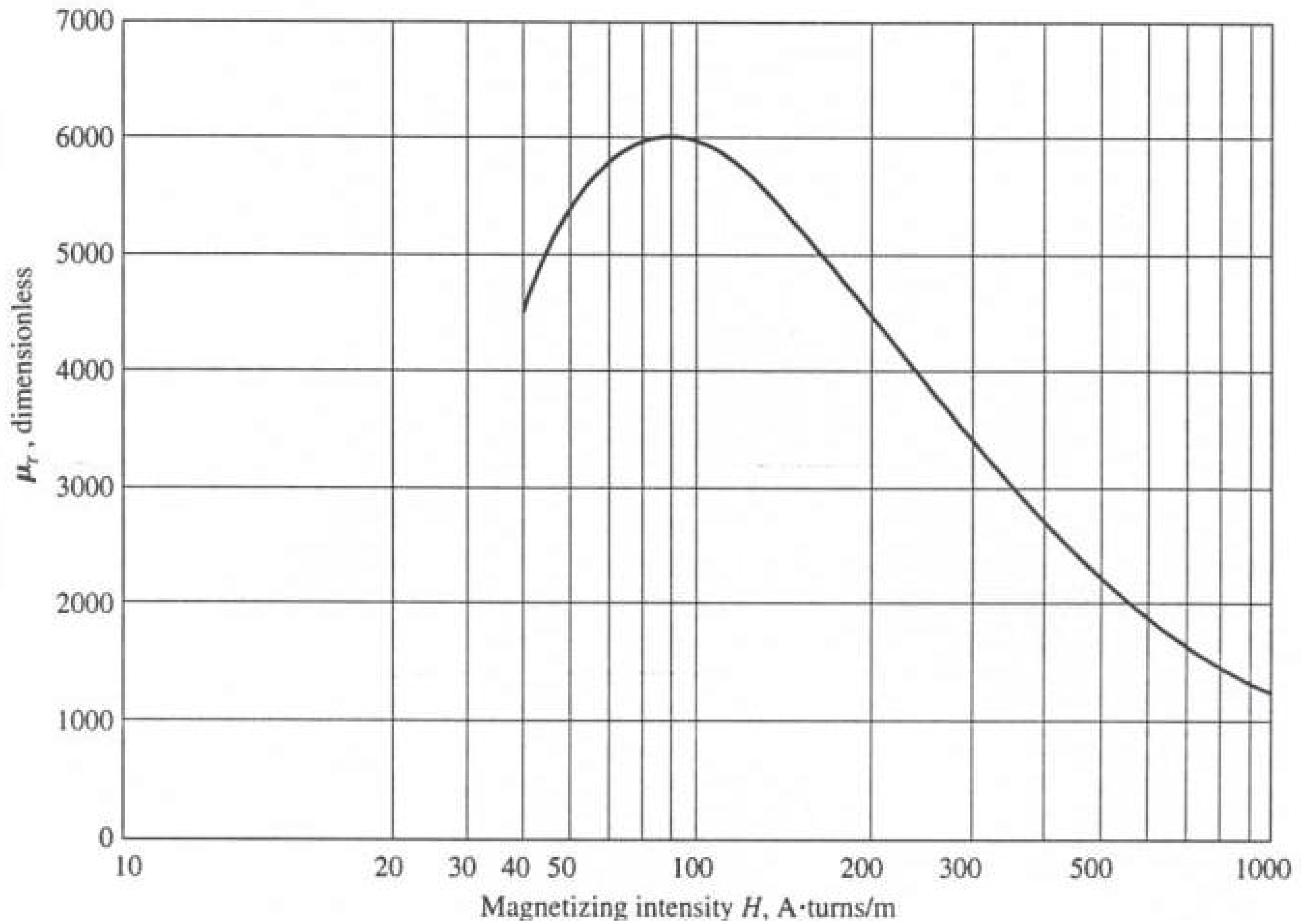
$$\Phi = BA$$

Saturation Curve or Magnetization Curve

- The relationship between B and H has the same shape as the relationship between flux and mmf.
- The slope of the B vs. H curve at any value of H is, by definition, the permeability of the core at that H.



Magnetization Curve for a Typical Piece of Steel



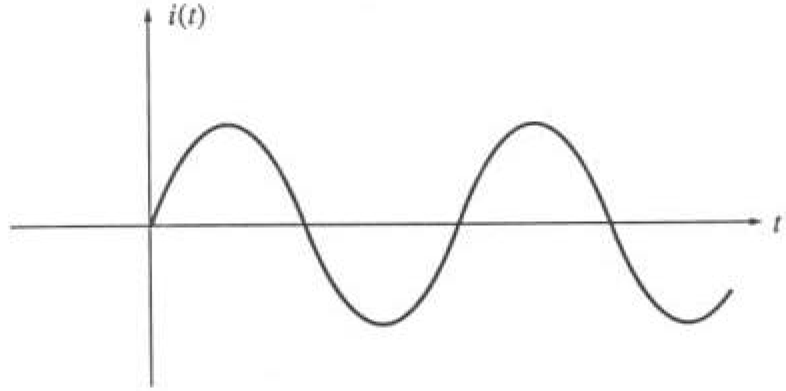
Relative Permeability vs. H Curve for a Typical Piece of Steel

- Most real machines operate near the knee of the magnetization curve; the flux in their cores is not linearly related to the mmf producing it.
- Why does the magnetization curve have this shape?
  - Microscopically, ferromagnetic materials have been found to be divided into magnetic domains wherein all magnetic moments (dipoles) are aligned. Each domain acts as a small permanent magnet. The direction of this alignment will differ from one domain to another; domains are oriented randomly within the material.
  - When a ferromagnetic material is subjected to an external field, it causes domains that happen to point in the direction of the field to grow at the expense of domains pointed in other directions. It is a positive feedback effect!

- This is known as domain wall motion. As the strength of the magnetic field increases, the aligned domains continue to grow in a nearly linear fashion. Whole domains that are aligned in the wrong direction eventually reorient themselves as a unit to line up with the field. A nearly linear B-H curve results.
- Soon the ability of the aligned domains to take from the unaligned domains starts to slow. This gives rise to the knee of the B-H curve and saturation is beginning.
- Finally, when nearly all the atoms and domains in the iron are lined up with the external field, any further increase in the mmf can cause only the same flux increase that it would in free space. Once everything is aligned, there can be no more feedback effect to strengthen the field. The material is saturated with flux. Slope of B-H curve is μ_0 .

New Experiment

Instead of applying a direct current to the windings on the core, apply an alternating current and observe what happens. Assume that B and H are initially both zero.



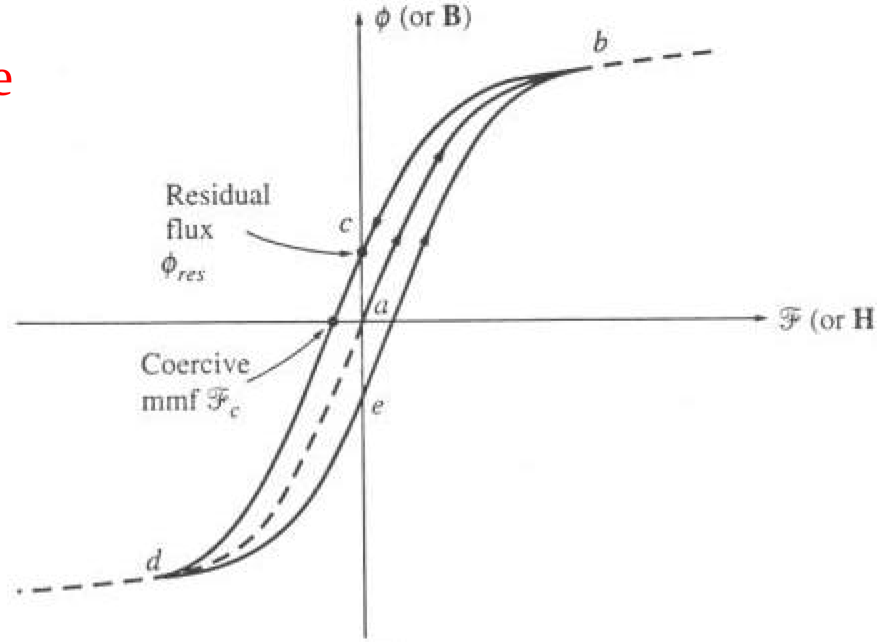
(a)

After several cycles, a steady-state condition is reached.

Hysteresis Loop

Path b-c-d-e-b

(double-valued function)



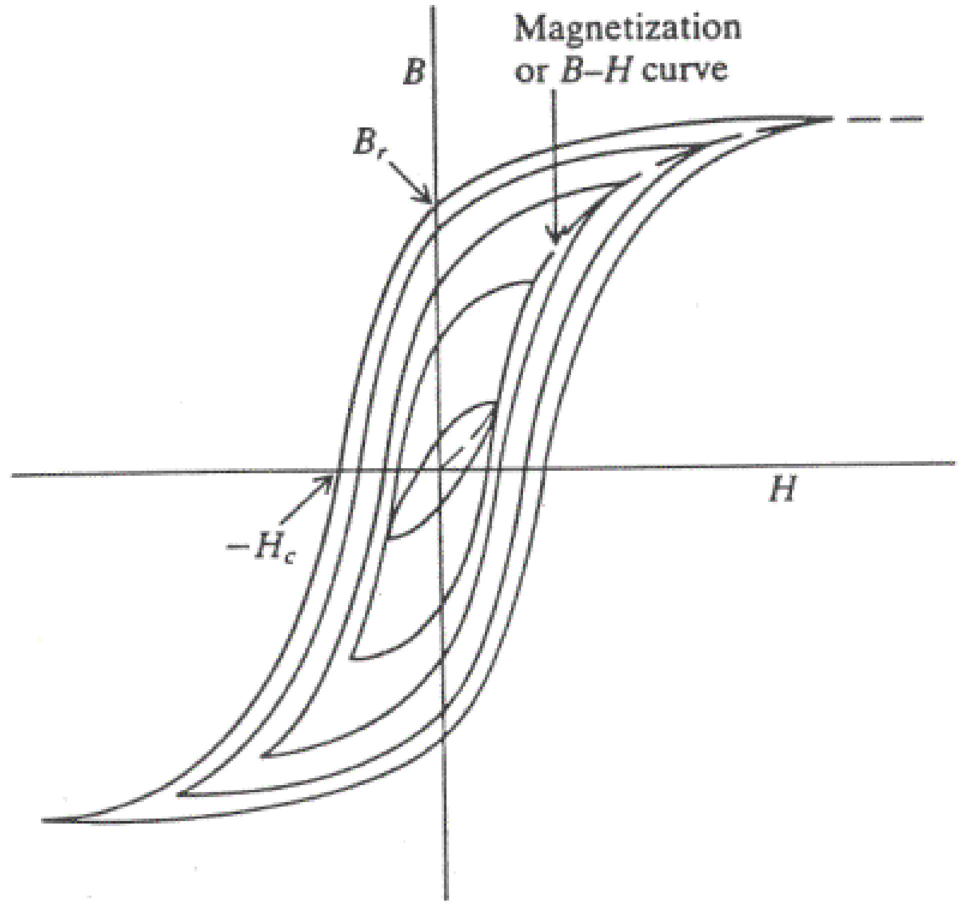
(b)

- Observations
  - The amount of flux present in the core depends not only on the amount of current applied to the windings of the core, but also on the previous history of the flux in the core.
  - The dependence on the preceding flux history and the resulting failure to retrace flux paths is called *hysteresis*.
  - Path *bcdeb* traced out on the magnetization curve as the applied current changes is called a *hysteresis loop*.
  - If a large mmf is first applied to the core and then removed, the flux path in the core will be *abc*. When the mmf is removed, the flux in the core does not go to zero. Instead, a magnetic field is left in the core. This magnetic field is called the *residual flux*. It is in precisely this manner that permanent magnets are produced.

- To force the flux to zero, an amount of mmf known as the *coercive magnetomotive force* must be applied to the core in the opposite direction.

- Why does hysteresis occur?
  - In simple terms, the growth of aligned domains for an incremental change in H in one direction is not equal to the growth of oppositely aligned domains if this change in H were suddenly reversed.

The diagram shows the effect of the size of mmf excursions on the magnitude of the hysteresis loop.



Family of Steady-State Hysteresis Loops

- Hysteresis Loss
  - The fact that turning domains in a ferromagnetic material requires energy leads to a common type of energy loss in all machines and transformers.
  - The hysteresis loss in an iron core is the energy required to accomplish the reorientation of domains during each cycle of the alternating current applied to the core. The area enclosed in the hysteresis loop formed by applying an alternating current to the core is directly proportional to the energy lost in a given *ac* cycle.
  - The smaller the applied mmf excursions on the core, the smaller the area of the resulting hysteresis loop and so the smaller the resulting losses.

– This energy causes a rise in the temperature of the magnetic material and the power associated with this energy loss is called *hysteresis loss*.

- Eddy Currents
  – The mechanism of eddy current losses is explained by Faraday's Law. A time-changing flux induces voltage within a ferromagnetic core.
  – When a solid block of magnetic material is subjected to an alternating field intensity, the resulting alternating flux induces current in the solid magnetic material which will circulate in a loop perpendicular to the flux density inducing it. These are called *eddy currents*.
  – There are two undesirable side effects from eddy currents:

- First, the mmf established by these circulating currents opposes the mmf produced by the winding, and this opposition is greatest at the center of the material because that tends to be also the center of the current loops. Thus, the flux would tend not to flow through the center of the solid magnetic member, thereby not utilizing the full benefits of the ferromagnetic material.
- Second, there is a I^2R loss associated with these eddy currents flowing in a resistive material, called eddy current loss, which is dissipated as heat.

– These two adverse effects can be minimized in several ways; the most common way is to build the ferromagnetic core of laminations insulated from each other and oriented in the direction of the magnetic field. These thin strips offer a much smaller area in which the eddy currents can flow, hence smaller currents and smaller losses result.

- Core Losses
  - The core losses associated with ferromagnetic materials are the combination of the hysteresis and eddy current losses.
  - Electromagnetic devises are designed to minimize these losses; however, they are always present.
  - We can often take them into account in a linear system analysis by representing their effects on the system by a resistance.

Faraday's Law and Lenz's Law

- Now focus on the various ways in which an existing magnetic field can effect its surroundings.
- Faraday's Law — Maxwell's 1st Equation $\Delta \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}$
  - If a flux passes through a turn of a coil of wire, a voltage will be induced in the turn of wire that is directly proportional to the rate of change in the flux with respect to time.

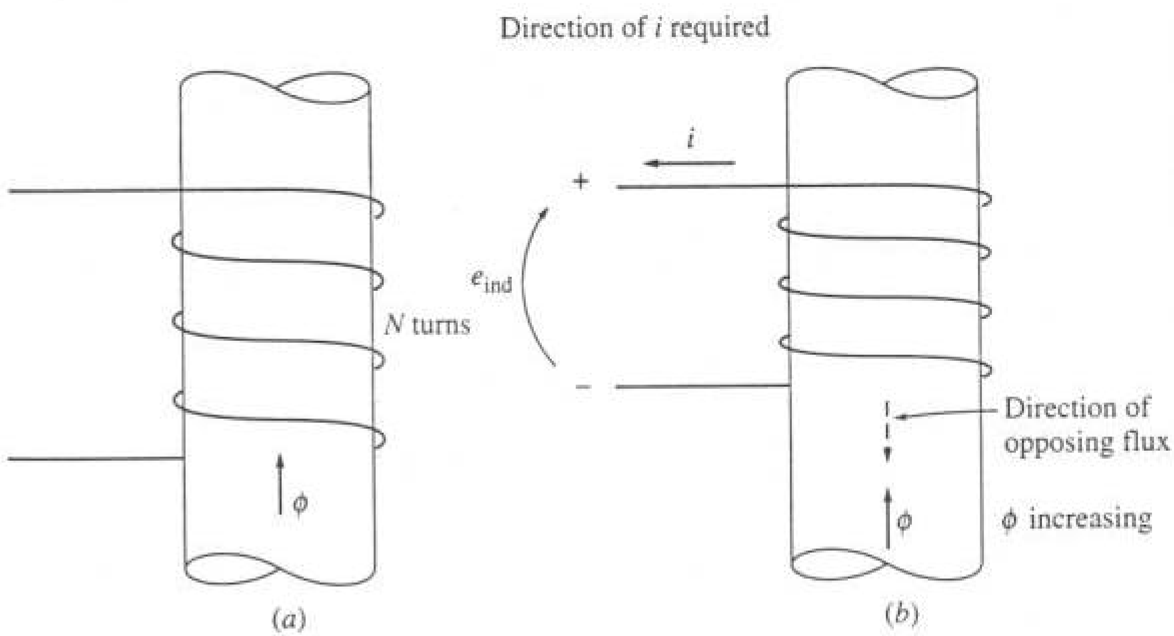
$$e_{ind} = -\frac{d\phi}{dt}$$

  - If the coil has N turns and the same flux passes through all of them, then

$$e_{ind} = -N\frac{d\phi}{dt}$$

- It is the fundamental law of transformer operation.
- The minus sign is an expression of Lenz's Law.
  - The direction of the voltage buildup in the coil is such that if the coil ends were short-circuited, it would produce current that would cause a flux opposing the original flux change. Since the induced voltage opposes the change that causes it, a minus sign is included.
  - The minus sign is often left out, as the polarity of the resulting voltage can be determined from physical considerations.
- Practical Problem
  - Equation assumes that exactly the same flux is present in each turn of the coil. What about leakage flux?

Meaning of Lenz's Law



(a) A coil enclosing an increasing magnetic flux
(b) Determining the resulting voltage polarity

– If the windings are tightly coupled, so that the vast majority of the flux passing through one turn of the coil does indeed pass through all of them, then the equation will give valid answers.
– If the leakage is quite high or if extreme accuracy is required, a different expression is needed.

$$e_{ind} = \frac{d(\phi_i)}{dt} \quad i^{th} \text{ turn of the coil}$$

$$e_{ind} = \sum_{i=1}^{N} e_i = \sum_{i=1}^{N} \frac{d(\phi_i)}{dt} = \frac{d}{dt}\left(\sum_{i=1}^{N} \phi_i\right)$$

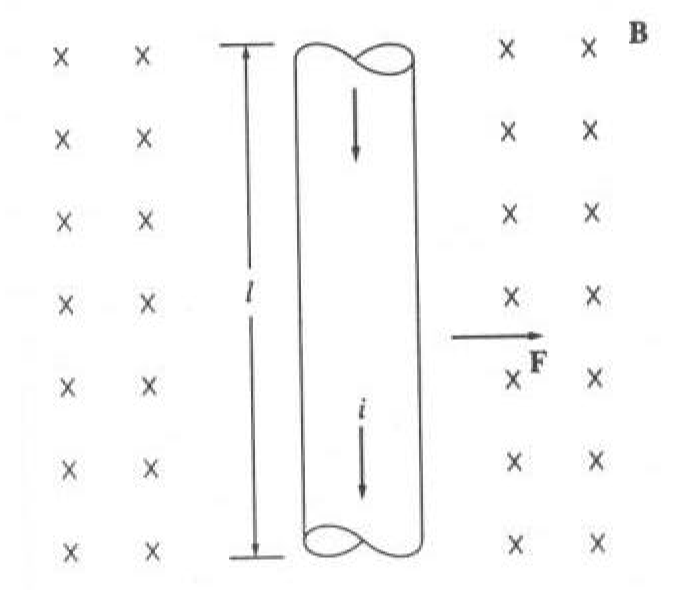
$$e_{ind} = \frac{d\lambda}{dt} \quad \text{where } \lambda = \sum_{i=1}^{N} \phi_i = \text{ flux linkage of coil (Wb-turns)}$$

Production of an Induced Force on a Wire

- A magnetic field induces a force on a current-carrying wire within the field. $\vec{F} = i\left(\vec{\ell} \times \vec{B}\right)$
- The direction of the force is given by the right-hand rule.
- The magnitude of the force is given by $F = i\ell B \sin\theta$ where θ is the angle between the wire and the flux density vector.
- The induction of a force in a wire by a current in the presence of a magnetic field is the basis of motor action.

$\vec{B}$
points into the page

A current-carrying wire in the presence of a magnetic field



$$\vec{F} = i\left(\vec{\ell} \times \vec{B}\right)$$

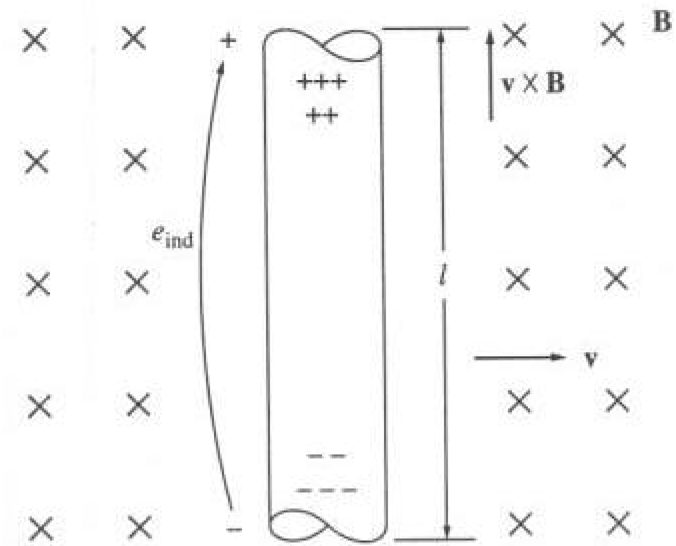
Induced Voltage on a Conductor Moving in a Magnetic Field

- If a wire with the proper orientation moves through a magnetic field, a voltage is induced in it. The voltage induced in the wire is given by $e_{ind} = \left(\vec{v} \times \vec{B}\right) \bullet \vec{\ell}$
- Vector $\vec{\ell}$ points along the direction of the wire toward the end making the smallest angle with respect to the vector $\vec{v} \times \vec{B}$.
- The voltage in the wire will be built up so that the positive end is in the direction of the vector $\vec{v} \times \vec{B}$.
- The induction of voltages in a wire moving in a magnetic field is the basis of generator action.

$\vec{B}$

points into the page

A conductor moving in the presence of a magnetic field



$$e_{ind} = (\vec{v} \times B) \bullet \vec{\ell}$$

Linear DC Machine – A Simple Example

- A linear dc machine is about the simplest and easiest-to-understand version of a dc machine, yet it operates according to the same principles and exhibits the same behavior as real generators and motors.

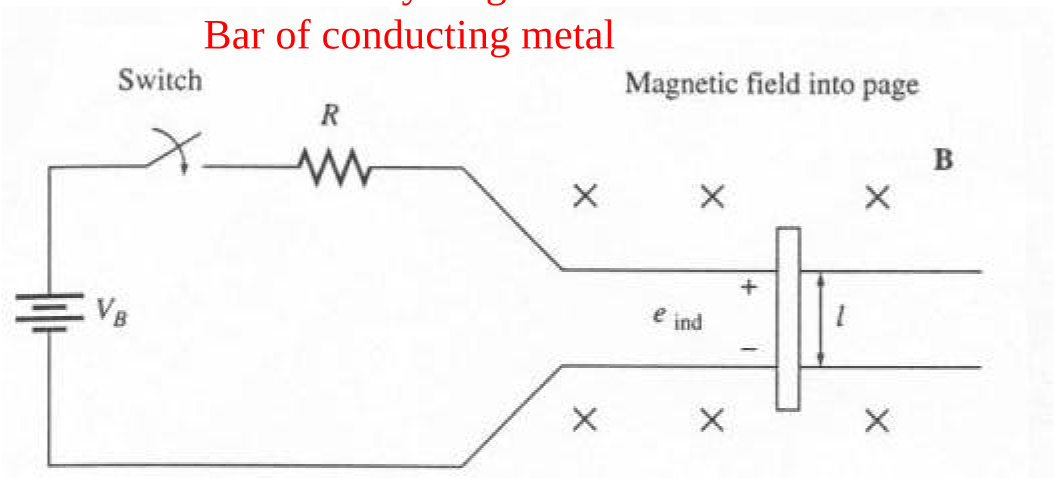
$$\vec{F} = i\left(\vec{\ell} \times \vec{B}\right)$$

$$e_{ind} = \left(\vec{v} \times \vec{B}\right) \bullet \vec{\ell}$$

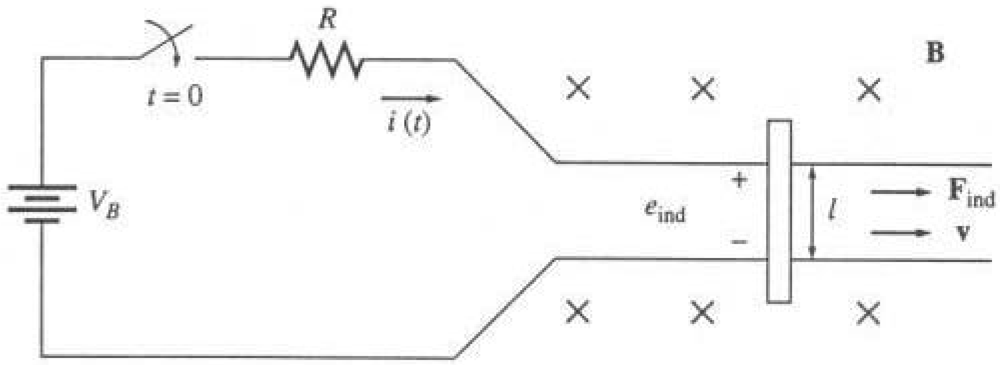
$$V_B - iR - e_{ind} = 0$$

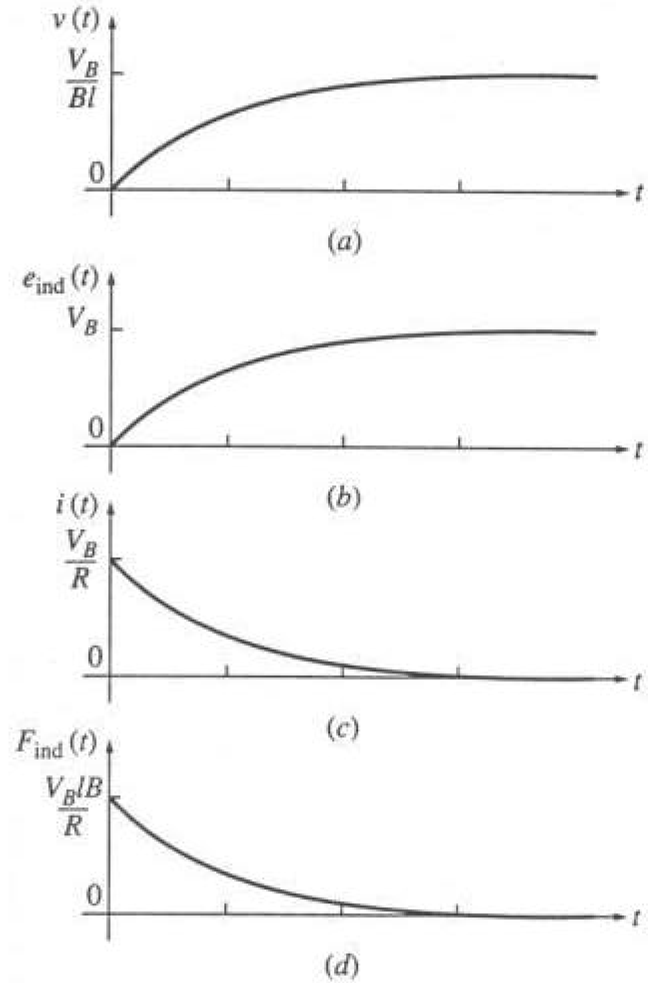
$$F_{net} = ma$$

Smooth frictionless rails
Uniform-density magnetic field
Bar of conducting metal



Starting the Linear DC Machine





Closing the switch produces a current flow $i = \frac{V_B}{R}$

The current flow produces a force on the bar given by $F = i\ell B$

The bar accelerates to the right, producing an induced voltage e_{ind} as it speeds up.

This induced voltage reduces the current flow $i = \frac{\left(V_B - e_{ind}\uparrow\right)}{R}$

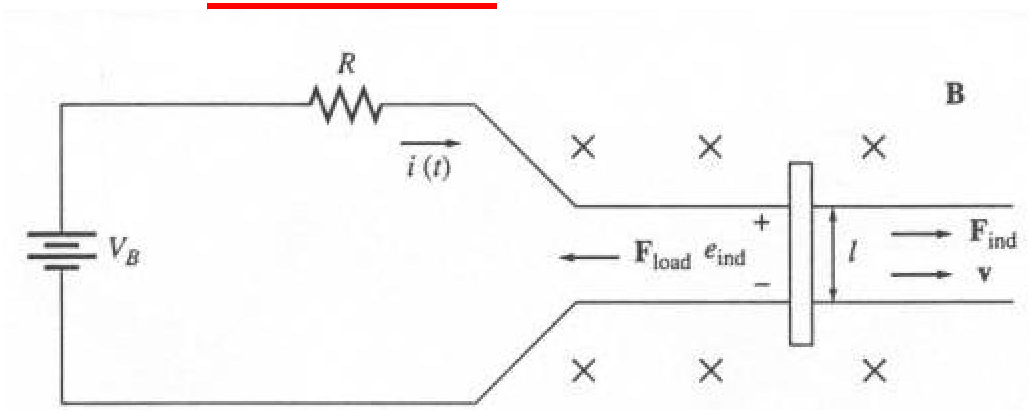
The induced force is thus decreased until eventually F = 0. At that point, $e_{ind} = V_B$ and i = 0, and the bar moves at a constant no-load speed.

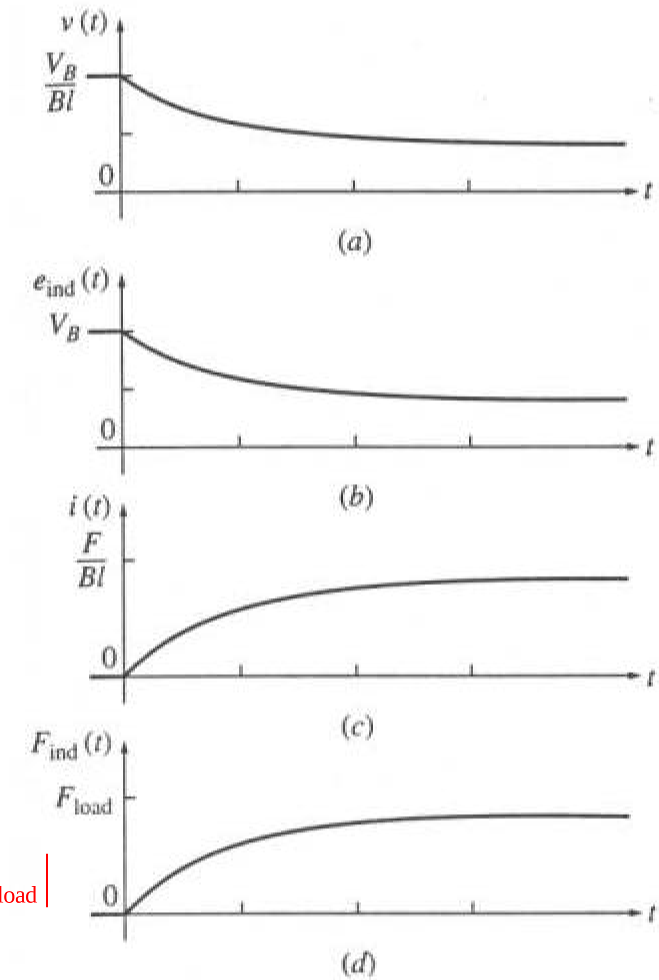
$F = i\downarrow \ell B$

$v_{ss} = \frac{V_B}{B\ell}$

The Linear DC Machine as a Motor

Apply an external load
Assume machine is initially running at no-load SS conditions





A force F_{load} is applied opposite to the direction of motion, which causes a net force F_{net} opposite to the direction of motion.

The resulting acceleration is negative, so the bar slows down. $a = \frac{F_{net}}{m}$

$$e_{ind} = v \downarrow \ell B$$

The voltage e_{ind} falls, and so i increases.

$$i = \frac{(V_B - e_{ind} \downarrow)}{R}$$

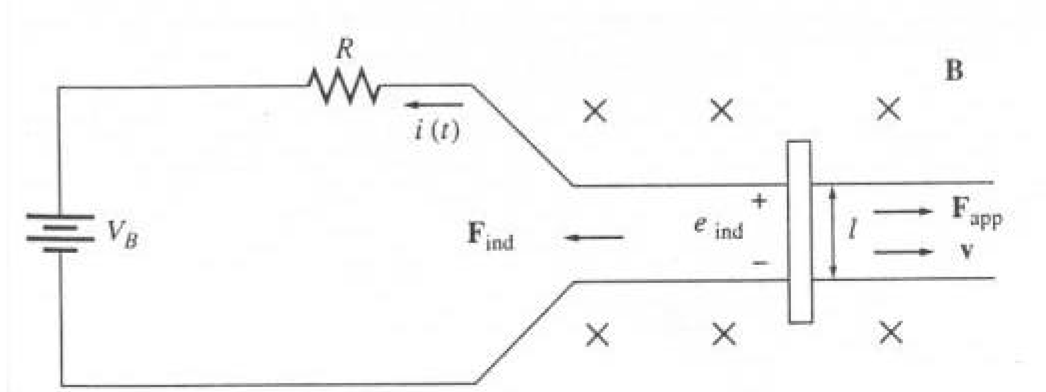
The induced force F_{ind} increases until, at a lower speed, $|F_{ind}| = |F_{load}|$

$$F_{ind} = i \uparrow \ell B$$

An amount of electric power equal to $e_{ind}i$ is now being converted to mechanical power equal to $F_{ind}v$.

The Linear DC Machine as a Generator

Apply a force in the direction of motion
Assume machine is initially running at no-load SS conditions



A force F_{app} is applied in the direction of motion; F_{net} is in the direction of motion.
Acceleration is positive, so the bar speeds up. $a = \frac{F_{net}}{m}$

$$e_{ind} = v \uparrow B\ell$$

The voltage e_{ind} increases, and so i increases.

$$i = \frac{\left(e_{ind} \uparrow - V_B\right)}{R}$$

The induced force F_{ind} increases until, at a higher speed, $\left|F_{ind}\right| = \left|F_{app}\right|$

$$F_{ind} = i \uparrow \ell B$$

An amount of mechanical power equal to $F_{ind}v$ is now being converted to electric power $e_{ind}i$, and the machine is acting a a generator.

- Observations
  - The same machine acts as both motor and generator.
    - Generator: externally applied forces are in the direction of motion
    - Motor: externally applied forces are opposite to the direction of motion
  - Electrically
    - $e_{ind} > V_B$, machine acts as a generator
    - $e_{ind} < V_B$, machine acts as a motor
  - Whether the machine is a motor or a generator, both induced force (motor action) and induced voltage (generator action) are present at all times.
  - This machine was a generator when it moved rapidly and a motor when it moved more slowly, but whether it was a motor or a generator, it always moved in the same direction.

Example Problem

- A linear dc machine has a battery voltage of 120 V, an internal resistance of 0.3 Ω, and a magnetic flux density of 0.1 T.
  - What is the machine's maximum starting current? What is its steady-state velocity at no load?
  - Suppose that a 30-N force pointing to the right were applied to the bar. What would the steady-state speed be? How much power would the bar be producing or consuming? How much power would the battery be producing or consuming? Explain the difference between these two figures. Is this machine acting as a motor or as a generator?

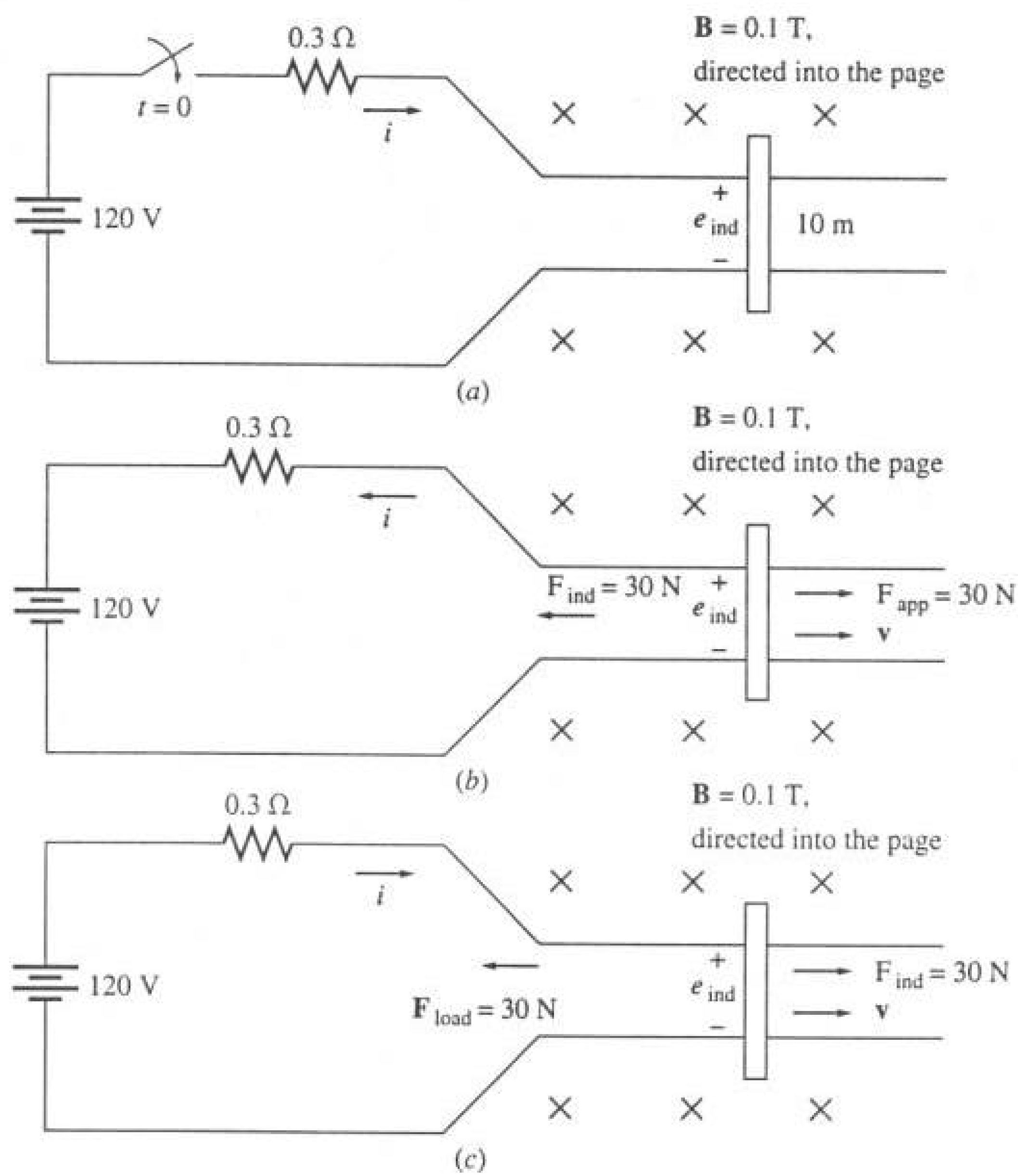
- Now suppose a 30-N force pointing to the left were applied to the bar. What would the new steady-state speed be? Is this machine a motor or a generator now?
- Assume that a force pointing to the left is applied to the bar. Calculate the speed of the bar as a function of the force for values from 0 N to 50 N in 10-N steps. Plot the velocity of the bar versus the applied force.
- Assume that the bar is unloaded and that it suddenly runs into a region where the magnetic field is weakened to 0.08 T. How fast will the bar go now?

Example Problem

Starting Conditions

Operating as a Generator

Operating as a Motor



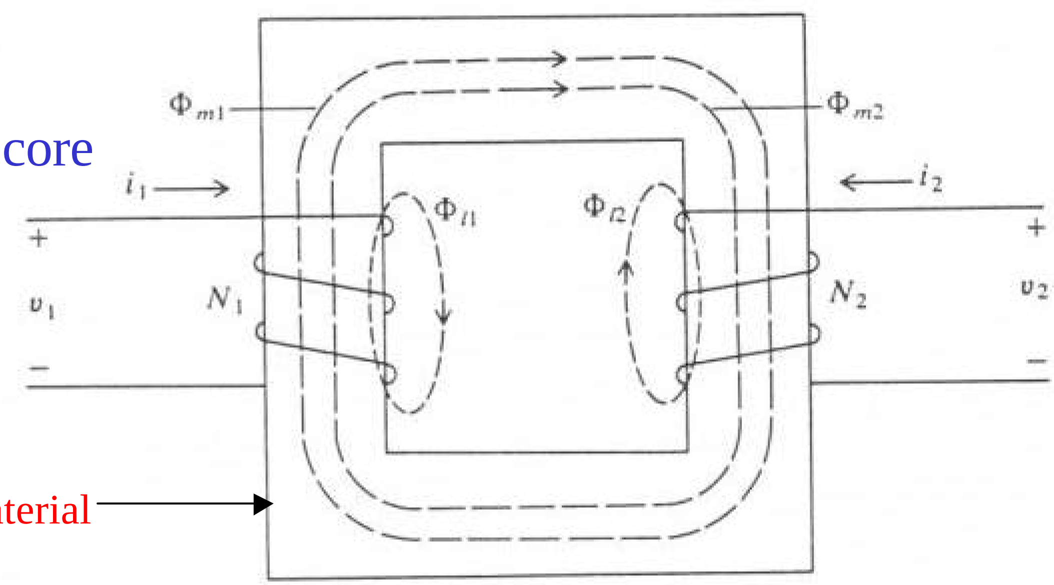
Stationary Magnetically Coupled Circuits

- Magnetically-coupled electric circuits are central to the operation of transformers and electromechanical motion devices.
- In transformers, stationary circuits are magnetically coupled for the purpose of changing the voltage and current levels.
- In electromechanical motion devices, circuits in relative motion are magnetically coupled for the purpose of transferring energy between the mechanical and electrical systems.

- <u>Goal</u>:
  - Establish the equations that describe the behavior of magnetically coupled circuits
  - Express the equations in a form convenient for analysis
- Consider first two stationary electric circuits which are magnetically coupled.

Magnetically-Coupled Circuits:
two windings on a common core

ferromagnetic material



- Some comments:
  - Ideal transformer
    - only the turns ratio is considered

$$v_2 = \frac{N_2}{N_1} v_1$$

$$i_2 = -\frac{N_1}{N_2} i_1$$

  - This treatment is often not sufficient for a detailed analysis of transformers, and it is seldom appropriate in the analysis of electromechanical motion devices, since an air gap is necessary for motion to occur; hence, windings are not as tightly coupled and leakage flux must be taken into account.
- Flux produced by each winding can be separated into two components:
  - leakage component and magnetizing component

- Leakage Flux
  - The leakage flux associated with a given winding links only that winding
- Magnetizing Flux
  - The magnetizing flux, whether it is due to current in winding 1 or winding 2, links both windings
- The flux linking each winding is expressed as:

$$\phi_1 = \phi_{\ell 1} + \phi_{m1} + \phi_{m2}$$

$$\phi_2 = \phi_{\ell 2} + \phi_{m2} + \phi_{m1}$$

  - Leakage flux is produced by current flowing in a winding and it links only the turns of that winding
  - Magnetizing flux is produced by current flowing in a winding and it links all the turns of both windings

- This is an idealization of the actual magnetic system
  - All of the leakage flux will not link all the turns of the winding producing it; so the leakage fluxes are really "equivalent" leakage fluxes
  - All the magnetizing flux of one winding will not link all of the turns of the other winding; N_1 and N_2 are often considered to be "equivalent" number of turns rather than the actual number.
- The voltage equations may be expressed as:

$$v_1 = r_1 i_1 + \frac{d\lambda_1}{dt}$$

$$v_2 = r_2 i_2 + \frac{d\lambda_2}{dt}$$

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} r_1 & 0 \\ 0 & r_2 \end{bmatrix} \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} + \frac{d}{dt} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}$$

$$\lambda_1 = N_1 \phi_1$$

$$\lambda_2 = N_2 \phi_2$$

- Assume that the magnetic system is linear; apply Ohm's Law for magnetic circuits to express the fluxes.

Typically
$$\Re_{\ell 1} \gg \Re_m$$
$$\Re_{\ell 2} \gg \Re_m$$

$$\phi_{\ell 1} = \frac{N_1 i_1}{\Re_{\ell 1}}$$

$$\phi_{m1} = \frac{N_1 i_1}{\Re_m}$$

$$\phi_{\ell 2} = \frac{N_2 i_2}{\Re_{\ell 2}}$$

$$\phi_{m2} = \frac{N_2 i_2}{\Re_m}$$

$$\phi_1 = \phi_{\ell 1} + \phi_{m1} + \phi_{m2} = \frac{N_1 i_1}{\Re_{\ell 1}} + \frac{N_1 i_1}{\Re_m} + \frac{N_2 i_2}{\Re_m}$$

$$\phi_2 = \phi_{\ell 2} + \phi_{m2} + \phi_{m1} = \frac{N_2 i_2}{\Re_{\ell 2}} + \frac{N_2 i_2}{\Re_m} + \frac{N_1 i_1}{\Re_m}$$

$\Re_m$ can be computed with sufficient accuracy

$\Re_\ell$ is usually approximated from test data

$$\lambda_1 = N_1\phi_1$$

$$= \frac{N_1^2}{\Re_{\ell 1}} i_1 + \frac{N_1^2}{\Re_m} i_1 + \frac{N_1 N_2}{\Re_m} i_2$$

$$\lambda_2 = N_2\phi_2$$

$$= \frac{N_2^2}{\Re_{\ell 2}} i_2 + \frac{N_2^2}{\Re_m} i_2 + \frac{N_2 N_1}{\Re_m} i_1$$

- When the magnetic system is linear, the flux linkages are generally expressed in terms of inductances and the currents.
  - The self inductances are:

$$L_{11} = \frac{N_1^2}{\Re_{\ell 1}} + \frac{N_1^2}{\Re_m} = L_{\ell 1} + L_{m1}$$

$$L_{22} = \frac{N_2^2}{\Re_{\ell 2}} + \frac{N_2^2}{\Re_m} = L_{\ell 2} + L_{m2}$$

– We see that: $$\frac{L_{m2}}{N_2^2} = \frac{L_{m1}}{N_1^2}$$

– The mutual inductances are:

$$\left.\begin{aligned} L_{12} &= \frac{N_1 N_2}{\Re_m} \\ L_{21} &= \frac{N_2 N_1}{\Re_m} \end{aligned}\right\} \quad L_{12} = L_{21}$$

– In this situation, with the assumed positive direction for current flow and the manner in which the windings are wound, the mutual inductances are positive. If, however, the assumed positive directions of current were such that φ_{m1} opposed φ_{m2} , then the mutual inductances would be negative.

– We see that: $$L_{12} = \frac{N_2}{N_1} L_{m1} = \frac{N_1}{N_2} L_{m2}$$

- Therefore $$\lambda_1 = L_{11}i_1 + L_{12}i_2$$ $$\lambda_2 = L_{21}i_1 + L_{22}i_2$$
  - L_{11} and L_{22} are always positive
  - $L_{12} = L_{21}$ may be positive or negative
- The voltage equations $$v_1 = r_1 i_1 + \frac{d\lambda_1}{dt}$$ $$v_2 = r_2 i_2 + \frac{d\lambda_2}{dt}$$

  (already derived) may be used for purposes of analysis. However, it is customary to perform a change of variables.

$$\lambda_1 = L_{11}i_1 + L_{12}i_2$$
$$= \left(L_{\ell 1} + L_{m1}\right)i_1 + L_{12}i_2$$
$$= L_{\ell 1}i_1 + L_{m1}\left(i_1 + \frac{N_2}{N_1}i_2\right)$$

$$\lambda_2 = L_{21}i_1 + L_{22}i_2$$
$$= \left(L_{\ell 2} + L_{m2}\right)i_2 + L_{21}i_1$$
$$= L_{\ell 2}i_2 + L_{m2}\left(\frac{N_1}{N_2}i_1 + i_2\right)$$

Two possibilities for substitute variables: $\frac{N_2}{N_1}i_2$ or $\frac{N_1}{N_2}i_1$

Let $i_2' = \frac{N_2}{N_1}i_2$ and then $N_1 i_2' = N_2 i_2$

This current i_2', when flowing through winding 1, produces the same mmf as the actual i_2 flowing through winding 2. This is said to be referring the current in winding 2 to winding 1 or to a winding with N_1 turns, whereupon winding 1 becomes the reference winding.

- We want the instantaneous power to be unchanged by this substitution of variables. Therefore

$$v_2' i_2' = v_2 i_2$$

$$v_2' = v_2 \frac{i_2}{i_2'}$$

$$v_2' = \frac{N_1}{N_2} v_2$$

- Flux linkages, which have units of volts-second, are related to the substitute flux linkages in the same way as voltages.

$$\lambda_2' = \frac{N_1}{N_2} \lambda_2$$

$$\lambda_1 = L_{\ell 1} i_1 + L_{m1}\left(i_1 + \frac{N_2}{N_1} i_2 \right)$$

$$i_2' = \frac{N_2}{N_1} i_2$$

$$\boxed{\lambda_1 = L_{\ell 1} i_1 + L_{m1}\left(i_1 + i_2' \right)}$$

$$L_{11} = L_{\ell 1} + L_{m1}$$

$$\boxed{\lambda_1 = L_{11} i_1 + L_{m1} i_2'}$$

$$\lambda_2 = L_{\ell 2} i_2 + L_{m2}\left(\frac{N_1}{N_2} i_1 + i_2 \right)$$

$$i_2' = \frac{N_2}{N_1} i_2 \qquad \frac{L_{m2}}{N_2^2} = \frac{L_{m1}}{N_1^2}$$

$$\lambda_2' = \frac{N_1}{N_2} \lambda_2 \qquad L_{\ell 2}' \equiv \left(\frac{N_1}{N_2} \right)^2 L_{\ell 2}$$

$$\boxed{\lambda_2' = L_{\ell 2}' i_2' + L_{m1}\left(i_1 + i_2' \right)}$$

$$L_{22} = L_{\ell 2} + L_{m2}$$

$$\frac{L_{m2}}{N_2^2} = \frac{L_{m1}}{N_1^2}$$

$$L_{22}' \equiv \left(\frac{N_1}{N_2} \right)^2 L_{22} = L_{\ell 2}' + L_{m1}$$

$$\boxed{\lambda_2' = L_{m1} i_1 + L_{22}' i_2'}$$

$$\lambda_1 = L_{11} i_1 + L_{m1} i_2'$$

$$\lambda_2' = L_{m1} i_1 + L_{22}' i_2'$$

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} r_1 & 0 \\ 0 & r_2 \end{bmatrix} \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} + \frac{d}{dt} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}$$

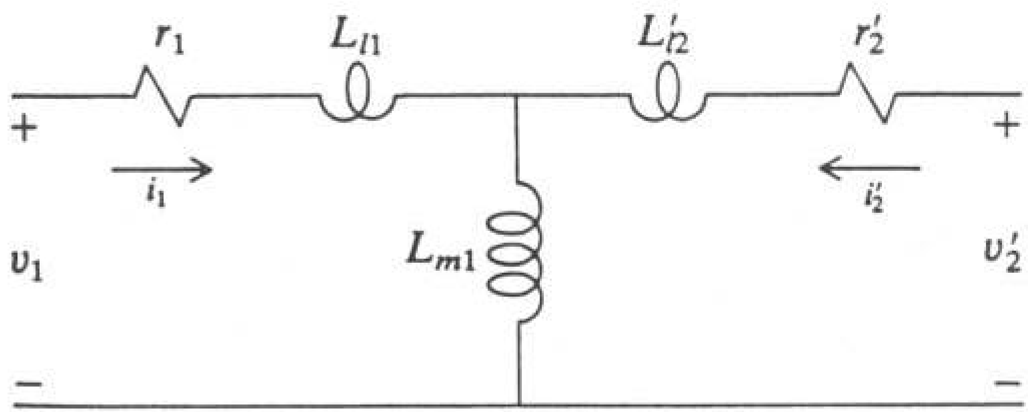
$$i_2' = \frac{N_2}{N_1} i_2$$

$$v_2' = \frac{N_1}{N_2} v_2$$

$$\begin{bmatrix} v_1 \\ v_2' \end{bmatrix} = \begin{bmatrix} r_1 & 0 \\ 0 & r_2' \end{bmatrix} \begin{bmatrix} i_1 \\ i_2' \end{bmatrix} + \frac{d}{dt} \begin{bmatrix} \lambda_1 \\ \lambda_2' \end{bmatrix}$$

$$r_2' \equiv \left(\frac{N_1}{N_2} \right)^2 r_2$$

Equivalent T circuit with winding 1 selected as the reference winding



This method may be extended to include any number of windings wound on the same core.

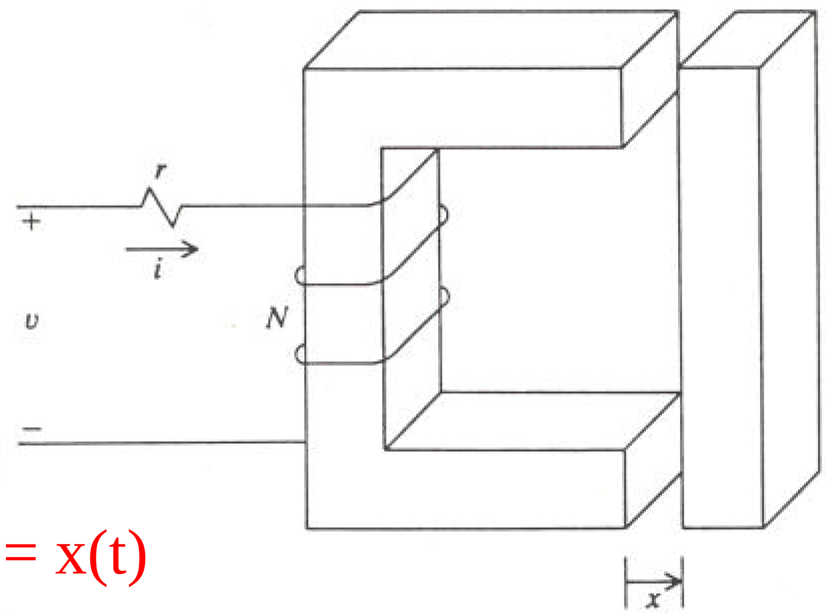
- We can now more fully appreciate the assumption that for an ideal transformer only the turns-ratio is considered.
  - the resistances and the leakage inductances are neglected
  - It is assumed that the magnetizing inductance is large so that the magnetizing current $i_1 + i_2'$ is negligibly small
- This section forms the basis of the equivalent circuits for many types of electric machines. Using a turns-ratio to refer the voltages and currents of rotor circuits of electric machines to a winding with the same number of turns as the stator windings is common practice.

Magnetic Systems with Mechanical Motion

- We introduce three elementary electromechanical systems for the purpose of establishing the voltage equations and expressions for the self- and mutual-inductances, thereby setting the stage for deriving the relationships for determining the electromagnetic force or torque established in electromechanical systems.
- The three systems are:
  - elementary version of an electromagnet
  - rotational device referred to as a reluctance machine
  - rotational device which has two windings

Elementary Electromagnet

- The system consists of:
  - stationary core with a winding of N turns
  - block of magnetic material is free to slide relative to the stationary member



$x = x(t)$

$$v = ri + \frac{d\lambda}{dt}$$ voltage equation that describes the electric system

$$\lambda = N\phi$$
$$\phi = \phi_\ell + \phi_m$$
$$\phi_\ell = \text{ leakage flux}$$
$$\phi_m = \text{ magnetizing flux}$$

flux linkages

(the magnetizing flux is common to both stationary and rotating members)

$$\phi_\ell = \frac{Ni}{\Re_\ell}$$
$$\phi_m = \frac{Ni}{\Re_m}$$

If the magnetic system is considered to be linear (saturation neglected), then, as in the case of stationary coupled circuits, we can express the fluxes in terms of reluctances.

$$\lambda = \left(\frac{N^2}{\Re_\ell} + \frac{N^2}{\Re_m} \right) i = \left(L_\ell + L_m \right) i$$

flux linkages

L_ℓ = leakage inductance

L_m = magnetizing inductance

$$\Re_m = \Re_i + 2\Re_g$$

reluctance of the magnetizing path

$\Re_i$ — total reluctance of the magnetic material of the stationary and movable members

$\Re_g$ — reluctance of one of the air gaps

$$\Re_i = \frac{\ell_i}{\mu_{ri}\mu_0 A_i}$$

$$\Re_g = \frac{x}{\mu_0 A_g}$$

Assume that the cross-sectional areas of the stationary and movable members are equal and of the same material

$$A_g = A_i$$

This may be somewhat of an oversimplification, but it is sufficient for our purposes.

$$\Re_m = \Re_i + 2\Re_g = \frac{1}{\mu_0 A_i}\left(\frac{\ell_i}{\mu_{ri}} + 2x\right)$$

$$L_m = \frac{N^2}{\frac{1}{\mu_0 A_i}\left(\frac{\ell_i}{\mu_{ri}} + 2x\right)}$$

Assume that the leakage inductance is constant.

The magnetizing inductance is clearly a function of displacement.

$x = x(t)$ and $L_m = L_m(x)$

When dealing with linear magnetic circuits wherein mechanical motion is not present, as in the case of a transformer, the change of flux linkages with respect to time was simply L(di/dt). This is not the case here.

$$\lambda(i,x) = L(x)i = \left[L_\ell + L_m(x)\right]i$$

$$\frac{d\lambda(i,x)}{dt} = \frac{\partial\lambda}{\partial i}\frac{di}{dt} + \frac{\partial\lambda}{\partial x}\frac{dx}{dt}$$

The inductance is a function of x(t).

$$v = ri + \left[L_\ell + L_m(x)\right]\frac{di}{dt} + i\frac{dL_m(x)}{dx}\frac{dx}{dt}$$

The voltage equation is a nonlinear differential equation.

$$L_m(x) = \frac{N^2}{\frac{1}{\mu_0 A_i}\left(\frac{\ell_i}{\mu_{ri}} + 2x\right)}$$

Let's look at the magnetizing inductance again.

$$L_m(x) = \frac{k}{k_0 + x} \qquad \begin{cases} k = \dfrac{N^2 \mu_0 A_i}{2} \\ k_0 = \dfrac{\ell_i}{2\mu_{ri}} \end{cases}$$

$$L_m(0) = \frac{k}{k_0} = \frac{N^2 \mu_0 \mu_{ri} A_i}{\ell_i}$$

$$L_m(x) \cong \frac{k}{x} \quad \text{for } x > 0$$

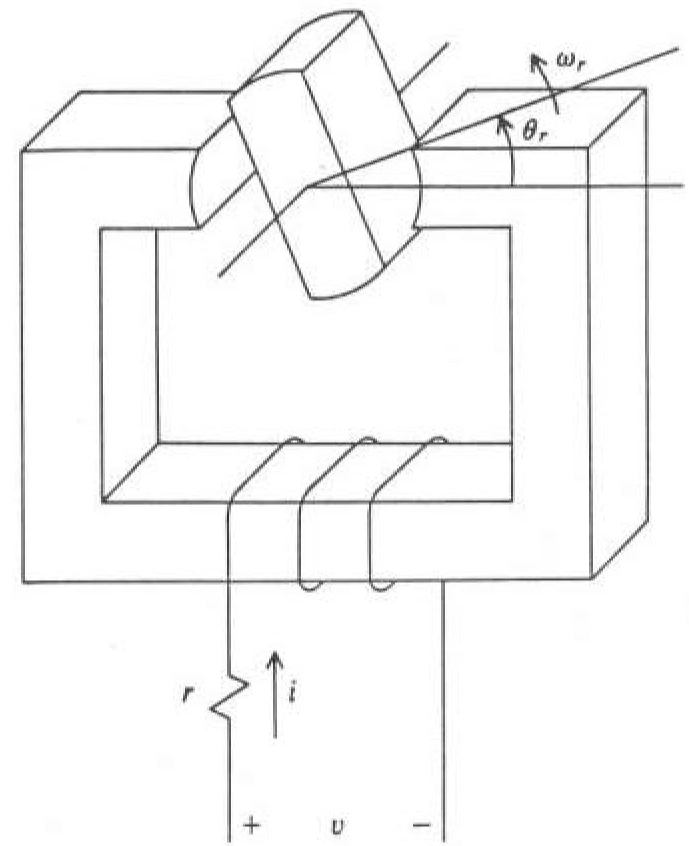
Elementary Reluctance Machine

- The machine consists of:
  - stationary core with a winding of N turns
  - moveable member which rotates

θ_r = angular displacement

ω_r = angular velocity

$$\theta_r = \int_0^t \omega_r(\xi)\,d\xi + \theta_r(0)$$



$$v = ri + \frac{d\lambda}{dt}$$

voltage equation

$$\phi = \phi_\ell + \phi_m$$

$\phi_\ell =$ leakage flux

$\phi_m =$ magnetizing flux

$$\lambda = (L_\ell + L_m)i$$

It is convenient to express the flux linkages as the product of the sum of the leakage inductance and the magnetizing inductance and the current in the winding.

$L_\ell =$ constant (independent of θ_r)

$L_m =$ periodic function of θ_r

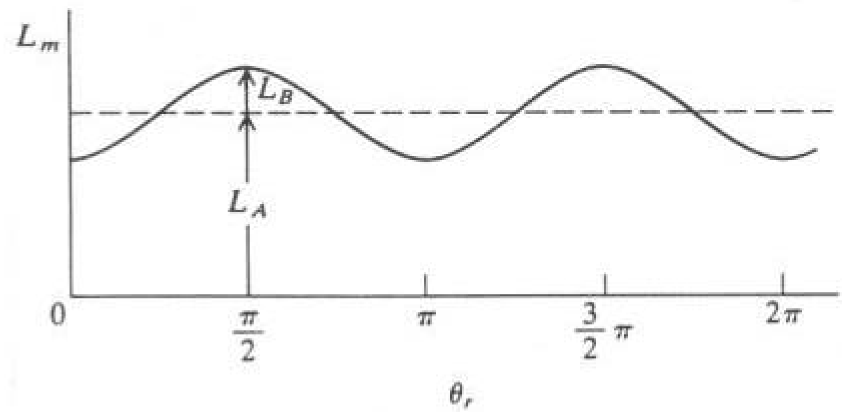
$$L_m = L_m(\theta_r)$$

$$L_m(0) = \frac{N^2}{\Re_m(0)} \quad \Rightarrow \quad \begin{cases} \Re_m \text{ is maximum} \\ L_m \text{ is minimum} \end{cases}$$

$$L_m\left(\frac{\pi}{2}\right) = \frac{N^2}{\Re_m\left(\frac{\pi}{2}\right)} \quad \Rightarrow \quad \begin{cases} \Re_m \text{ is minimum} \\ L_m \text{ is maximum} \end{cases}$$

The magnetizing inductance varies between maximum and minimum positive values twice per revolution of the rotating member.

Assume that this variation may be adequately approximated by a sinusoidal function.



$$L_m(\theta_r) = L_A - L_B \cos(2\theta_r)$$

$$L_m(0) = L_A - L_B$$

$$L_m\left(\frac{\pi}{2}\right) = L_A + L_B$$

$$L_A > L_B$$

$$L_A = \text{ average value}$$

$$L(\theta_r) = L_\ell + L_m(\theta_r)$$
$$= L_\ell + L_A - L_B \cos(2\theta_r)$$

$$v = ri + \left[L_\ell + L_m(\theta_r)\right]\frac{di}{dt} + i\frac{dL_m(\theta_r)}{d\theta_r}\frac{d\theta_r}{dt}$$

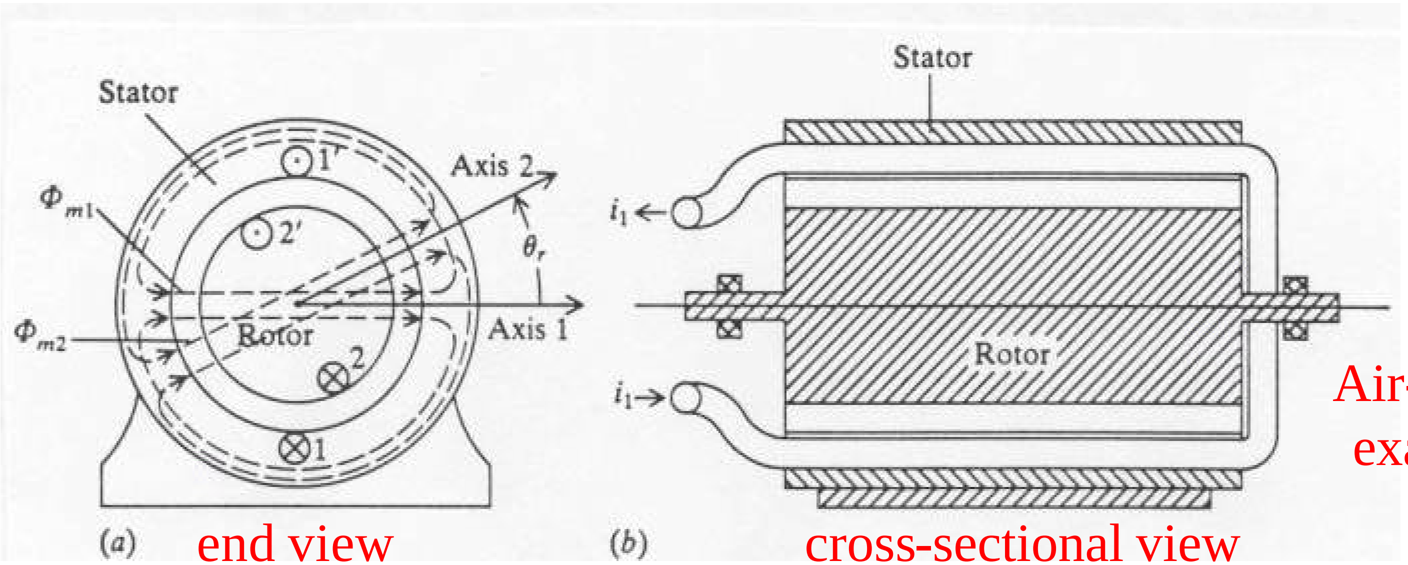
voltage equation

Windings in Relative Motion

- The rotational device shown will be used to illustrate windings in relative motion.

Winding 1: N_1 turns on stator
Winding 2: N_2 turns on rotor

Assume that the turns are concentrated in one position.



end view

cross-sectional view

Air-gap size is exaggerated.

$$v_1 = r_1 i_1 + \frac{d\lambda_1}{dt}$$

$$v_2 = r_2 i_2 + \frac{d\lambda_2}{dt}$$

voltage equations

$$\lambda_1 = L_{11} i_i + L_{12} i_2$$

$$\lambda_2 = L_{21} i_i + L_{22} i_2$$

The magnetic system is assumed linear.

$$L_{11} = L_{\ell 1} + L_{m1} = \frac{N_1^2}{\Re_{\ell 1}} + \frac{N_1^2}{\Re_m}$$

$$L_{22} = L_{\ell 2} + L_{m2} = \frac{N_2^2}{\Re_{\ell 2}} + \frac{N_2^2}{\Re_m}$$

The self-inductances L_{11} and L_{22} are constants and may be expressed in terms of leakage and magnetizing inductances.

$\Re_m$ is the reluctance of the complete magnetic path of φ_{m1} and φ_{m2} , which is through the rotor and stator iron and twice across the air gap.

Let's now consider L_{12}.

$$\theta_r = \text{ angular displacement}$$

$$\omega_r = \text{ angular velocity}$$

$$\theta_r = \int_0^t \omega_r(\xi)\, d\xi + \theta_r(0)$$

When θ_r is zero, then the coupling between windings 1 and 2 is maximum. The magnetic system of winding 1 aids that of winding 2 with positive currents assumed. Hence the mutual inductance is positive.

$$L_{12}(0) = \frac{N_1 N_2}{\Re_m}$$

When θ_r is $\pi/2$, the windings are orthogonal. The mutual coupling is zero.

$$L_{12}\left(\frac{\pi}{2}\right) = 0$$

Assume that the mutual inductance may be adequately predicted by:

$$L_{12}(\theta_r) = L_{sr}\cos(\theta_r)$$

$$L_{sr} = \frac{N_1 N_2}{\Re_m}$$

L_{sr} is the amplitude of the sinusoidal mutual inductance between the stator and rotor windings.

$$v_1 = r_1 i_1 + \frac{d\lambda_1}{dt}$$

$$v_2 = r_2 i_2 + \frac{d\lambda_2}{dt}$$

In writing the voltage equations, the total derivative of the flux linkages is required.

Principles of Electromechanical Energy Conversion

- Why do we study this?
  - Electromechanical energy conversion theory is the cornerstone for the analysis of electromechanical motion devices.
  - The theory allows us to express the electromagnetic force or torque in terms of the device variables such as the currents and the displacement of the mechanical system.
  - Since numerous types of electromechanical devices are used in motion systems, it is desirable to establish methods of analysis which may be applied to a variety of electromechanical devices rather than just electric machines.

- Plan
  - Establish analytically the relationships which can be used to express the electromagnetic force or torque.
  - Develop a general set of formulas which are applicable to all electromechanical systems with a single mechanical input.
  - Detailed analysis of:
    - Elementary electromagnet
    - Elementary single-phase reluctance machine
    - Windings in relative motion

Lumped Parameters vs. Distributed Parameters

- If the physical size of a device is small compared to the wavelength associated with the signal propagation, the device may be considered lumped and a lumped (network) model employed.

$$\lambda = \frac{v}{f} \quad \begin{cases} \lambda = \text{wavelength (distance/cycle)} \\ v = \text{velocity of wave propagation (distance/second)} \\ f = \text{signal frequency (Hz)} \end{cases}$$

- Consider the electrical portion of an audio system:
  - 20 to 20,000 Hz is the audio range

$$\lambda = \frac{186{,}000 \text{ miles/second}}{20{,}000 \text{ cycles/second}} = 9.3 \text{ miles/cycle}$$

Conservative Force Field

- A force field acting on an object is called *conservative* if the work done in moving the object from one point to another is independent of the path joining the two points.

$$\vec{F} = F_1\hat{i} + F_2\hat{j} + F_3\hat{k}$$

$$\int_C \vec{F}\cdot\overrightarrow{dr} \text{ is independent of path if and only if } \nabla\times\vec{F} = 0 \text{ or } \vec{F} = \nabla\phi$$

$\vec{F}\cdot\overrightarrow{dr}$ is an exact differential

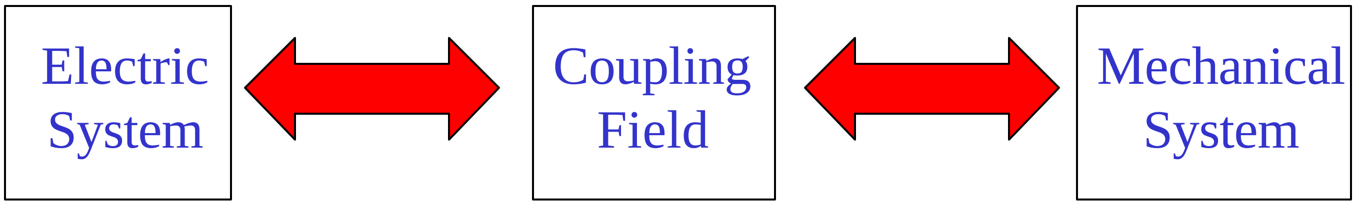
$F_1dx + F_2dy + F_3dz = d\phi$ where $\phi(x,y,z)$

$$\int_{(x_1,y_1,z_1)}^{(x_2,y_2,z_2)} \vec{F}\cdot\overrightarrow{dr} = \int_{(x_1,y_1,z_1)}^{(x_2,y_2,z_2)} d\phi = \phi(x_2,y_2,z_2) - \phi(x_1,y_1,z_1)$$

Energy Balance Relationships

- Electromechanical System
  - Comprises
    - Electric system
    - Mechanical system
    - Means whereby the electric and mechanical systems can interact
  - Interactions can take place through any and all electromagnetic and electrostatic fields which are common to both systems, and energy is transferred as a result of this interaction.
  - Both electrostatic and electromagnetic coupling fields may exist simultaneously and the system may have any number of electric and mechanical subsystems.

- **Electromechanical System in Simplified Form:**



  - Neglect electromagnetic radiation
  - Assume that the electric system operates at a frequency sufficiently low so that the electric system may be considered as a lumped-parameter system

- **Energy Distribution**

$$W_E = W_e + W_{eL} + W_{eS}$$

$$W_M = W_m + W_{mL} + W_{mS}$$

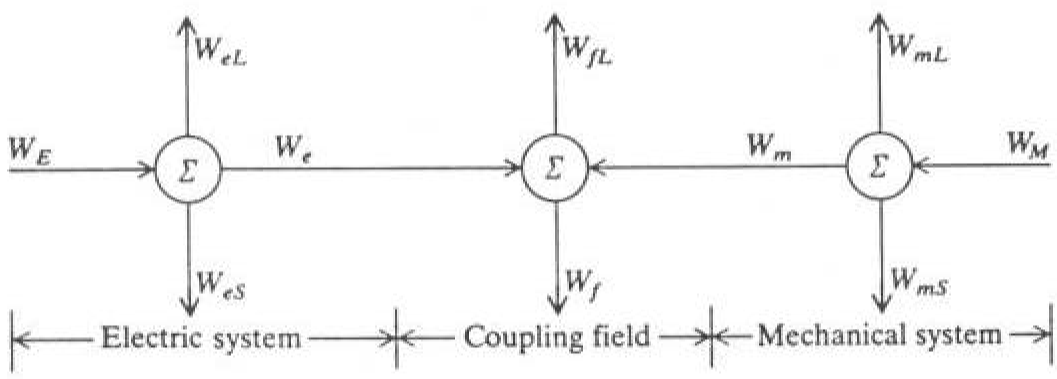
  - W_E = total energy supplied *by* the electric source (+)
  - W_M = total energy supplied *by* the mechanical source (+)

- W_{eS} = energy stored in the electric or magnetic fields which are not coupled with the mechanical system
- W_{eL} = heat loss associated with the electric system, excluding the coupling field losses, which occurs due to:
  - the resistance of the current-carrying conductors
  - the energy dissipated in the form of heat owing to hysteresis, eddy currents, and dielectric losses external to the coupling field
- W_e = energy transferred to the coupling field by the electric system
- W_{mS} = energy stored in the moving member and the compliances of the mechanical system
- W_{mL} = energy loss of the mechanical system in the form of heat due to friction
- W_m = energy transferred to the coupling field by the mechanical system

- $W_F = W_f + W_{fL}$ = total energy transferred to the coupling field
  - W_f = energy stored in the coupling field
  - W_{fL} = energy dissipated in the form of heat due to losses within the coupling field (eddy current, hysteresis, or dielectric losses)
- <u>Conservation of Energy</u>

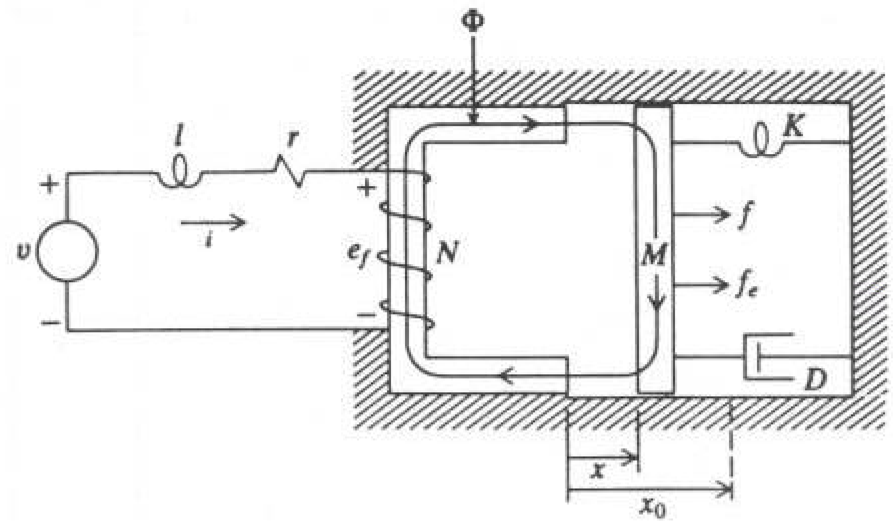
$$W_f + W_{fL} = (W_E - W_{eL} - W_{eS}) + (W_M - W_{mL} - W_{mS})$$

$$W_f + W_{fL} = W_e + W_m$$

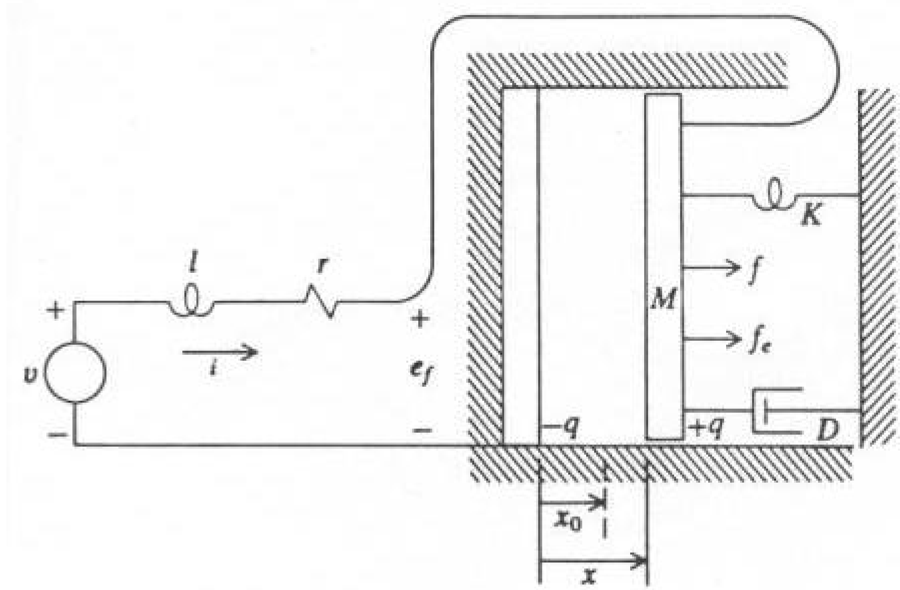


- The actual process of converting electric energy to mechanical energy (or vice versa) is independent of:
  - The loss of energy in either the electric or the mechanical systems (W_{eL} and W_{mL})
  - The energies stored in the electric or magnetic fields which are not in common to both systems (W_{eS})
  - The energies stored in the mechanical system (W_{mS})
- If the losses of the coupling field are neglected, then the field is <u>conservative</u> and $W_f = W_e + W_m$.
- <u>Consider two examples of elementary electromechanical systems</u>
  - Magnetic coupling field
  - Electric field as a means of transferring energy

v = voltage of electric source
f = externally-applied mechanical force
f_e = electromagnetic or electrostatic force
r = resistance of the current-carrying conductor
ℓ = inductance of a linear (conservative) electromagnetic system which does not couple the mechanical system
M = mass of moveable member
K = spring constant
D = damping coefficient
x_0 = zero force or equilibrium position of the mechanical system (f_e = 0, f = 0)



Electromechanical System with Magnetic Field



Electromechanical System with Electric Field

$$v = ri + \ell \frac{di}{dt} + e_f$$

voltage equation that describes the electric systems; e_f is the voltage drop due to the coupling field

$$f = M\frac{d^2x}{dt^2} + D\frac{dx}{dt} + K(x - x_0) - f_e$$

Newton's Law of Motion

$$W_E = \int (vi)\,dt$$

$$W_M = \int (f)\,dx = \int \left(f\frac{dx}{dt} \right) dt$$

Since power is the time rate of energy transfer, this is the total energy supplied by the electric and mechanical sources

$$v = ri + \ell \frac{di}{dt} + e_f$$

$$W_E = \int (vi)\,dt$$

$$W_E = r\int (i^2)\,dt + \ell \int \left(i\frac{di}{dt} \right) dt + \int (e_f i)\,dt$$

$$= W_{eL} + W_{eS} + W_e$$

$$f = M\frac{d^2x}{dt^2} + D\frac{dx}{dt} + K\left(x - x_0\right) - f_e$$

$$W_M = \int (f)dx = \int \left(f\frac{dx}{dt} \right) dt$$

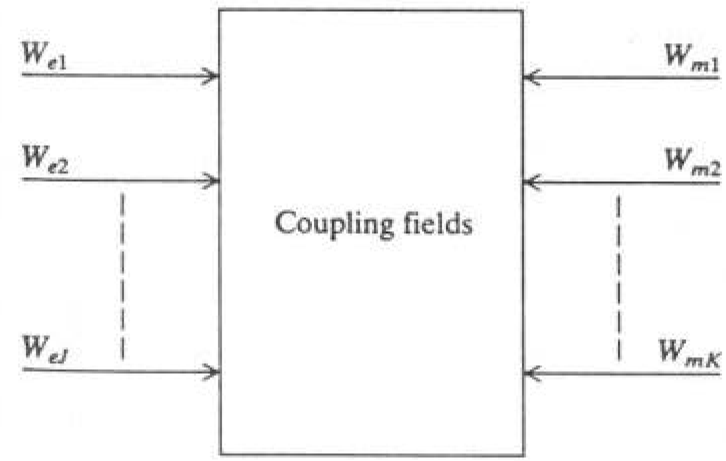
$$W_M = M\int \left(\frac{d^2x}{dt^2} \right) dx + D\int \left(\frac{dx}{dt} \right)^2 dt + K\int (x - x_0)dx - \int (f_e)dx$$

Σ W_{mS} $\qquad$ W_{mL} $\qquad$ W_m

$$W_f = W_e + W_m = \int (e_f i)dt - \int (f_e)dx$$

total energy transferred to the coupling field

- These equations may be readily extended to include an electromechanical system with any number of electrical and mechanical inputs and any number of coupling fields.



- We will consider devices with only one mechanical input, but with possibly multiple electric inputs. In all cases, however, the multiple electric inputs have a common coupling field.

$$W_f = \sum_{j=1}^{J} W_{ej} + \sum_{k=1}^{K} W_{mk}$$

Total energy supplied to the coupling field

$$\sum_{j=1}^{J} W_{ej} = \int \sum_{j=1}^{J} e_{fj} i_j dt$$

Total energy supplied to the coupling field from the electric inputs

$$\sum_{k=1}^{K} W_{mk} = -\int \sum_{k=1}^{K} f_{ek} dx_k$$

Total energy supplied to the coupling field from the mechanical inputs

$$W_f = \int \sum_{j=1}^{J} e_{fj} i_j dt - \int f_e dx$$

$$dW_f = \sum_{j=1}^{J} e_{fj} i_j dt - f_e dx$$

With one mechanical input and multiple electric inputs, the energy supplied to the coupling field, in both integral and differential form

Energy in Coupling Field

- We need to derive an expression for the energy stored in the coupling field before we can solve for the electromagnetic force f_e.
- We will neglect all losses associated with the electric or magnetic coupling field, whereupon the field is assumed to be conservative and the energy stored therein is a function of the state of the electrical and mechanical variables and not the manner in which the variables reached that state.
- This assumption is not as restrictive as it might first appear.

- The ferromagnetic material is selected and arranged in laminations so as to minimize the hysteresis and eddy current losses.
- Nearly all of the energy stored in the coupling field is stored in the air gap of the electromechanical device. Air is a conservative medium and all of the energy stored therein can be returned to the electric or mechanical systems.

- We will take advantage of the conservative field assumption in developing a mathematical expression for the field energy. We will fix mathematically the position of the mechanical system associated with the coupling field and then excite the electric system with the displacement of the mechanical system held fixed.

- During the excitation of the electric inputs, $dx = 0$, hence, W_m is zero even though electromagnetic and electrostatic forces occur.
- Therefore, with the displacement held fixed, the energy stored in the coupling field during the excitation of the electric inputs is equal to the energy supplied to the coupling field by the electric inputs.
- With $dx = 0$, the energy supplied from the electric system is:

$$W_f = \int \sum_{j=1}^{J} e_{fj} i_j dt - \int f_e \underbrace{dx}_{0}$$

$$W_f = \int \sum_{j=1}^{J} e_{fj} i_j dt$$

- For a singly excited electromagnetic system:

$$e_f = \frac{d\lambda}{dt}$$

$$W_f = \int (i) d\lambda \quad \text{with } dx = 0$$

$$W_f = \int (i) d\lambda$$

Area represents energy stored in the field at the instant when $\lambda = \lambda_a$ and $i = i_a$.

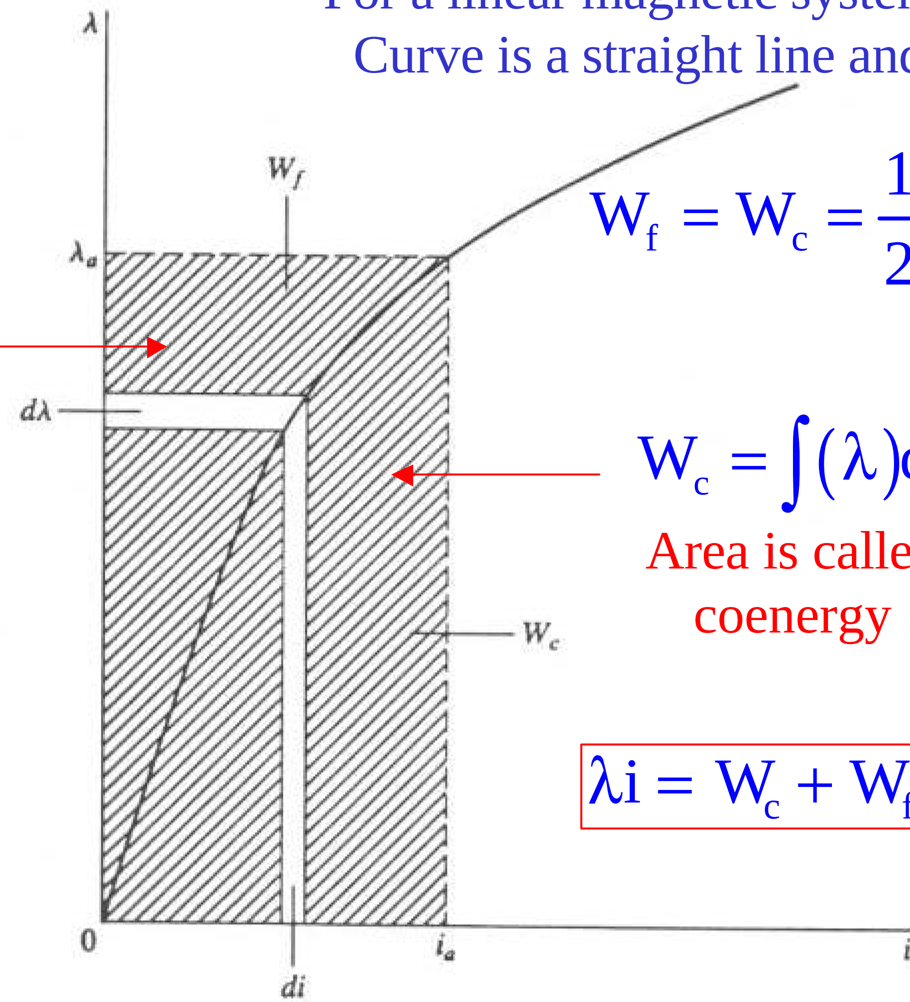
For a linear magnetic system: Curve is a straight line and

$$W_f = W_c = \frac{1}{2}\lambda i$$

$$W_c = \int (\lambda) di$$

Area is called coenergy

$$\lambda i = W_c + W_f$$



Graph
Stored energy and coenergy in a magnetic field of a singly excited electromagnetic device

- The $l i$ relationship need not be linear, it need only be single-valued, a property which is characteristic to a conservative or lossless field.
- Also, since the coupling field is conservative, the energy stored in the field with $l = l_a$ and $i = i_a$ is independent of the excursion of the electrical and mechanical variables before reaching this state.
- The displacement x defines completely the influence of the mechanical system upon the coupling field; however, since l and i are related, only one is needed in addition to x in order to describe the state of the electromechanical system.

- If *i* and *x* are selected as the independent variables, it is convenient to express the field energy and the flux linkages as

$$W_f = W_f(i, x)$$

$$\lambda = \lambda(i, x)$$

$$d\lambda = \frac{\partial \lambda(i,x)}{\partial i} di + \frac{\partial \lambda(i,x)}{\partial x} dx$$

$$d\lambda = \frac{\partial \lambda(i,x)}{\partial i} di \quad \text{with } dx = 0$$

$$W_f = \int (i) d\lambda = \int i \frac{\partial \lambda(i,x)}{\partial i} di = \int_0^i \xi \frac{\partial \lambda(\xi, x)}{\partial \xi} d\xi$$

Energy stored in the field of a singly excited system

- The coenergy in terms of *i* and *x* may be evaluated as

$$W_c(i,x) = \int \lambda(i,x)\,di = \int_0^i \lambda(\xi,x)\,d\xi$$

- For a linear electromagnetic system, the *l i* plots are straightline relationships. Thus, for the singly excited magnetically linear system, $\lambda(i,x) = L(x)i$, where *L(x)* is the inductance.
- Let's evaluate $W_f(i,x)$.

$$d\lambda = \frac{\partial\lambda(i,x)}{\partial i}di \quad \text{with } dx = 0$$

$$d\lambda = L(x)\,di$$

$$W_f(i,x) = \int_0^i \xi L(x)\,d\xi = \frac{1}{2}L(x)i^2$$

- The field energy is a state function and the expression describing the field energy in terms of the state variables is valid regardless of the variations in the system variables.
- W_f expresses the field energy regardless of the variations in *L(x)* and *i*. The fixing of the mechanical system so as to obtain an expression for the field energy is a mathematical convenience and not a restriction upon the result.

$$W_f(i,x) = \int_0^i \xi L(x) d\xi = \frac{1}{2} L(x) i^2$$

- In the case of a multiexcited electromagnetic system, an expression for the field energy may be obtained by evaluating the following relation with dx = 0:

$$W_f = \int \sum_{j=1}^{J} i_j d\lambda_j$$

- Since the coupling field is considered conservative, this expression may be evaluated independent of the order in which the flux linkages or currents are brought to their final values.
- Let's consider a doubly excited electric system with one mechanical input.

$$W_f(i_1, i_2, x) = \int \left[i_1 d\lambda_1(i_1, i_2, x) + i_2 d\lambda_2(i_1, i_2, x) \right] \quad \text{with } dx = 0$$

- The result is:

$$W_f\left(i_1,i_2,x\right)=\int_0^{i_1}\xi\frac{\partial\lambda_1\left(\xi,0,x\right)}{\partial\xi}d\xi+$$

$$\int_0^{i_2}\left[i_1\frac{\partial\lambda_1\left(i_1,\xi,x\right)}{\partial\xi}+\xi\frac{\partial\lambda_2\left(i_1,\xi,x\right)}{\partial\xi}\right]d\xi$$

- The first integral results from the first step of the evaluation with i_1 as the variable of integration and with $i_2 = 0$ and $di_2 = 0$. The second integral comes from the second step of the evaluation with i_1 equal to its final value ($di_1 = 0$) and i_2 as the variable of integration. The order of allowing the currents to reach their final state is irrelevant.

- Let's now evaluate the energy stored in a magnetically linear electromechanical system with two electrical inputs and one mechanical input.

$$\lambda_1(i_1, i_2, x) = L_{11}(x) i_1 + L_{12}(x) i_2$$
$$\lambda_2(i_1, i_2, x) = L_{21}(x) i_1 + L_{22}(x) i_2$$

- The self-inductances $L_{11}(x)$ and $L_{22}(x)$ include the leakage inductances.
- With the mechanical displacement held constant (dx = 0):

$$d\lambda_1(i_1, i_2, x) = L_{11}(x) di_1 + L_{12}(x) di_2$$
$$d\lambda_2(i_1, i_2, x) = L_{21}(x) di_1 + L_{22}(x) di_2$$

- Substitution into:

$$W_f(i_1, i_2, x) = \int_0^{i_1} \xi \frac{\partial \lambda_1(\xi, 0, x)}{\partial \xi} d\xi + \int_0^{i_2} \left[i_1 \frac{\partial \lambda_1(i_1, \xi, x)}{\partial \xi} + \xi \frac{\partial \lambda_2(i_1, \xi, x)}{\partial \xi} \right] d\xi$$

- Yields:

$$W_f(i_1, i_2, x) = \int_0^{i_1} \xi L_{11}(x) d\xi + \int_0^{i_2} \left[i_1 L_{12}(x) + \xi L_{22}(x) \right] d\xi = \frac{1}{2} L_{11}(x) i_1^2 + L_{12}(x) i_1 i_2 + \frac{1}{2} L_{22}(x) i_2^2$$

- It follows that the total field energy of a linear electromagnetic system with J electric inputs may be expressed as:

$$W_f\left(i_1,\ldots,i_j,x\right)=\frac{1}{2}\sum_{p=1}^{J}\sum_{q=1}^{J}L_{pq}i_p i_q$$

Electromagnetic and Electrostatic Forces

- Energy Balance Equation:

$$W_f = \int \sum_{j=1}^{J} e_{fj} i_j dt - \int f_e dx$$

$$dW_f = \sum_{j=1}^{J} e_{fj} i_j dt - f_e dx$$

$$f_e dx = \sum_{j=1}^{J} e_{fj} i_j dt - dW_f$$

- To obtain an expression for f_e, it is first necessary to express W_f and then take its total derivative. The total differential of the field energy is required here.

- The force or torque in any electromechanical system may be evaluated by employing: $dW_f = dW_e + dW_m$
- We will derive the force equations for electro-mechanical systems with one mechanical input and *J* electrical inputs.
- For an electromagnetic system: $f_e dx = \sum_{j=1}^{J} i_j d\lambda_j - dW_f$
- Select i_j and x as independent variables: $W_f = W_f\left(\vec{i}, x\right)$

$$\lambda_j = \lambda_j\left(\vec{i}, x\right)$$

$$dW_f = \sum_{j=1}^{J}\left[\frac{\partial W_f\left(\vec{i}, x\right)}{\partial i_j} di_j\right] + \frac{\partial W_f\left(\vec{i}, x\right)}{\partial x} dx$$

$$d\lambda_j = \sum_{n=1}^{J}\left[\frac{\partial \lambda_j\left(\vec{i}, x\right)}{\partial i_n} di_n\right] + \frac{\partial \lambda_j\left(\vec{i}, x\right)}{\partial x} dx$$

- The summation index n is used so as to avoid confusion with the subscript j since each $d\,l_j$ must be evaluated for changes in all currents to account for mutual coupling between electric systems.
- Substitution:

$$dW_f = \sum_{j=1}^{J}\left[\frac{\partial W_f\left(\vec{i},x\right)}{\partial i_j}di_j\right] + \frac{\partial W_f\left(\vec{i},x\right)}{\partial x}dx$$

$$d\lambda_j = \sum_{n=1}^{J}\left[\frac{\partial \lambda_j\left(\vec{i},x\right)}{\partial i_n}di_n\right] + \frac{\partial \lambda_j\left(\vec{i},x\right)}{\partial x}dx$$

into

$$f_e dx = \sum_{j=1}^{J} i_j d\lambda_j - dW_f$$

- Result:

$$f_e(\vec{i},x)dx = \sum_{j=1}^{J} i_j \left\{ \sum_{n=1}^{J} \left[\frac{\partial \lambda j(\vec{i},x)}{\partial i_n} di_n \right] + \frac{\partial \lambda_j(\vec{i},x)}{\partial x} dx \right\}$$

$$-\sum_{j=1}^{J} \left[\frac{\partial W_f(\vec{i},x)}{\partial i_j} di_j \right] + \frac{\partial W_f(\vec{i},x)}{\partial x} dx$$

$$f_e(\vec{i},x)dx = \left\{ \sum_{j=1}^{J} \left[i_j \frac{\partial \lambda_j(\vec{i},x)}{\partial x} \right] - \frac{\partial W_f(\vec{i},x)}{\partial x} \right\} dx$$

$$+\sum_{j=1}^{J} \left\{ i_j \sum_{n=1}^{J} \left[\frac{\partial \lambda_j(\vec{i},x)}{\partial i_n} di_n \right] - \frac{\partial W_f(\vec{i},x)}{\partial i_j} di_j \right\}$$

- This equation is satisfied provided that:

$$f_e\left(\vec{i},x\right)=\sum_{j=1}^{J}\left[i_j\frac{\partial\lambda_j\left(\vec{i},x\right)}{\partial x}\right]-\frac{\partial W_f\left(\vec{i},x\right)}{\partial x}$$

$$0=\sum_{j=1}^{J}\left\{i_j\sum_{n=1}^{J}\left[\frac{\partial\lambda_j\left(\vec{i},x\right)}{\partial i_n}di_n\right]-\frac{\partial W_f\left(\vec{i},x\right)}{\partial i_j}di_j\right\}$$

- The first equation can be used to evaluate the force on the mechanical system with *i* and *x* selected as independent variables.

- We can incorporate an expression for coenergy and obtain a second force equation:

$$W_c = \sum_{j=1}^{J} i_j \lambda_j - W_f$$

- Since *i* and *x* are independent variables, the partial derivative with respect to *x* is:

$$\frac{\partial W_c\left(\vec{i}, x\right)}{\partial x} = \sum_{j=1}^{J}\left[i_j \frac{\partial \lambda_j\left(\vec{i}, x\right)}{\partial x}\right] - \frac{\partial W_f\left(\vec{i}, x\right)}{\partial x}$$

- Substitution:

$$f_e\left(\vec{i}, x\right) = \sum_{j=1}^{J}\left[i_j \frac{\partial \lambda_j\left(\vec{i}, x\right)}{\partial x}\right] - \frac{\partial W_f\left(\vec{i}, x\right)}{\partial x} = \frac{\partial W_c\left(\vec{i}, x\right)}{\partial x}$$

- Note:
  - Positive f_e and positive dx are in the same direction
  - If the magnetic system is linear, $W_c = W_f$.
- Summary:

$$f_e\left(\vec{i},x\right)=\sum_{j=1}^{J}\left[i_j\frac{\partial\lambda_j\left(\vec{i},x\right)}{\partial x}\right]-\frac{\partial W_f\left(\vec{i},x\right)}{\partial x}$$

$$f_e\left(\vec{i},x\right)=\frac{\partial W_c\left(\vec{i},x\right)}{\partial x}$$

$$T_e\left(\vec{i},\theta\right)=\sum_{j=1}^{J}\left[i_j\frac{\partial\lambda_j\left(\vec{i},\theta\right)}{\partial\theta}\right]-\frac{\partial W_f\left(\vec{i},\theta\right)}{\partial\theta}$$

$f_e \Rightarrow T_e$

$x \Rightarrow \theta$

$$T_e\left(\vec{i},\theta\right)=\frac{\partial W_c\left(\vec{i},\theta\right)}{\partial\theta}$$

- By a similar procedure, force equations may be derived with flux linkages $\lambda_1, \ldots, \lambda_j$ of the J windings and x as independent variables. The relations, given without proof, are:

$$f_e\left(\vec{\lambda}, x\right) = -\sum_{j=1}^{J}\left[\lambda_j \frac{\partial i_j\left(\vec{\lambda}, x\right)}{\partial x}\right] + \frac{\partial W_c\left(\vec{\lambda}, x\right)}{\partial x}$$

$$f_e\left(\vec{\lambda}, x\right) = -\frac{\partial W_f\left(\vec{\lambda}, x\right)}{\partial x}$$

$$T_e\left(\vec{\lambda}, \theta\right) = -\sum_{j=1}^{J}\left[\lambda_j \frac{\partial i_j\left(\vec{\lambda}, \theta\right)}{\partial \theta}\right] + \frac{\partial W_c\left(\vec{\lambda}, \theta\right)}{\partial \theta}$$

$$T_e\left(\vec{\lambda}, \theta\right) = -\frac{\partial W_f\left(\vec{\lambda}, \theta\right)}{\partial \theta}$$

- One may prefer to determine the electromagnetic force or torque by starting with the relationship

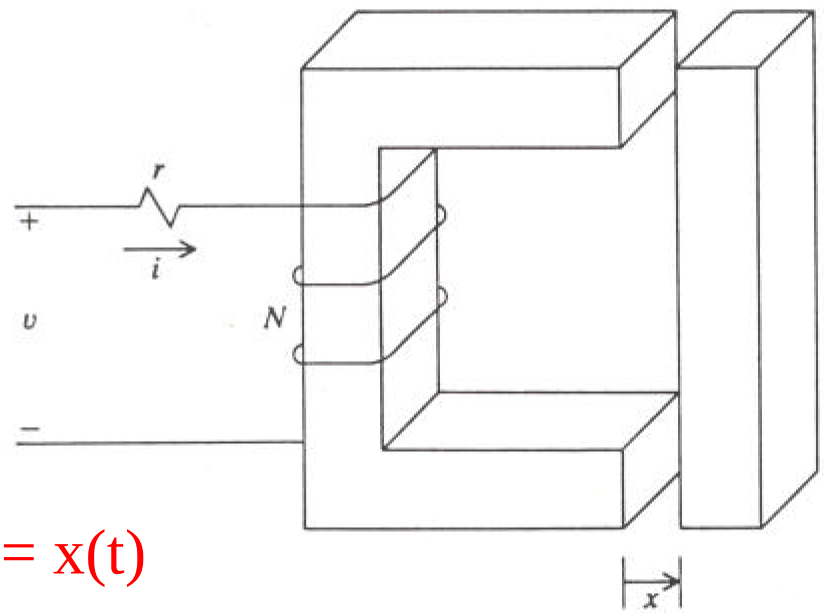
$$dW_f = dW_e + dW_m$$

rather than by selecting a formula.

- Example:
  - Given: $\lambda = \left[1 + a(x)\right] i^2$
  - Find $f_e(i,x)$

Elementary Electromagnet

- The system consists of:
  - stationary core with a winding of N turns
  - block of magnetic material is free to slide relative to the stationary member



$x = x(t)$

$$v = ri + \frac{d\lambda}{dt}$$

voltage equation that describes the electric system

$$\lambda = N\phi$$

$$\phi = \phi_\ell + \phi_m$$

ϕ_ℓ = leakage flux

ϕ_m = magnetizing flux

flux linkages

(the magnetizing flux is common to both stationary and rotating members)

$$\phi_\ell = \frac{Ni}{\Re_\ell}$$

$$\phi_m = \frac{Ni}{\Re_m}$$

If the magnetic system is considered to be linear (saturation neglected), then, as in the case of stationary coupled circuits, we can express the fluxes in terms of reluctances.

$$\lambda = \left(\frac{N^2}{\Re_\ell} + \frac{N^2}{\Re_m} \right) i = \left(L_\ell + L_m \right) i$$

flux linkages

$L_\ell =$ leakage inductance

$L_m =$ magnetizing inductance

$$\Re_m = \Re_i + 2\Re_g$$

reluctance of the magnetizing path

$\Re_i$ — total reluctance of the magnetic material of the stationary and movable members

$\Re_g$ — reluctance of one of the air gaps

$$\Re_i = \frac{\ell_i}{\mu_{ri}\mu_0 A_i}$$

$$\Re_g = \frac{x}{\mu_0 A_g}$$

Assume that the cross-sectional areas of the stationary and movable members are equal and of the same material

$$A_g = A_i$$

This may be somewhat of an oversimplification, but it is sufficient for our purposes.

$$\Re_m = \Re_i + 2\Re_g$$

$$= \frac{1}{\mu_0 A_i}\left(\frac{\ell_i}{\mu_{ri}} + 2x\right)$$

$$L_m = \frac{N^2}{\frac{1}{\mu_0 A_i}\left(\frac{\ell_i}{\mu_{ri}} + 2x\right)}$$

Assume that the leakage inductance is constant.
The magnetizing inductance is clearly a function of displacement.
$x = x(t)$ and $L_m = L_m(x)$

When dealing with linear magnetic circuits wherein mechanical motion is not present, as in the case of a transformer, the change of flux linkages with respect to time was simply L(di/dt). This is not the case here.

$$\lambda(i,x) = L(x)i = [L_\ell + L_m(x)]i$$

$$\frac{d\lambda(i,x)}{dt} = \frac{\partial\lambda}{\partial i}\frac{di}{dt} + \frac{\partial\lambda}{\partial x}\frac{dx}{dt}$$

The inductance is a function of x(t).

$$v = ri + [L_\ell + L_m(x)]\frac{di}{dt} + i\frac{dL_m(x)}{dx}\frac{dx}{dt}$$

The voltage equation is a nonlinear differential equation.

$$L_m(x) = \frac{N^2}{\frac{1}{\mu_0 A_i}\left(\frac{\ell_i}{\mu_{ri}} + 2x\right)}$$

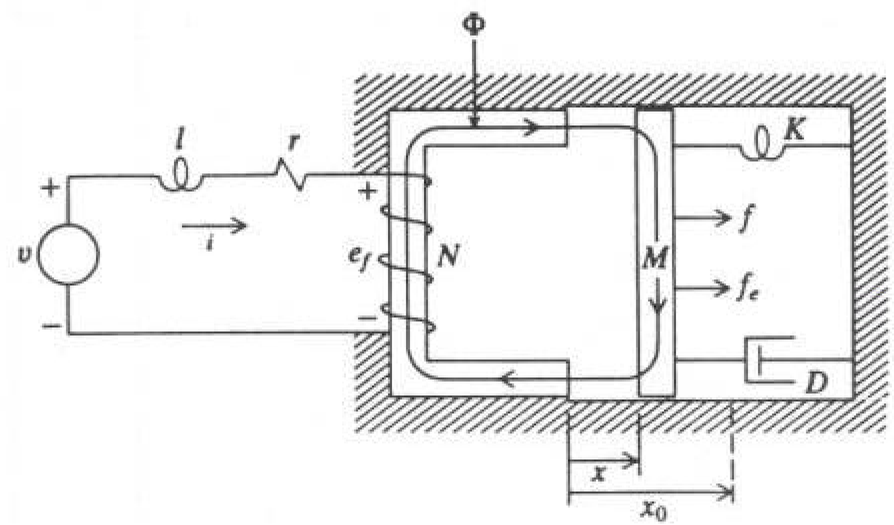
Let's look at the magnetizing inductance again.

$$L_m(x) = \frac{k}{k_0 + x} \qquad \begin{cases} k = \dfrac{N^2\mu_0 A_i}{2} \\[2ex] k_0 = \dfrac{\ell_i}{2\mu_{ri}} \end{cases}$$

$$L_m(0) = \frac{k}{k_0} = \frac{N^2\mu_0\mu_{ri}A_i}{\ell_i}$$

$$L_m(x) \cong \frac{k}{x} \quad \text{for } x > 0$$

Detailed diagram of electromagnet for further analysis



Electromagnet

$$L_m(x) \cong \frac{k}{x} \quad \text{for } x > 0$$

Use this approximation

$$L(x) \cong L_\ell + L_m(x) = L_\ell + \frac{k}{x} \quad \text{for } x > 0$$

$$\lambda(i,x) = L(x)i = \left[L_\ell + L_m(x)\right]i$$

The system is magnetically linear: $W_f(i,x) = W_c(i,x) = \frac{1}{2}L(x)i^2$

$$f_e(\vec{i},x) = \sum_{j=1}^{J}\left[i_j \frac{\partial \lambda_j(\vec{i},x)}{\partial x}\right] - \frac{\partial W_f(\vec{i},x)}{\partial x}$$

$$f_e(\vec{i},x) = \frac{\partial W_c(\vec{i},x)}{\partial x}$$

$$f_e(i,x) = \frac{1}{2}i^2 \frac{\partial L(x)}{\partial x} = -\frac{ki^2}{2x^2}$$

- The force f_e is always negative; it pulls the moving member to the stationary member. In other words, an electromagnetic force is set up so as to minimize the reluctance (maximize the inductance) of the magnetic system.
- Equations of motion:

$$v = ri + \ell \frac{di}{dt} + e_f$$

$$f = M\frac{d^2x}{dt^2} + D\frac{dx}{dt} + K\left(x - x_0\right) - f_e$$

Steady-State Operation
(if *v* and *f* are constant)

$$v = ri$$

$$f = K\left(x - x_0\right) - f_e$$

Steady-State Operation of an Electromagnet

$$f = K(x - x_0) - f_e$$

$$-f_e = f - K(x - x_0)$$

$$-\left(-\frac{ki^2}{2x^2}\right) = f - K(x - x_0)$$

Parameters:

$r = 10\ \Omega$

$K = 2667$ N/m

$x_0 = 3$ mm

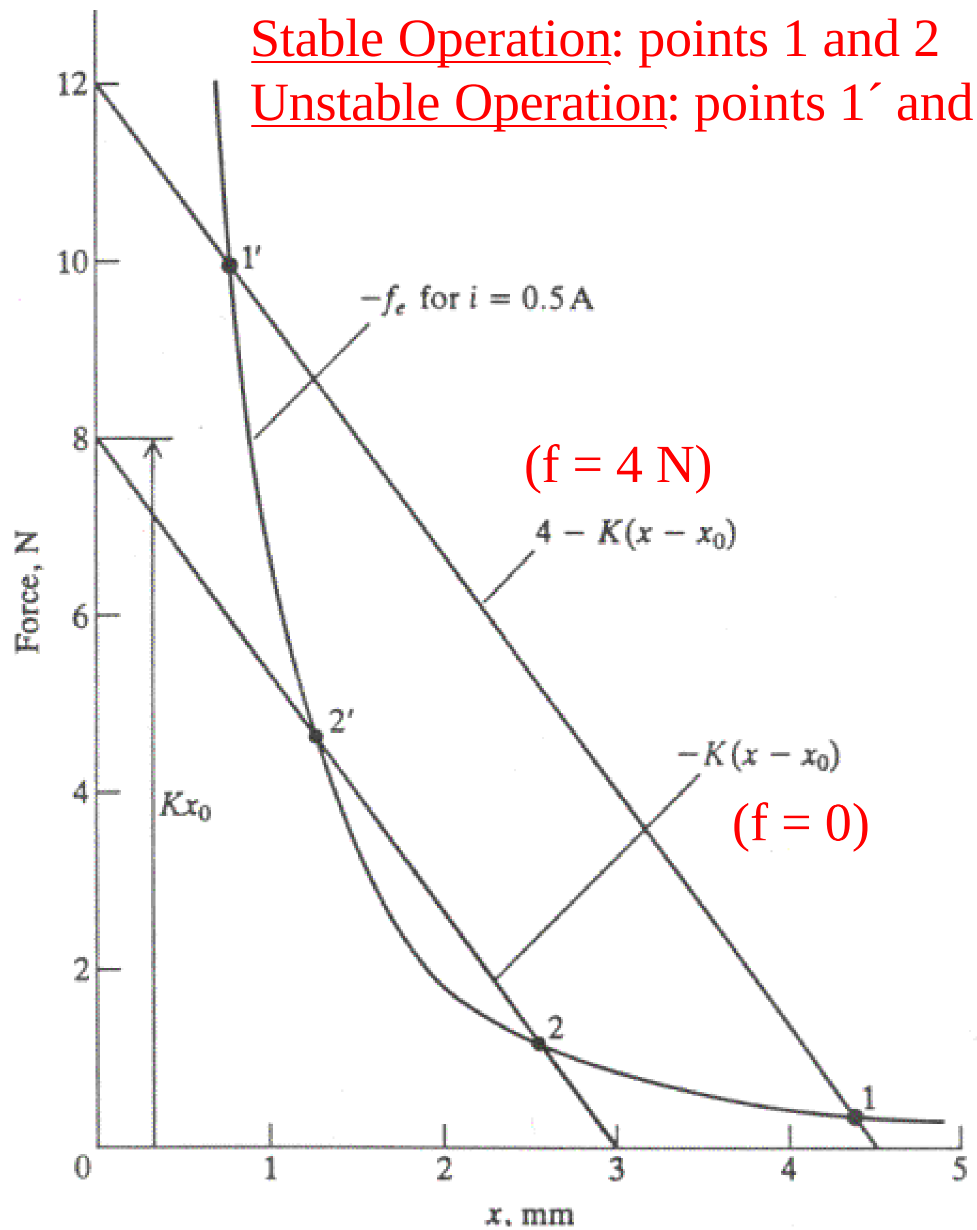
$k = 6.283\text{E-}5$ H m

$v = 5$ V

$i = 0.5$ A

Stable Operation: points 1 and 2

Unstable Operation: points 1´ and 2´



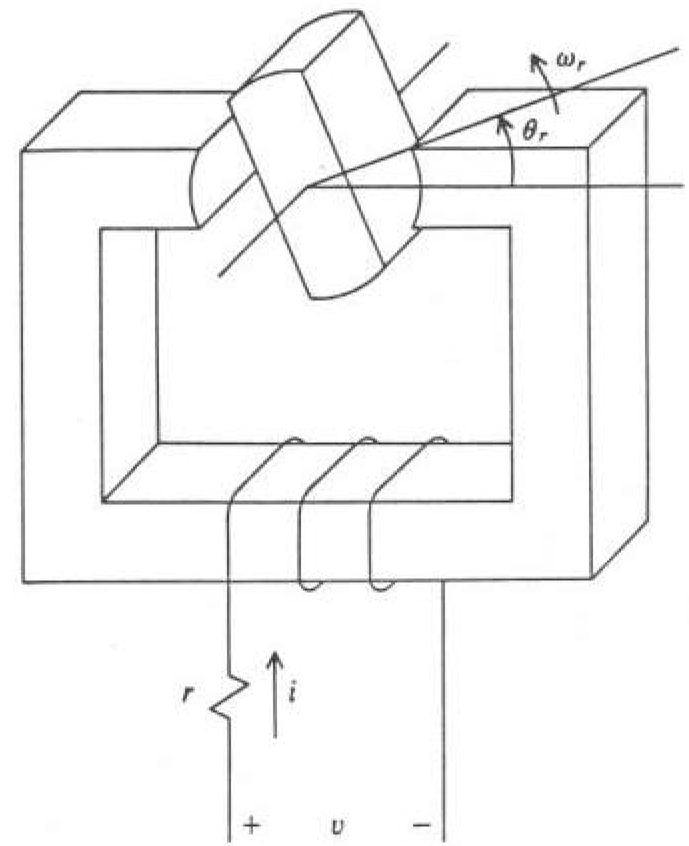
Single-Phase Reluctance Machine

- The machine consists of:
  - stationary core with a winding of N turns
  - moveable member which rotates

θ_r = angular displacement

ω_r = angular velocity

$$\theta_r = \int_0^t \omega_r(\xi)\,d\xi + \theta_r(0)$$



$$v = ri + \frac{d\lambda}{dt}$$

voltage equation

$$\phi = \phi_\ell + \phi_m$$

$\phi_\ell =$ leakage flux

$\phi_m =$ magnetizing flux

$$\lambda = \left(L_\ell + L_m\right) i$$

It is convenient to express the flux linkages as the product of the sum of the leakage inductance and the magnetizing inductance and the current in the winding.

$L_\ell =$ constant (independent of θ_r)

$L_m =$ periodic function of θ_r

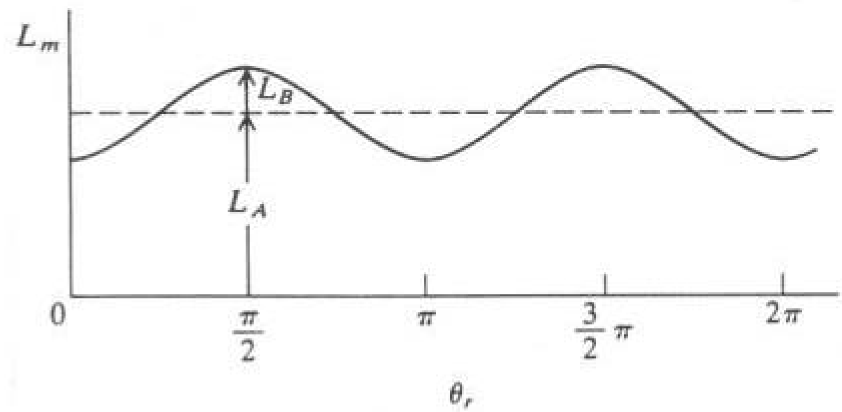
$$L_m = L_m(\theta_r)$$

$$L_m(0) = \frac{N^2}{\Re_m(0)} \quad \Rightarrow \quad \begin{cases} \Re_m \text{ is maximum} \\ L_m \text{ is minimum} \end{cases}$$

$$L_m\left(\frac{\pi}{2}\right) = \frac{N^2}{\Re_m\left(\frac{\pi}{2}\right)} \quad \Rightarrow \quad \begin{cases} \Re_m \text{ is minimum} \\ L_m \text{ is maximum} \end{cases}$$

The magnetizing inductance varies between maximum and minimum positive values twice per revolution of the rotating member.

Assume that this variation may be adequately approximated by a sinusoidal function.



$$L_m(\theta_r) = L_A - L_B \cos(2\theta_r)$$

$$L_m(0) = L_A - L_B$$

$$L_m\left(\frac{\pi}{2}\right) = L_A + L_B$$

$$L_A > L_B$$

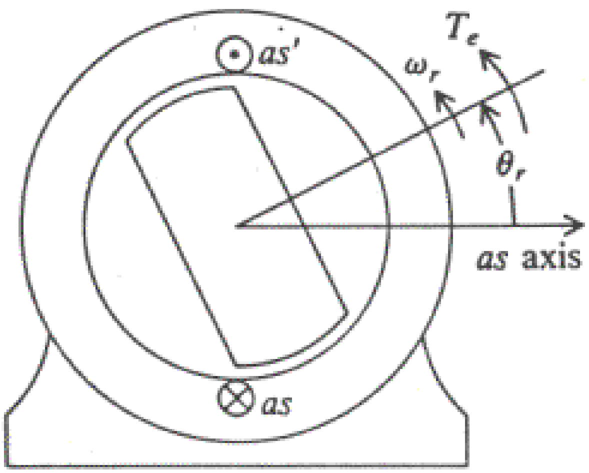
$$L_A = \text{ average value}$$

$$L(\theta_r) = L_\ell + L_m(\theta_r)$$
$$= L_\ell + L_A - L_B \cos(2\theta_r)$$

$$v = ri + \left[L_\ell + L_m(\theta_r)\right]\frac{di}{dt} + i\frac{dL_m(\theta_r)}{d\theta_r}\frac{d\theta_r}{dt}$$

voltage equation

- This elementary two-pole single-phase reluctance machine is shown in a slightly different form. Winding 1 is now winding *as* and the stator has been changed to depict more accurately the configuration common for this device.



$$v_{as} = r_s i_{as} + \frac{d\lambda_{as}}{dt}$$

$$\lambda_{as} = L_{asas} i_{as}$$

$$L_{asas} = L_{\ell s} + L_A - L_B \cos(2\theta_r)$$

$$\theta_r = \int_0^t \omega_r(\xi)d\xi + \theta_r(0)$$

r_s = resistance of *as* winding

L_{asas} = self-inductance of *as* winding

$L_{\ell s}$ = leakage inductance

- Electromagnetic torque:
  - Magnetic system is linear, hence $W_f = W_c$.

$$W_c\left(i_{as},\theta_r\right)=\frac{1}{2}\left(L_{\ell s}+L_A-L_B\cos\left(2\theta_r\right)\right)i_{as}^2$$

$$T_e\left(\vec{i},\theta\right)=\sum_{j=1}^{J}\left[i_j\frac{\partial\lambda_j\left(\vec{i},\theta\right)}{\partial\theta}\right]-\frac{\partial W_f\left(\vec{i},\theta\right)}{\partial\theta}$$

$$T_e\left(\vec{i},\theta\right)=\frac{\partial W_c\left(\vec{i},\theta\right)}{\partial\theta}$$

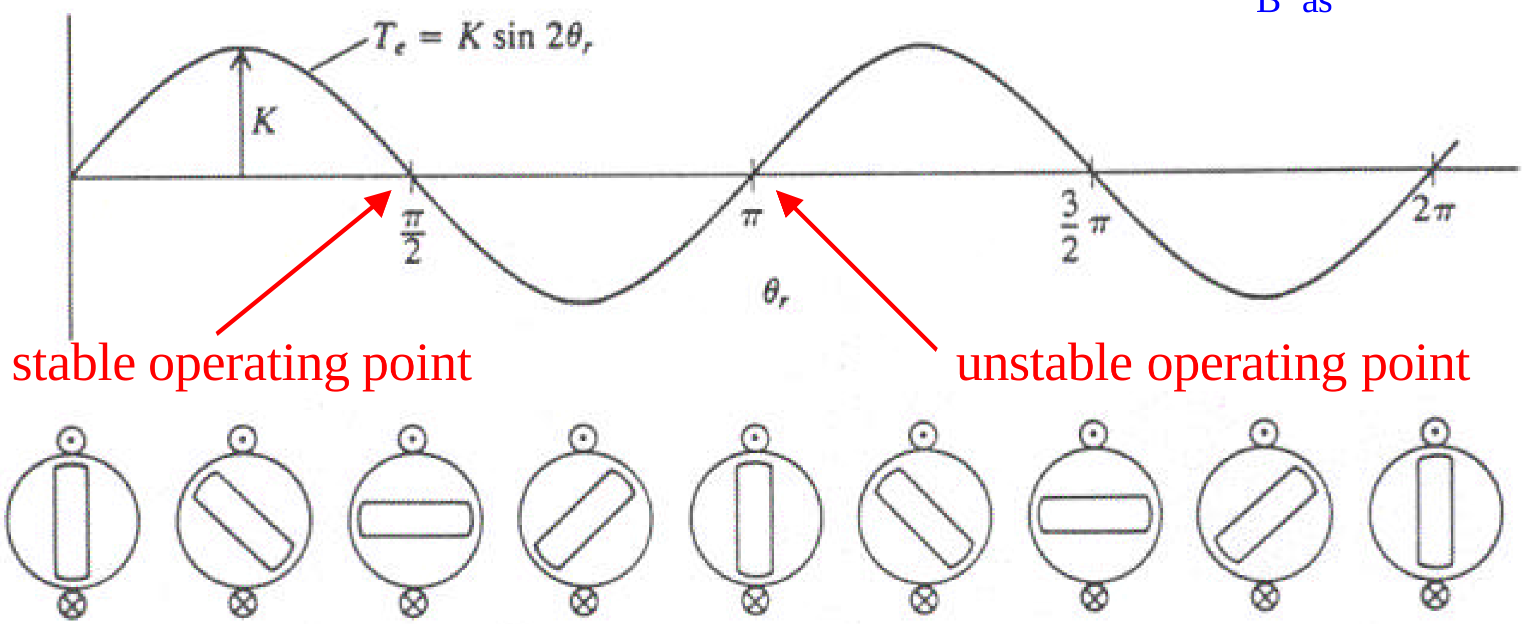
$$T_e\left(i_{as},\theta_r\right)=L_B i_{as}^2\sin\left(2\theta_r\right)$$

Valid for both transient and steady-state operation

- Consider steady-state operation: i_{as} is constant

$$T_e = K \sin(2\theta_r)$$

$$K = L_B i_{as}^2$$



Electromagnetic torque versus angular displacement of a single-phase reluctance machine with constant stator current

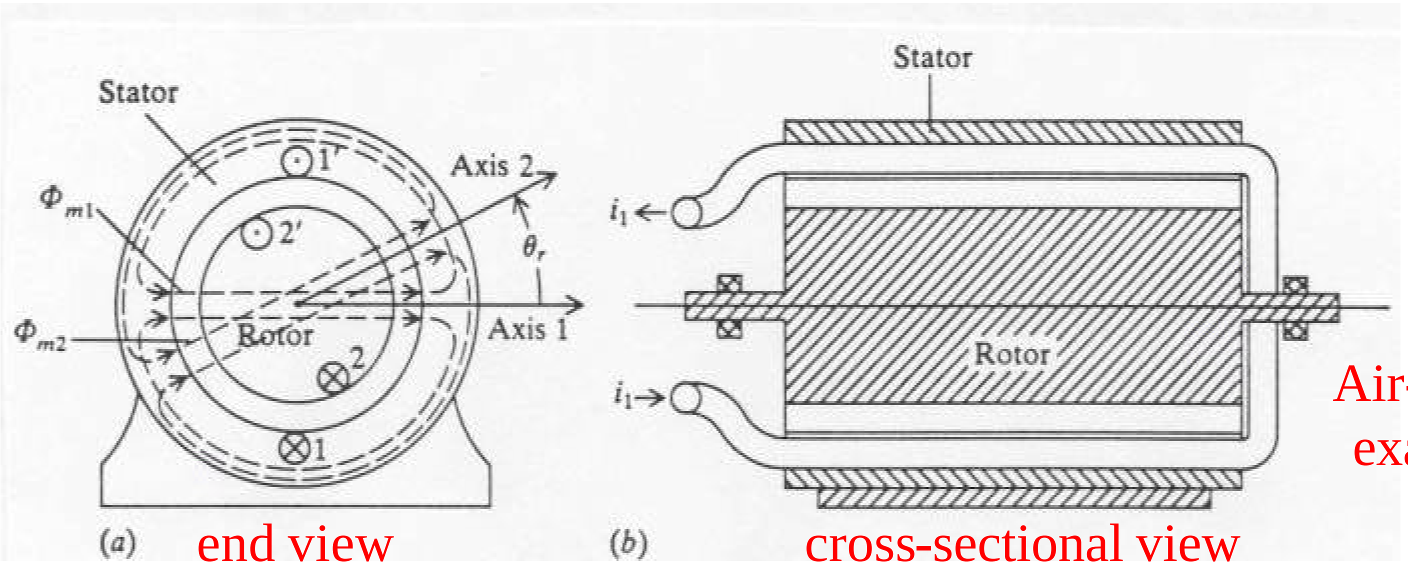
- Although the operation of a single-phase reluctance machine with a constant current is impracticable, it provides a basic understanding of reluctance torque, which is the operating principle of variable-reluctance stepper motors.
- In its simplest form, a variable-reluctance stepper motor consists of three cascaded, single-phase reluctance motors with rotors on a common shaft and arranged so that their minimum reluctance paths are displaced from each other.

Windings in Relative Motion

- The rotational device shown will be used to illustrate windings in relative motion.

Winding 1: N_1 turns on stator
Winding 2: N_2 turns on rotor

Assume that the turns are concentrated in one position.



end view

cross-sectional view

Air-gap size is exaggerated.

$$v_1 = r_1 i_1 + \frac{d\lambda_1}{dt}$$

$$v_2 = r_2 i_2 + \frac{d\lambda_2}{dt}$$

voltage equations

$$\lambda_1 = L_{11} i_i + L_{12} i_2$$

$$\lambda_2 = L_{21} i_i + L_{22} i_2$$

The magnetic system is assumed linear.

$$L_{11} = L_{\ell 1} + L_{m1} = \frac{N_1^2}{\Re_{\ell 1}} + \frac{N_1^2}{\Re_m}$$

$$L_{22} = L_{\ell 2} + L_{m2} = \frac{N_2^2}{\Re_{\ell 2}} + \frac{N_2^2}{\Re_m}$$

The self-inductances L_{11} and L_{22} are constants and may be expressed in terms of leakage and magnetizing inductances.

$\Re_m$ is the reluctance of the complete magnetic path of φ_{m1} and φ_{m2}, which is through the rotor and stator iron and twice across the air gap.

Let's now consider L_{12}.

$$\theta_r = \text{ angular displacement}$$

$$\omega_r = \text{ angular velocity}$$

$$\theta_r = \int_0^t \omega_r(\xi)\, d\xi + \theta_r(0)$$

When θ_r is zero, then the coupling between windings 1 and 2 is maximum. The magnetic system of winding 1 aids that of winding 2 with positive currents assumed. Hence the mutual inductance is positive.

$$L_{12}(0) = \frac{N_1 N_2}{\Re_m}$$

When θ_r is $\pi/2$, the windings are orthogonal. The mutual coupling is zero.

$$L_{12}\left(\frac{\pi}{2}\right) = 0$$

Assume that the mutual inductance may be adequately predicted by:

$$\begin{cases} L_{12}(\theta_r) = L_{sr}\cos(\theta_r) \\ L_{sr} = \dfrac{N_1 N_2}{\Re_m} \end{cases}$$

$$v_1 = r_1 i_1 + \frac{d\lambda_1}{dt}$$
$$v_2 = r_2 i_2 + \frac{d\lambda_2}{dt}$$

L_{sr} is the amplitude of the sinusoidal mutual inductance between the stator and rotor windings.

In writing the voltage equations, the total derivative of the flux linkages is required.

$$\lambda_1 = L_{11} i_1 + (L_{sr}\cos\theta_r) i_2$$
$$\lambda_2 = L_{22} i_2 + (L_{sr}\cos\theta_r) i_1$$

$$v_1 = r_1 i_1 + L_{11}\frac{di_1}{dt} + L_{sr}\cos\theta_r \frac{di_2}{dt} - i_2 \omega_r L_{sr}\sin\theta_r$$
$$v_2 = r_2 i_2 + L_{22}\frac{di_2}{dt} + L_{sr}\cos\theta_r \frac{di_1}{dt} - i_1 \omega_r L_{sr}\sin\theta_r$$

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} r_1 & 0 \\ 0 & r_2 \end{bmatrix} \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} + \frac{d}{dt} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}$$

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} L_{\ell 1} + L_{m1} & L_{sr} \cos\theta_r \\ L_{sr} \cos\theta_r & L_{\ell 2} + L_{m2} \end{bmatrix} \begin{bmatrix} i_{as} \\ i_{bs} \end{bmatrix}$$

Since the magnetic system is assumed to be linear:

$$W_f\left(i_1, i_2, \theta_r\right) = \frac{1}{2} L_{11} i_1^2 + L_{12} i_1 i_2 + \frac{1}{2} L_{22} i_2^2 = W_c\left(i_1, i_2, \theta_r\right)$$

$$T_e\left(\vec{i}, \theta\right) = \sum_{j=1}^{J} \left[i_j \frac{\partial \lambda_j\left(\vec{i}, \theta\right)}{\partial \theta} \right] - \frac{\partial W_f\left(\vec{i}, \theta\right)}{\partial \theta}$$

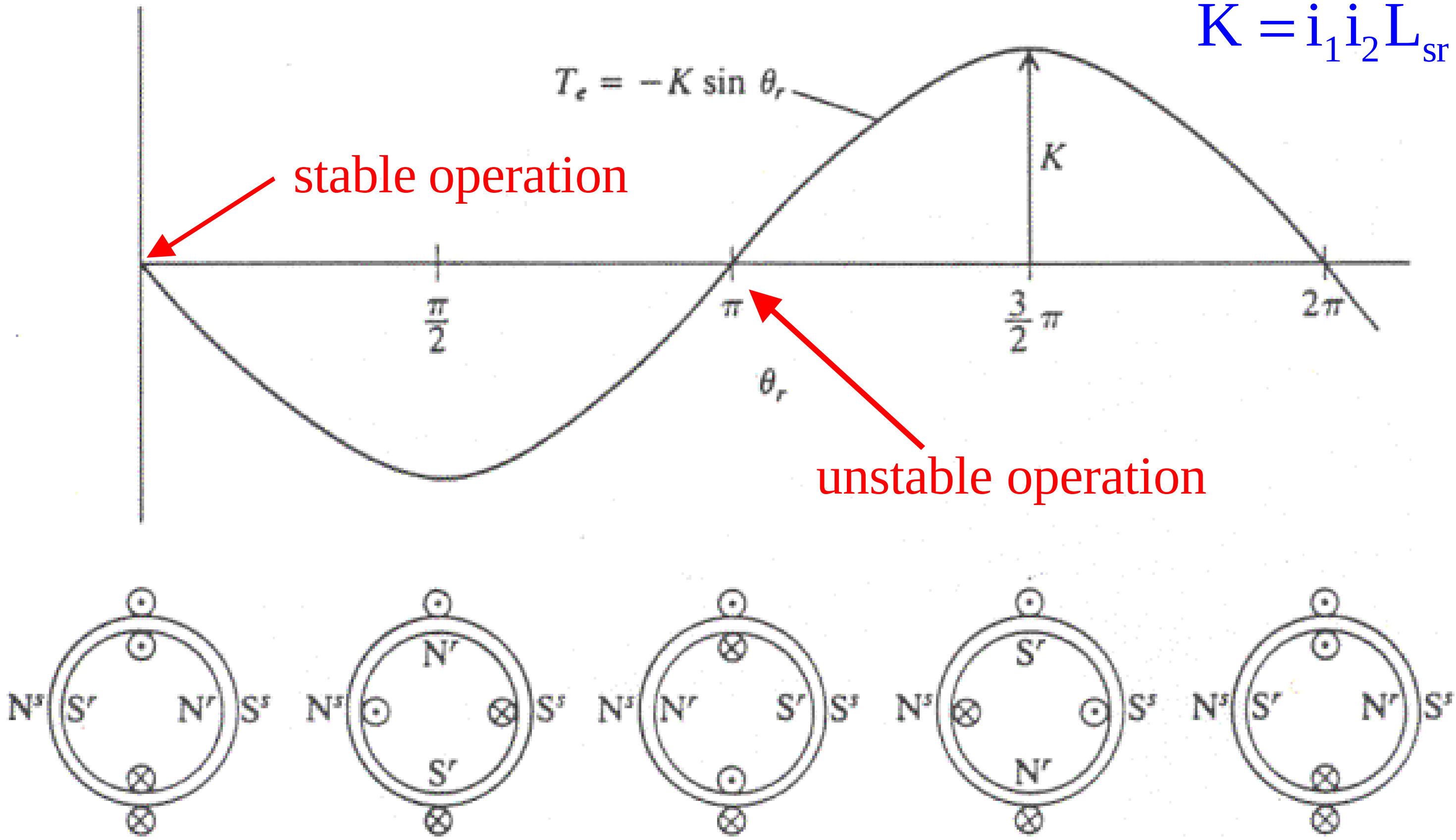
$$T_e\left(\vec{i}, \theta\right) = \frac{\partial W_c\left(\vec{i}, \theta\right)}{\partial \theta}$$

$$T_e\left(i_1, i_2, \theta_r\right) = -i_1 i_2 L_{sr} \sin\theta_r$$

- Consider the case where i_1 and i_2 are both positive and constant:

$$T_e = -K \sin \theta_r$$

$$K = i_1 i_2 L_{sr}$$



Electromagnetic torque versus angular displacement with constant winding currents

- Although operation with constant winding currents is somewhat impracticable, it does illustrate the principle of positioning of stepper motors with a permanent-magnet rotor which, in many respects, is analogous to holding i_2 constant on the elementary device considered here.

Sensors & Actuators in Mechatronics

MEAE 6960
Summer 2002

Assignment # 2

Problem # 1

(A) A resistor and an inductor are connected in series with R = 15Ω and L = 250 mH. Express the energy stored in the inductor and the energy dissipated by the resistor for $t > 0$ if $i(0) = 10$ A.

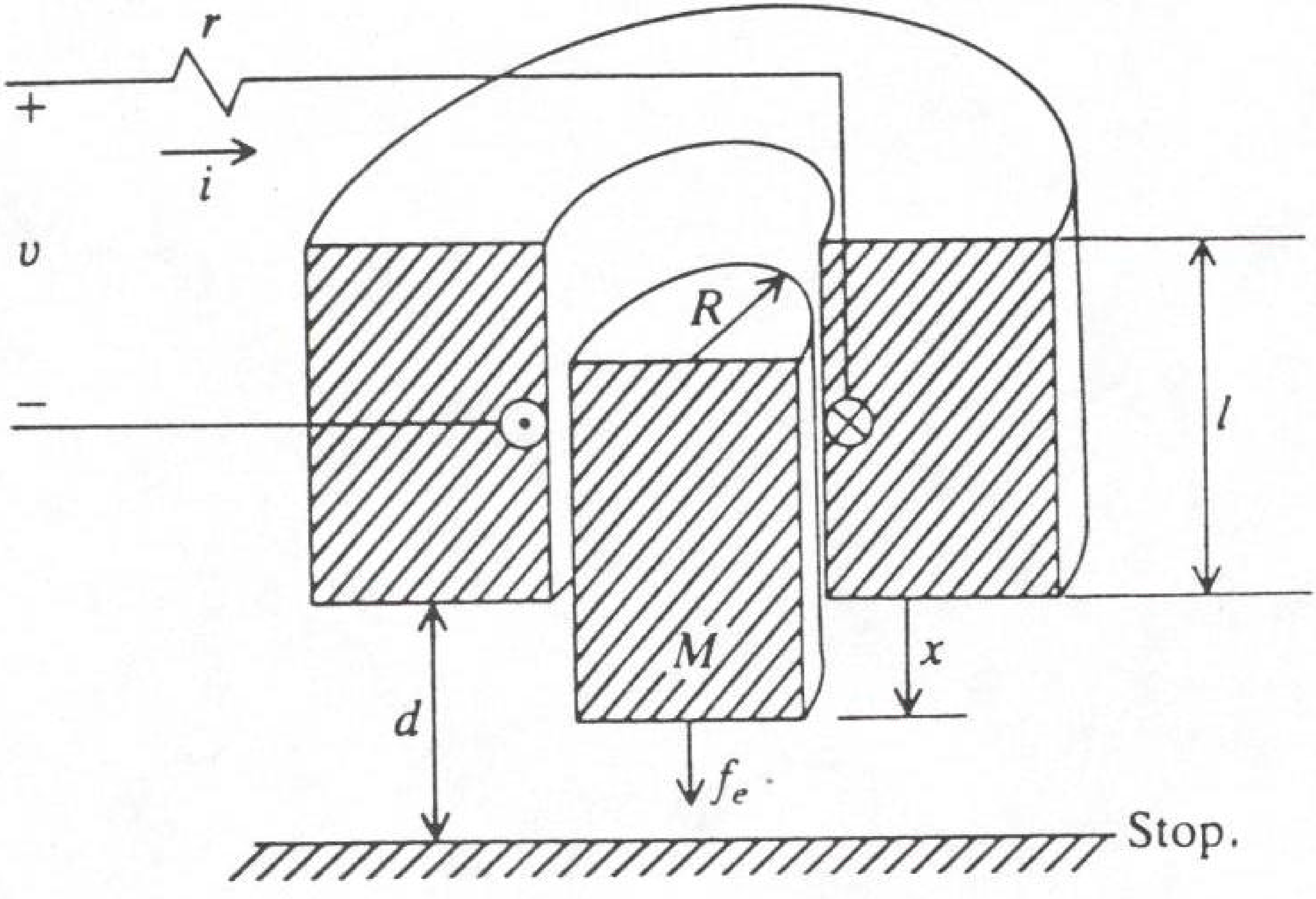
(B) Consider a simple horizontal spring-mass-damper system. At $t = 0$, $x(0) = x_0$ (rest position), and $dx/dt = 1.5$ m/s. Also, M = 0.8 kg, B = 10 Ns/m, and K = 120 N/m. For $t > 0$, express the energy stored in the spring, the kinetic energy of the mass, and the energy dissipated by the damper. You need not evaluate the integral expression for the energy dissipated by the damper.

Problem # 2

The plunger with mass *M* is free to move within an electromagnet. Although the winding of the electromagnet consists of many turns, only one is shown. The mechanical viscous damping coefficient B varies directly as the surface area of the plunger within the electromagnet. (a) Write the voltage equation for the electric system; (b) Write the dynamic equation for the mechanical system. Include the force of gravity; (c) Express the mechanical damping coefficient B; (d) Express the electromagnetic force f_e if

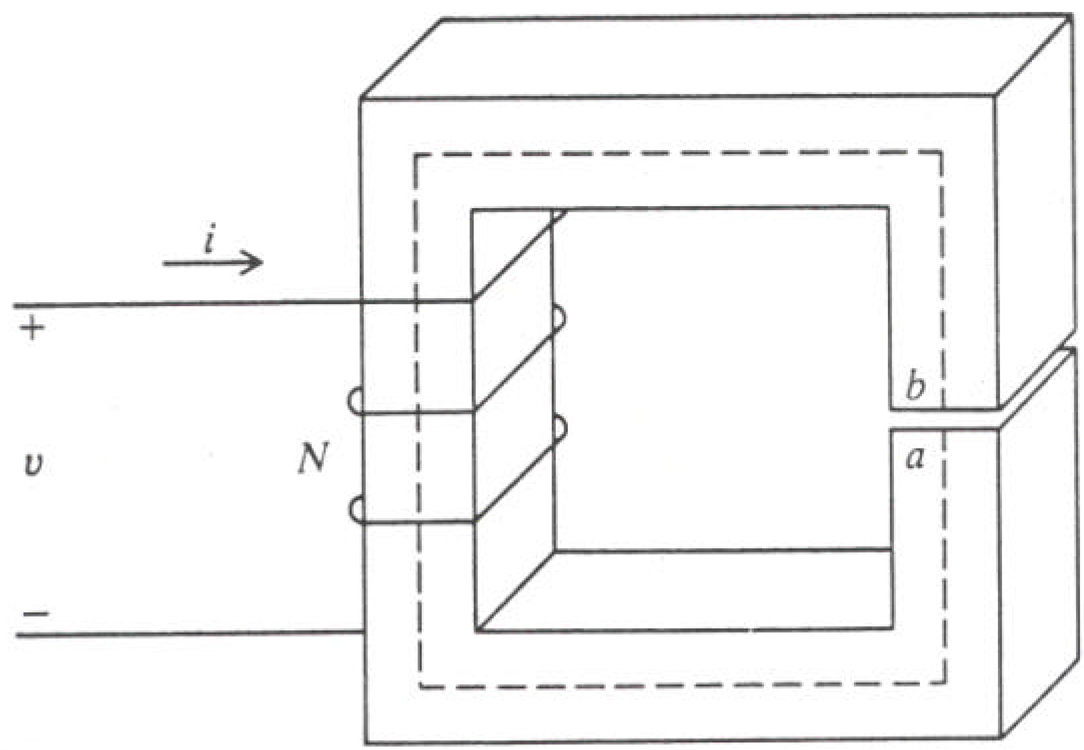
$$i = a\lambda^2 + b\lambda(x-d)^2$$

(e) Express the steady-state position x for a constant current flowing in the winding.


r
+
i
v
R
−
l
M
x
d
fe
Stop.

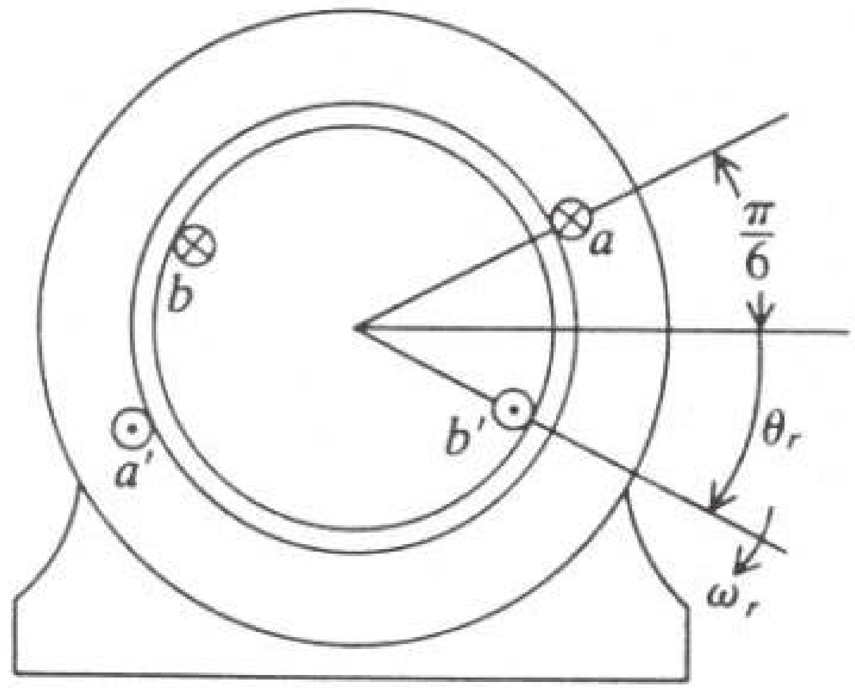
Problem #3

Express the force of attraction between the iron faces of the air gap in the figure shown in terms of N, the turns of the winding; i, the current flowing in the winding; m_0; A_g, the cross-sectional area of the air gap; and g, the air gap length. Neglect the reluctance of the iron.



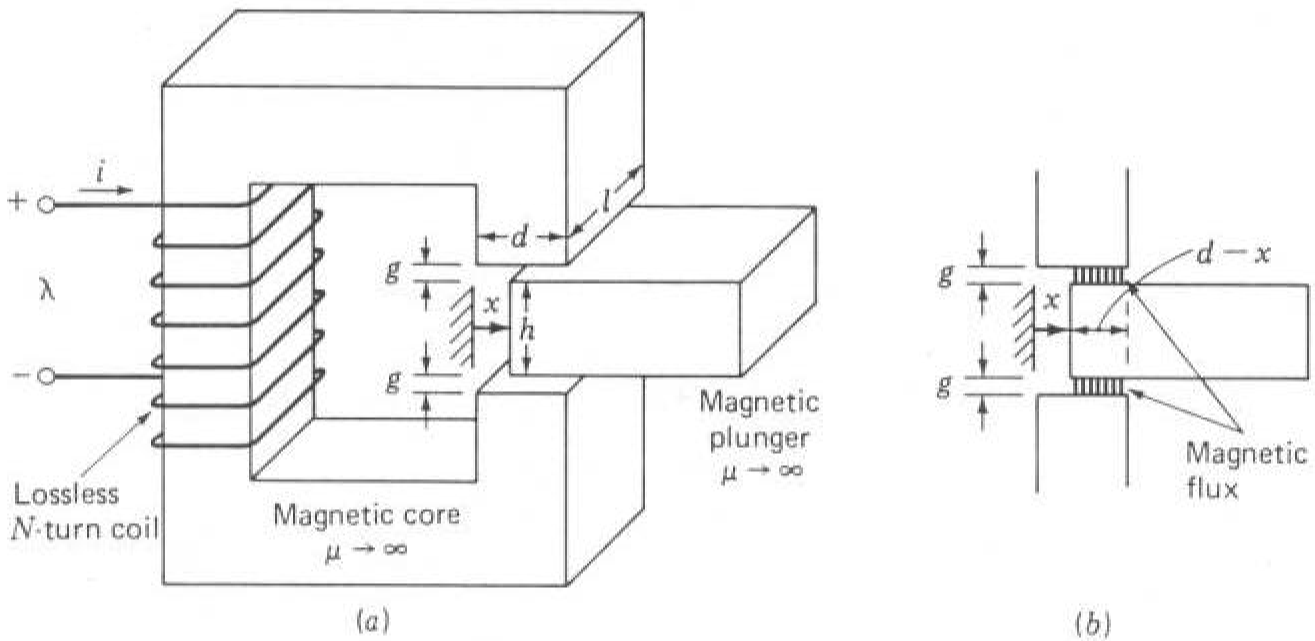
Problem #4

A two-winding device with clockwise rotation is shown. θ_r and ω_r are positive in the clockwise direction. The peak amplitude of the mutual inductance is L_{sr}. Express (a) the mutual inductance L_{ab} and (b) the electromagnetic torque T_e.



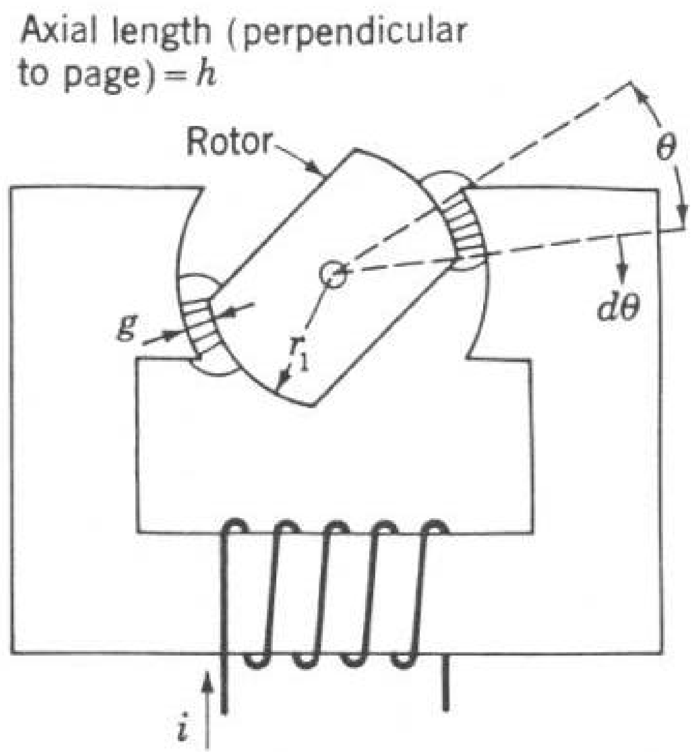
Problem # 5

The relay shown is made from infinitely permeable magnetic material with a movable plunger, also of infinitely permeable material. The height of the plunger is much greater than the air-gap length ($h >> g$). (a) Calculate the magnetic stored energy as a function of plunger position ($0 < x < d$) for N = 1000 turns, g = 0.002 m, d = 0.15 m, l = 0.1 m, and i = 10 A. (b) Find the force on the plunger as a function of x when the coil current is held constant at 10 A.



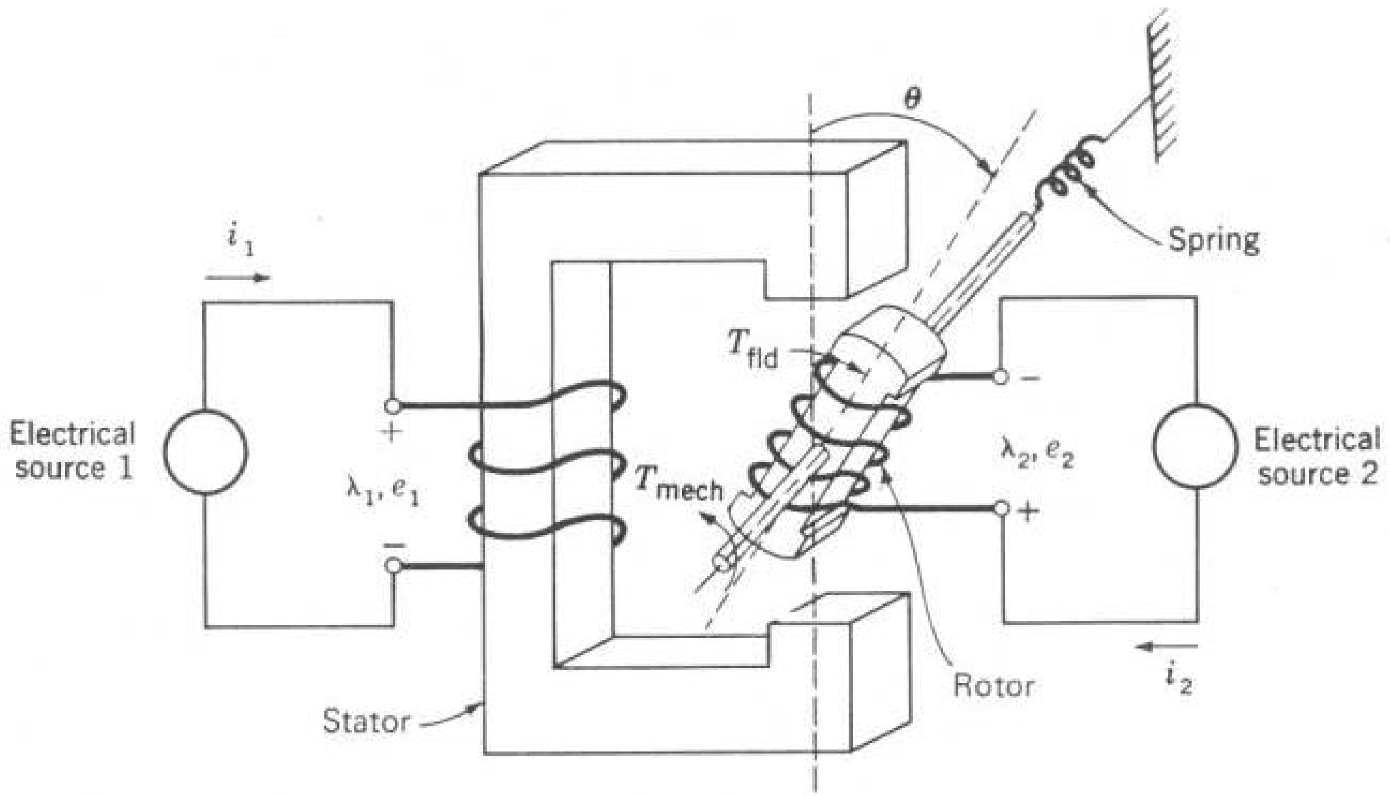
Problem # 6

The magnetic circuit shown is made of cast steel. The rotor is free to turn about a vertical axis. The dimensions are shown in the figure. Derive an expression for the torque acting on the rotor in terms of the dimensions and the magnetic field in the two air gaps. Neglect the effects of fringing.



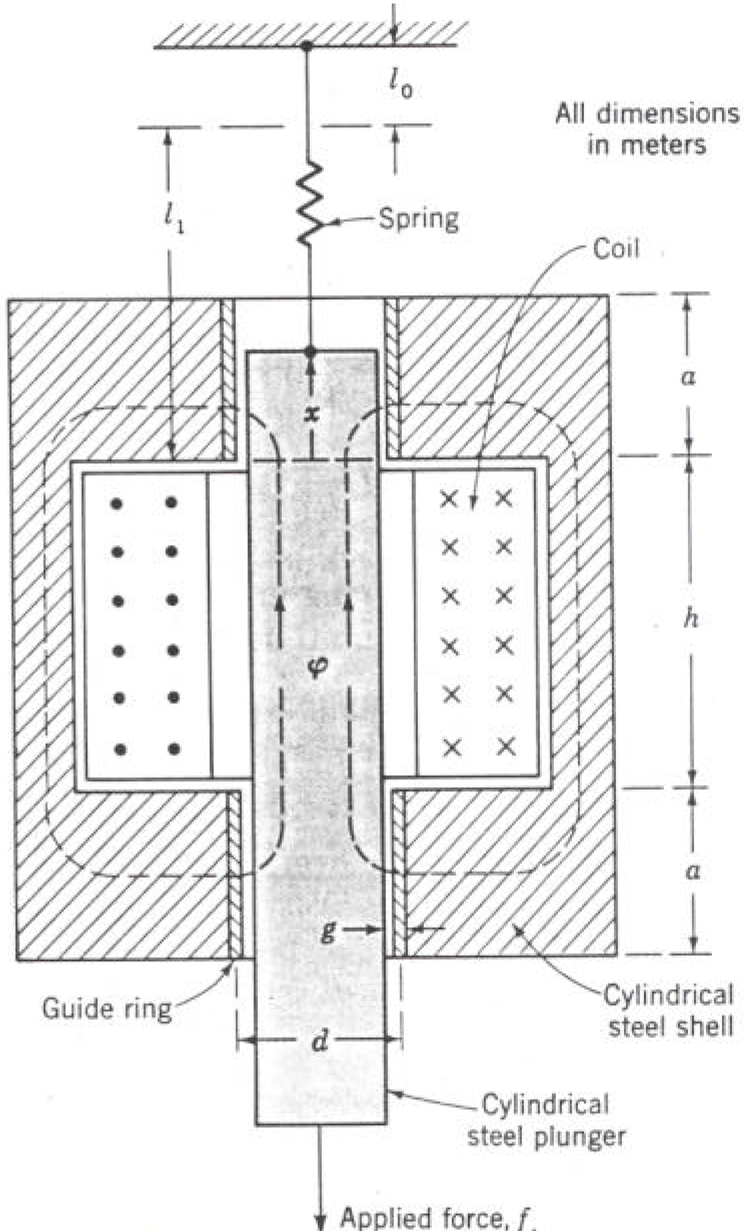
Problem # 7

In the system shown, the inductances in henrys are given as $L_{11} = (3 + \cos 2\theta) \times 10^{-3}$; $L_{12} = 0.1 \cos \theta$; $L_{22} = 30 + 10 \cos 2\theta$. Find the torque for current $i_1 = 1$ A and $i_2 = 0.01$ A.



Problem # 8

The figure shows in cross section a cylindrical solenoid magnet in which the cylindrical plunger of mass M moves vertically in brass guide rings of thickness g and mean diameter d. The permeability of brass is the same as that of free space. The plunger is supported by a spring whose spring constant is K. Its unstretched length is l_0. A mechanical load force f_t is applied to the plunger from the mechanical system connected to it. Assume that the frictional force is linearly proportional to the velocity and that the damping coefficient is B. The coil has N turns and resistance R. Its terminal voltage is v_t, and its current is i. The effects of magnetic leakage and reluctance of the steel are negligible. Derive the dynamic equations of motion of the electromechanical system.



l_0
All dimensions in meters
l_1
Spring
Coil
a
x
h
φ
a
g
Guide ring
d
Cylindrical steel shell
Cylindrical steel plunger
Applied force, f_t

Direct-Current Motors

- Introduction
- Elementary Direct-Current Machine
- Review: Windings in Relative Motion
- Voltage and Torque Equations
- Basic Types of Direct-Current Machines
  - Separate Winding Excitation (includes permanent magnet)
  - Shunt-Connected DC Machine
  - Series-Connected DC Machine
  - Compound-Connected DC Machine

- Time-Domain Block Diagrams, State Equations, and Transfer Functions
- Elementary Approach to Permanent-Magnet DC Motor Modeling
- Control of DC Motors
- Geared Systems, Optimum Gear Ratios, and Motor Selection
- Motor Selection Considerations

Introduction

- The Brushed DC motor is not as widely used today as in the past, but it is still being used, especially at the low-power level.
- Focus is on topics of interest to the mechatronics engineer
  - Shunt-connected dc motor
  - Permanent-magnet dc motor

  } similar operating characteristics
- A simplified method of analysis is used rather than an analysis wherein commutation is treated in detail. As a result the dc motor is the most straightforward to analyze of all the electromechanical devices.

Elementary Direct-Current Machine

- It is instructive to discuss the elementary two-pole dc machine prior to a formal analysis of the performance of a practical dc machine.
- The two-pole elementary machine is equipped with:
  - Field winding wound on stator poles
  - Rotor (or armature) coil
  - Commutator
    - Two semicircular copper segments mounted on the shaft at the end of the rotor and insulated from one another as well as from the iron of the rotor.
    - Each terminal of the rotor coil is connected to a copper segment.

- Stationary carbon brushes ride upon the copper segments whereby the rotor coil is connected to a stationary circuit by a near frictionless contact.

- The voltage equations for the field winding and rotor coil are:

$$v_f = r_f i_f + \frac{d\lambda_f}{dt}$$

$$v_{a-a'} = r_a i_{a-a'} + \frac{d\lambda_{a-a'}}{dt}$$

- The flux linkages can be expressed as:

$$\lambda_f = L_{ff} i_f + L_{fa} i_{a-a'}$$

$$\lambda_{a-a'} = L_{aa} i_{a-a'} + L_{af} i_f$$

- As a first approximation, the mutual inductance between the field winding and an armature coil is:

$$L_{af} = L_{fa} = -L\cos\theta_r \qquad \text{(L is a constant)}$$

- Commutation
  - As the rotor revolves, the action of the commutator is to switch the stationary terminals from one terminal of the rotor coil to the other. This switching occurs at: $\theta_r = 0, \pi, 2\pi, \ldots.$
  - At the instant of switching, each brush is in contact with both copper segments whereupon the rotor coil is short-circuited. It is desirable to commutate (short-circuit) the rotor coil at the instant the induced voltage is a minimum.
  - The waveform of the voltage induced in the open-circuited armature coil, during constant-speed operation with a constant field winding current, may be determined by setting $i_{a-a'} = 0$ and i_f = constant.

– Substitution:

$$\left.\begin{aligned}\lambda_{a-a'} &= L_{aa} i_{a-a'} + L_{af} i_f \\ L_{af} &= L_{fa} = -L\cos\theta_r\end{aligned}\right\} \Rightarrow \quad v_{a-a'} = r_a i_{a-a'} + \frac{d\lambda_{a-a'}}{dt}$$

– Result is an expression for the open-circuit voltage of coil a-a´ with the field current i_f a constant = I_f :

$$v_{a-a'} = \frac{d\theta_r}{dt} L I_f \sin\theta_r = \omega_r L I_f \sin\theta_r$$

ω_r = rotor speed

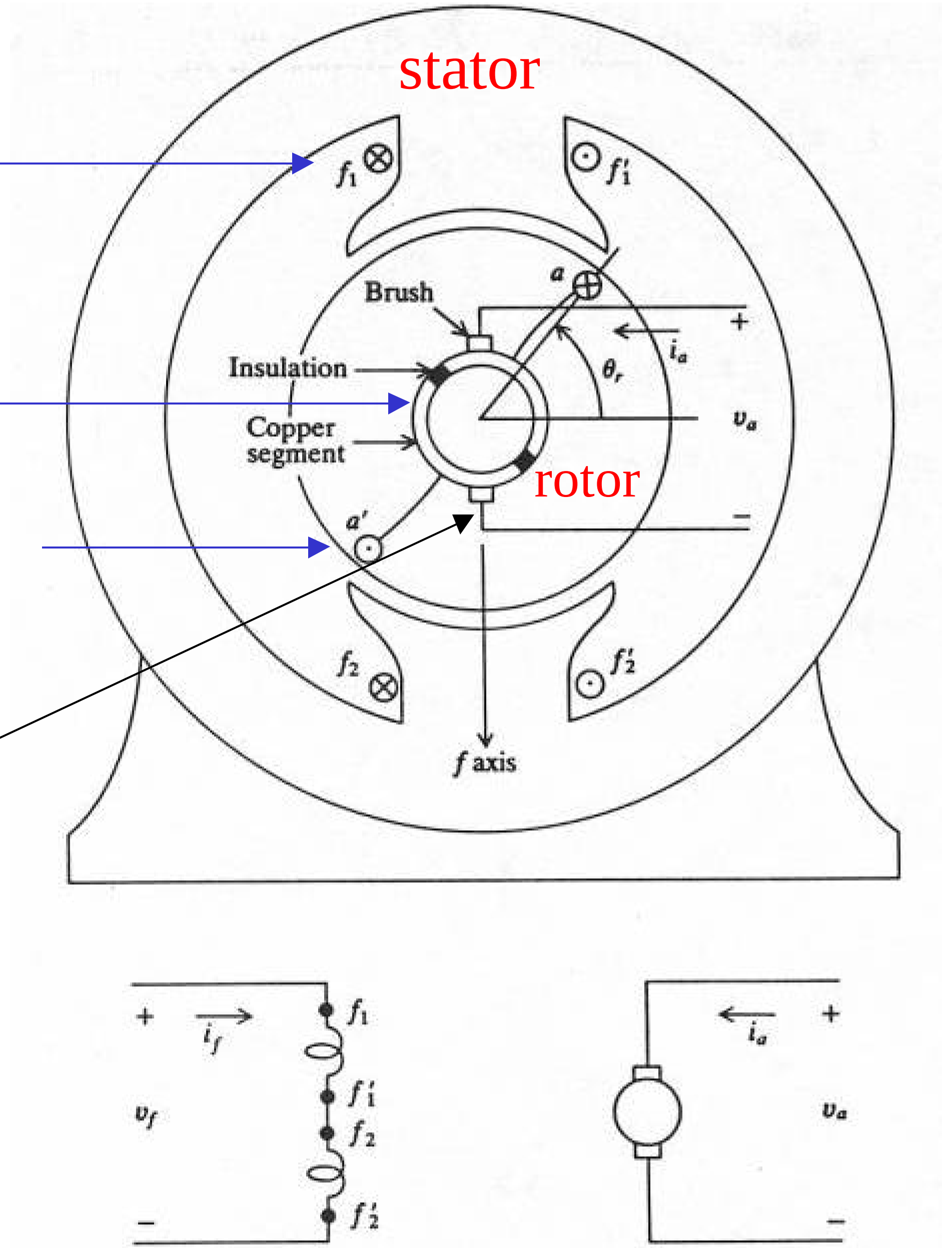
– This open-circuit coil voltage is zero at $\theta_r = 0, \pi, 2\pi, \ldots$ which is the rotor position during commutation.

field winding wound on stator poles
with resistance r_f

commutator

rotor (armature) coil
with resistance r_a

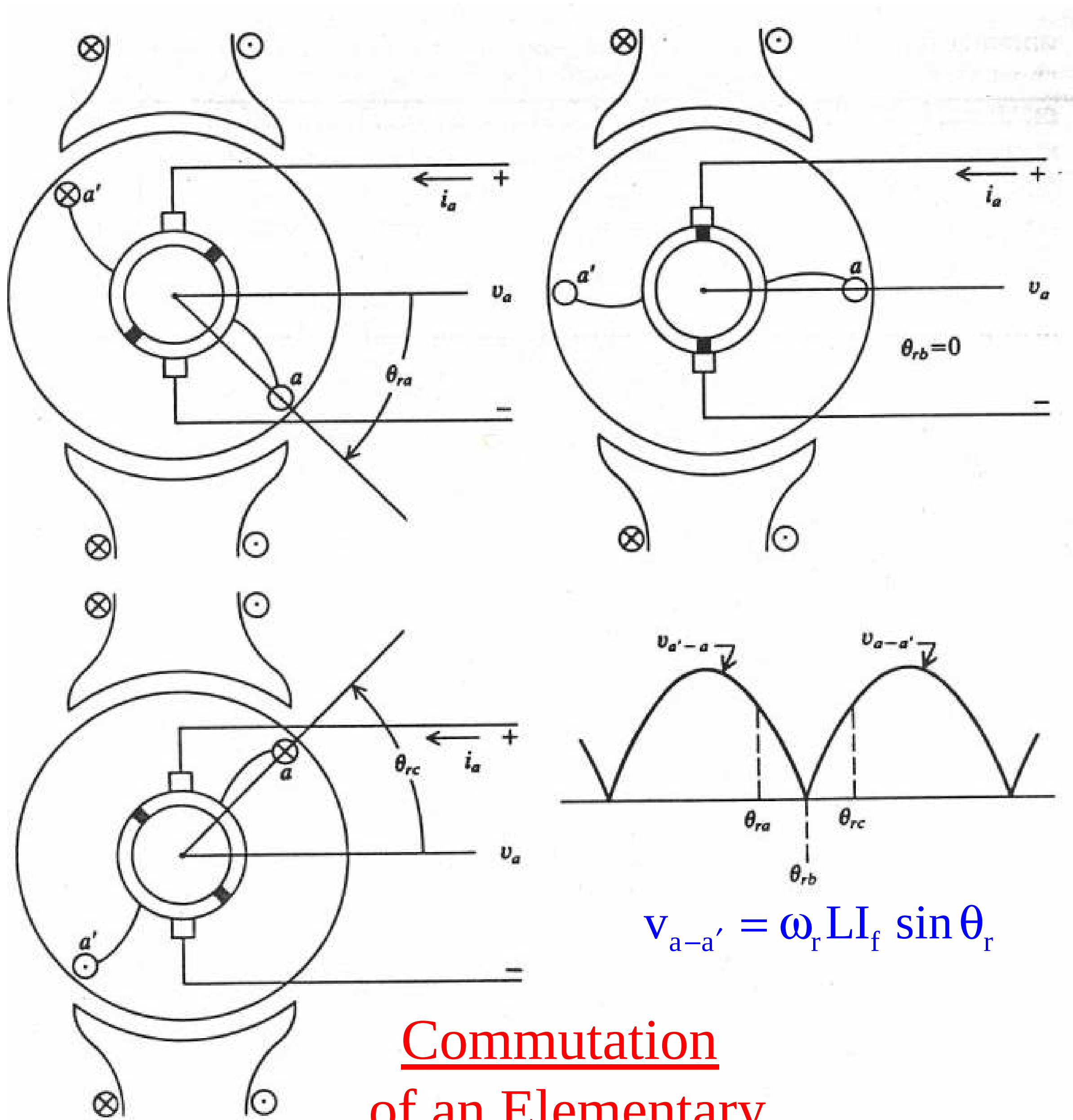
Brushes are stationary and
nearly frictionless



Elementary Two-Pole DC Machine

$\pi < \theta_r < 2\pi$
i_a positive:
down coil side *a′*
and out of coil side *a*

$0 < \theta_r < \pi$
i_a positive:
down coil side *a*
and out of coil side *a′*



$$v_{a-a'} = \omega_r L I_f \sin\theta_r$$

Commutation of an Elementary DC Machine

- The machine just analyzed is not a practicable machine. It could not be operated effectively as a motor supplied from a voltage source owing to the short-circuiting of the armature coil at each commutation.
- A practicable *dc* machine is shown with the rotor equipped with an *a* winding and an *A* winding. This requires some explanation.
- Here the parallel windings consist of only four coils. Usually the number of rotor coils is substantially more than four, thereby reducing the harmonic content of the open-circuit armature voltage.

- In this case, the rotor coils may be approximated as a uniformly distributed winding. Therein the rotor winding is considered as current sheets which are fixed in space due to the action of the commutator and which establish a magnetic axis positioned orthogonal to the magnetic axis of the field winding.

Follow the path of current through one of the parallel paths from one brush to the other.

The open-circuit or induced armature voltage is shown (plotted for one of the two parallel paths).

Positive current direction:

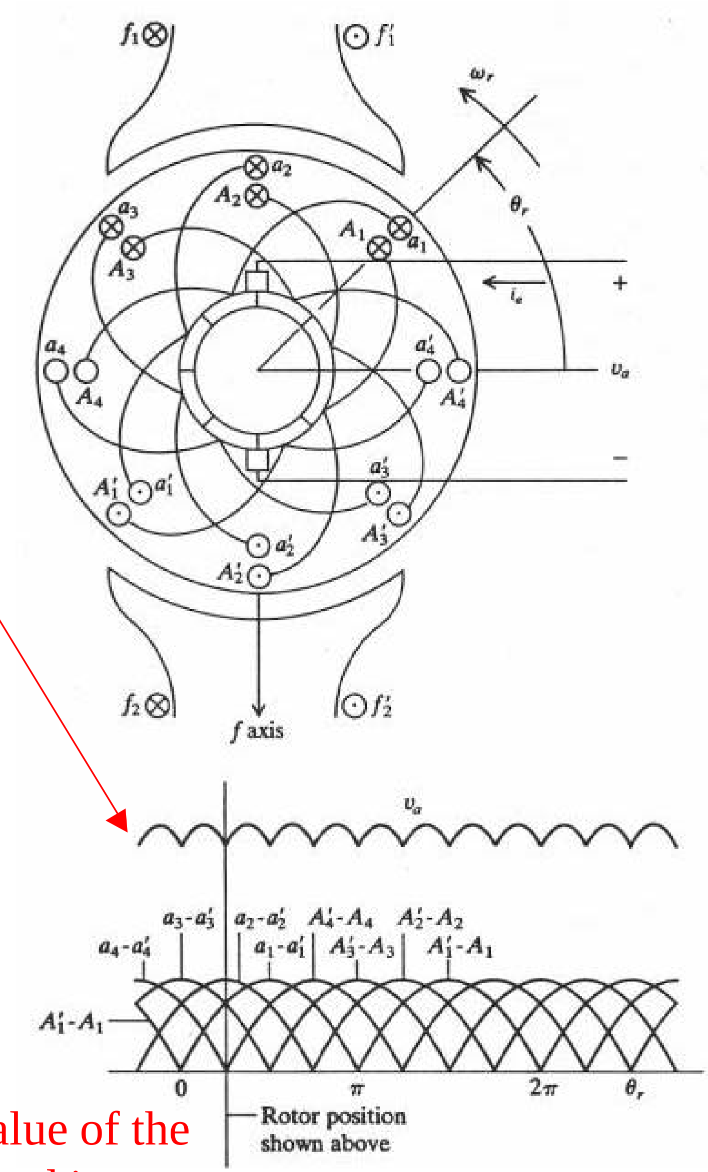
Into the paper in a_1, A_1; a_2 A_2; …

Out of the paper in a_1' A_1'; a_2' A_2'; …

Coils a_4-a_4' and A_4-A_4' are being commutated

Bottom brush short circuits a_4-a_4' coil

Top brush short circuits A_4-A_4' coil



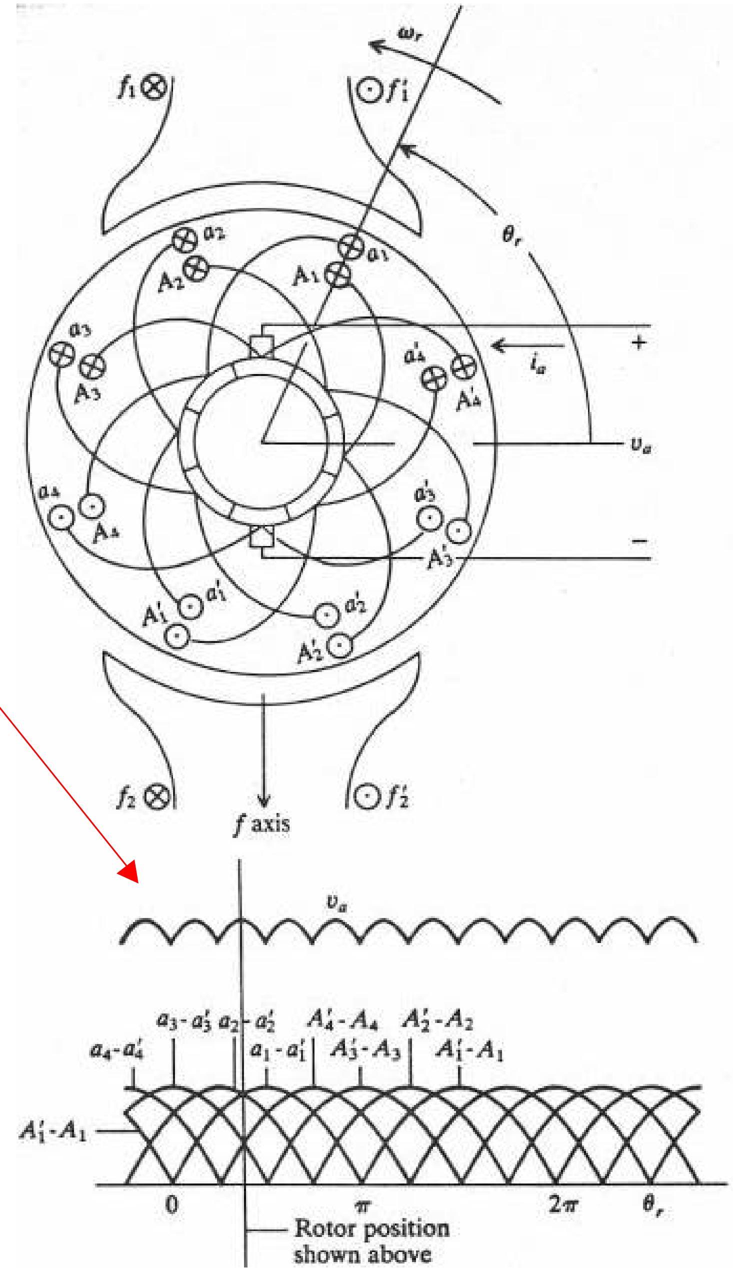
A DC Machine with Parallel Armature Windings

If the peak value of the voltage induced in one coil is 1 V, what is the max and min of v_a ?

The open-circuit or induced armature voltage is shown (plotted for one of the two parallel paths).

Top brush rides only on segment connecting A_3 and $A_4{}'$
Bottom brush rides only on segment connecting a_4 and $a_3{}'$

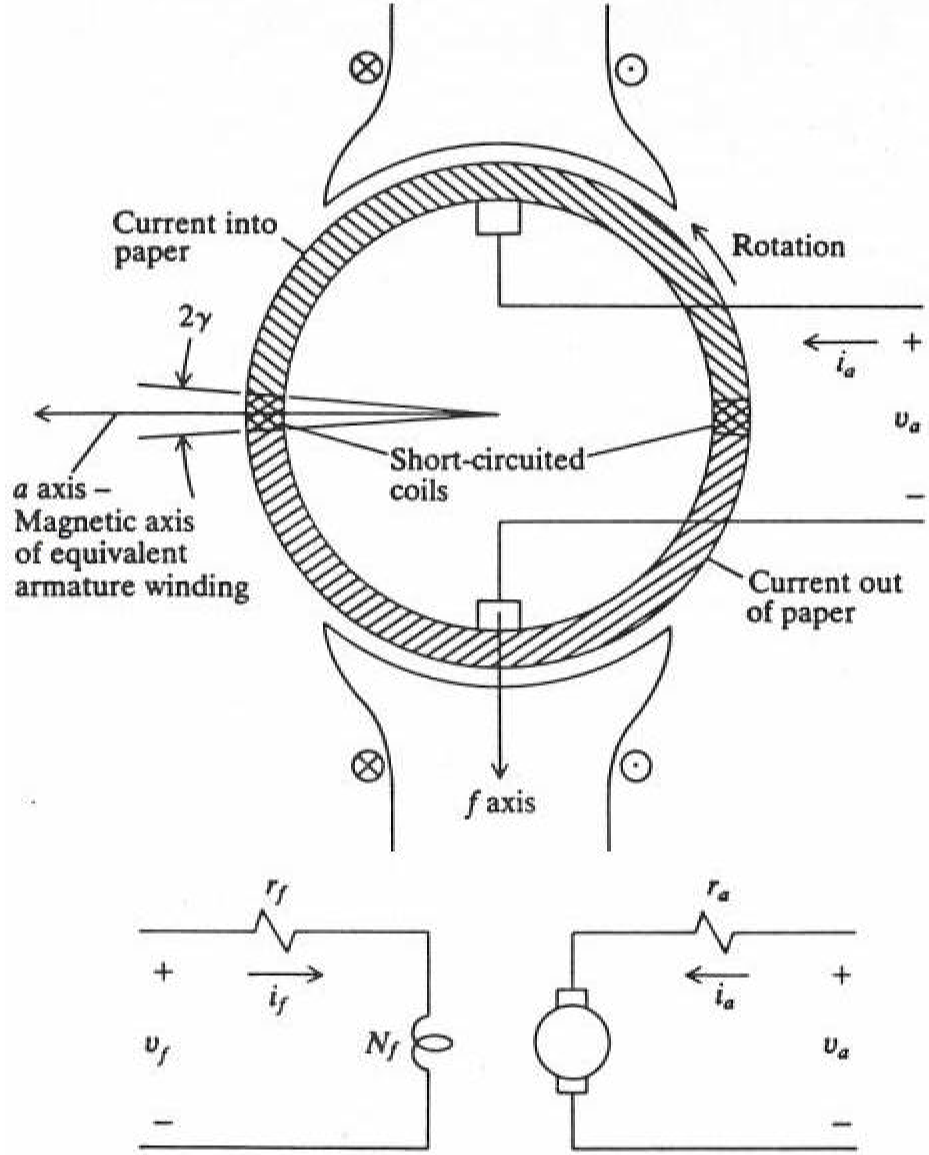
Current flows down the coils in the upper one half of the rotor and out of the coil sides in the bottom one half



A DC Machine with Parallel Armature Windings

Rotor advanced approximately 22.5º CCW

The brushes are shown depicted on the current sheet for the purpose of depicting commutation. The small angular displacement (2λ) designates the region of commutation wherein the coils are short-circuited.



Idealized DC Machine with Uniformly-Distributed Rotor Winding

- It is instructive to take a look at the arrangement of the armature windings and the method of commutation used in many of the low-power permanent-magnet dc motors.
- Small dc motors used in low-power control systems are often the permanent-magnet type, wherein a constant field flux is established by a permanent magnet rather than by a current flowing in a field winding.
- Three rotor positions of a typical low-power permanent-magnet dc motor are shown.

- Although we realize that in some cases it may be a rather crude approximation, we will consider the permanent-magnet dc motor as having current sheets on the armature with orthogonal armature and field magnetic axes.

Commutation of a Permanent-Magnet DC Motor

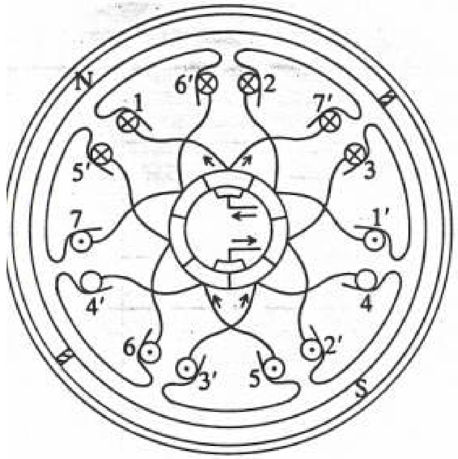
Figures:

(a) Winding 4 is being commutated

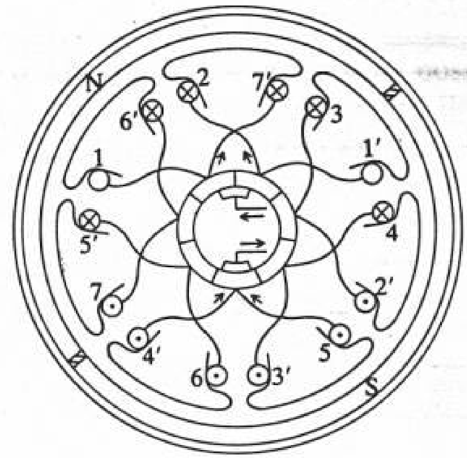
(b) Winding 1 is being commutated

(c) Winding 5 is being commutated

Rotor is turning in CCW direction.



(a)

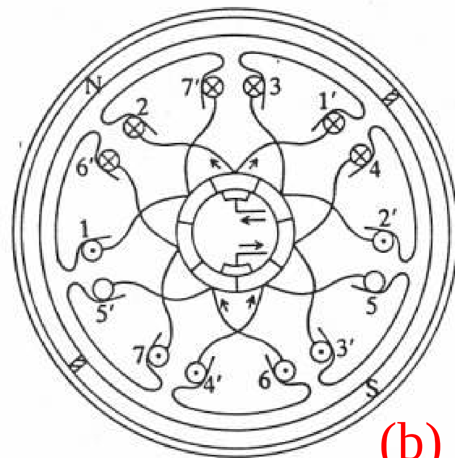


(b)

(a) → (b):
4 and 1 are being commutated

The armature windings consist of a large number of turns of fine wire, hence, each circle represents many conductors.

Note that the position of the brushes is shifted approximately 40° relative to the line drawn between the center of the N and S poles. This shift in the brushes was probably determined experimentally by minimizing brush arcing for normal load conditions.



(c)

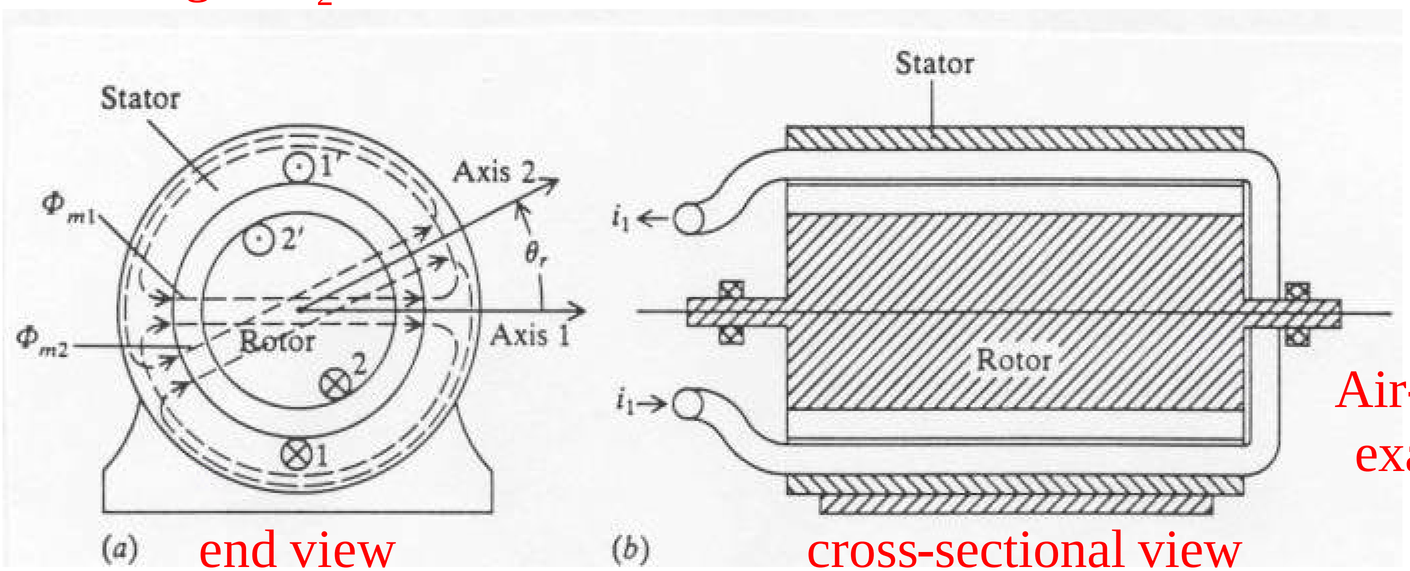
(b) → (c):
1 and 5
are being commutated

Review: Windings in Relative Motion

- The rotational device shown will be used to illustrate windings in relative motion.

Winding 1: N_1 turns on stator
Winding 2: N_2 turns on rotor

Assume that the turns are concentrated in one position.



end view

cross-sectional view

Air-gap size is exaggerated.

$$v_1 = r_1 i_1 + \frac{d\lambda_1}{dt}$$

$$v_2 = r_2 i_2 + \frac{d\lambda_2}{dt}$$

voltage equations

$$\lambda_1 = L_{11} i_i + L_{12} i_2$$

$$\lambda_2 = L_{21} i_i + L_{22} i_2$$

The magnetic system is assumed linear.

$$L_{11} = L_{\ell 1} + L_{m1} = \frac{N_1^2}{\Re_{\ell 1}} + \frac{N_1^2}{\Re_m}$$

$$L_{22} = L_{\ell 2} + L_{m2} = \frac{N_2^2}{\Re_{\ell 2}} + \frac{N_2^2}{\Re_m}$$

The self-inductances L_{11} and L_{22} are constants and may be expressed in terms of leakage and magnetizing inductances.

$\Re_m$ is the reluctance of the complete magnetic path of φ_{m1} and φ_{m2} , which is through the rotor and stator iron and twice across the air gap.

Let's now consider L_{12}.

$$\theta_r = \text{ angular displacement}$$

$$\omega_r = \text{ angular velocity}$$

$$\theta_r = \int_0^t \omega_r(\xi)\, d\xi + \theta_r(0)$$

When θ_r is zero, then the coupling between windings 1 and 2 is maximum. The magnetic system of winding 1 aids that of winding 2 with positive currents assumed. Hence the mutual inductance is positive.

$$L_{12}(0) = \frac{N_1 N_2}{\Re_m}$$

When θ_r is $\pi/2$, the windings are orthogonal. The mutual coupling is zero.

$$L_{12}\left(\frac{\pi}{2}\right) = 0$$

Assume that the mutual inductance may be adequately predicted by:

$$\begin{cases} L_{12}(\theta_r) = L_{sr}\cos(\theta_r) \\ L_{sr} = \dfrac{N_1 N_2}{\Re_m} \end{cases}$$

$$v_1 = r_1 i_1 + \frac{d\lambda_1}{dt}$$

$$v_2 = r_2 i_2 + \frac{d\lambda_2}{dt}$$

L_{sr} is the amplitude of the sinusoidal mutual inductance between the stator and rotor windings.

In writing the voltage equations, the total derivative of the flux linkages is required.

$$\lambda_1 = L_{11} i_1 + (L_{sr}\cos\theta_r) i_2$$

$$\lambda_2 = L_{22} i_2 + (L_{sr}\cos\theta_r) i_1$$

$$v_1 = r_1 i_1 + L_{11}\frac{di_1}{dt} + L_{sr}\cos\theta_r \frac{di_2}{dt} - i_2\omega_r L_{sr}\sin\theta_r$$

$$v_2 = r_2 i_2 + L_{22}\frac{di_2}{dt} + L_{sr}\cos\theta_r \frac{di_1}{dt} - i_1\omega_r L_{sr}\sin\theta_r$$

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} r_1 & 0 \\ 0 & r_2 \end{bmatrix} \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} + \frac{d}{dt} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}$$

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} L_{\ell 1} + L_{m1} & L_{sr} \cos\theta_r \\ L_{sr} \cos\theta_r & L_{\ell 2} + L_{m2} \end{bmatrix} \begin{bmatrix} i_{as} \\ i_{bs} \end{bmatrix}$$

Since the magnetic system is assumed to be linear:

$$W_f\left(i_1, i_2, \theta_r\right) = \frac{1}{2} L_{11} i_1^2 + L_{12} i_1 i_2 + \frac{1}{2} L_{22} i_2^2 = W_c\left(i_1, i_2, \theta_r\right)$$

$$T_e\left(\vec{i}, \theta\right) = \sum_{j=1}^{J} \left[i_j \frac{\partial \lambda_j\left(\vec{i}, \theta\right)}{\partial \theta} \right] - \frac{\partial W_f\left(\vec{i}, \theta\right)}{\partial \theta}$$

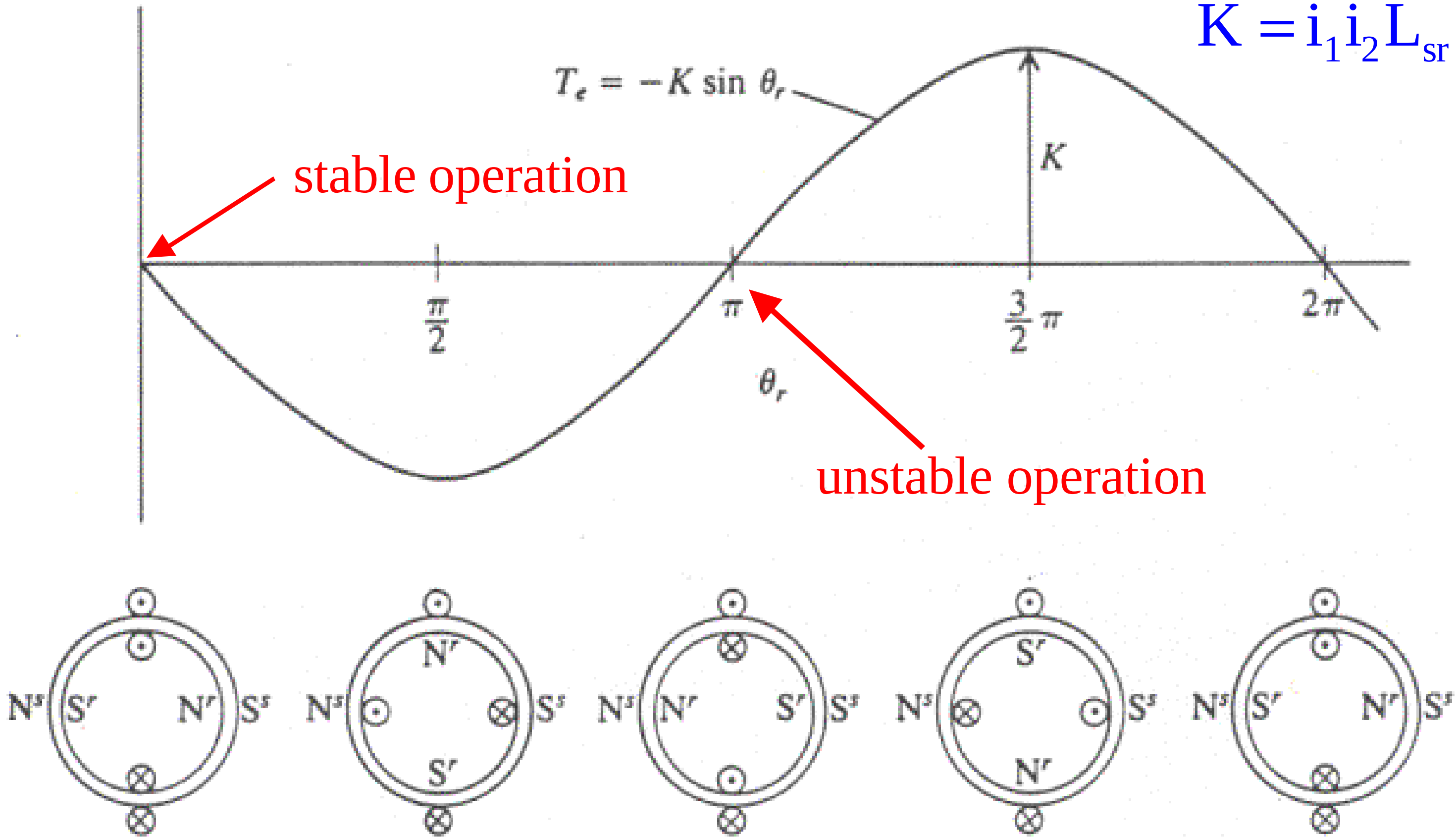
$$T_e\left(\vec{i}, \theta\right) = \frac{\partial W_c\left(\vec{i}, \theta\right)}{\partial \theta}$$

$$\Rightarrow \quad T_e\left(i_1, i_2, \theta_r\right) = -i_1 i_2 L_{sr} \sin\theta_r$$

- Consider the case where i_1 and i_2 are both positive and constant:

$$T_e = -K \sin \theta_r$$

$$K = i_1 i_2 L_{sr}$$



Electromagnetic torque versus angular displacement with constant winding currents

- In the diagram showing the positions of the poles of the magnetic system created by constant positive current flowing in the windings, it may appear at first that the N^s and S^s poles produced by positive current flowing in the stator winding are positioned incorrectly.
- Recall that flux issues from a north pole of a magnet into the air. Since the stator and rotor windings must each be considered as creating separate magnetic systems, we realize, by the right-hand rule, that flux issues from the N pole of the magnetic system established by the stator winding into the air gap. Similarly, flux produced by positive current in the rotor winding enters the air gap from the N pole of the magnetic system of the rotor.

- Here electromagnetic torque is produced in an attempt to align the magnetic systems established by currents flowing in the stator and rotor windings; in other words, to align the 1- and 2-axes.

- Although operation with constant winding currents is somewhat impracticable, it does illustrate the principle of positioning of stepper motors with a permanent-magnet rotor which, in many respects, is analogous to holding i_2 constant on the elementary device considered here.

Voltage and Torque Equations

- Rigorous derivations are possible, but they are lengthy!
- The armature coils revolve in a magnetic field established by a current flowing in the field winding.
- Voltage is induced in these coils by virtue of this rotation. However, the action of the commutator causes the armature coils to appear as a stationary winding with its magnetic axis orthogonal to the magnetic axis of the field winding.
- Consequently, voltages are not induced in one winding due to the time rate of change of the current flowing in the other (transformer action).

- <u>Field and armature voltage equations</u>:

$$\begin{bmatrix} v_f \\ v_a \end{bmatrix} = \begin{bmatrix} r_f + DL_{FF} & 0 \\ \omega_r L_{AF} & r_a + DL_{AA} \end{bmatrix} \begin{bmatrix} i_f \\ i_a \end{bmatrix}$$

$$D \equiv \frac{d}{dt}$$

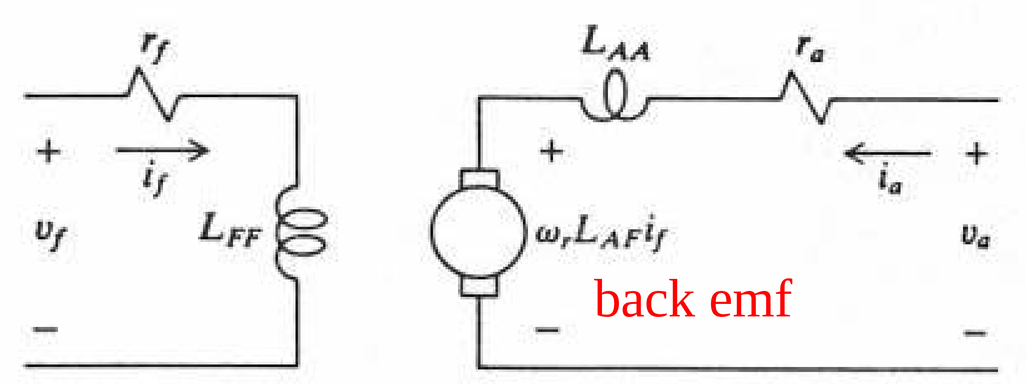
$$v_f = r_f i_f + L_{ff} \frac{di_f}{dt}$$

$$v_a = r_a i_a + L_{AA} \frac{di_a}{dt} + i_f \omega_r L_{AF}$$

L_{FF}	self-inductance of the field windings
L_{AA}	self-inductance of the armature windings
ω_r	rotor speed
L_{AF}	mutual inductance between the field and the rotating armature coils

<u>Equivalent Circuit</u>

(suggested by above equations)



- The voltage induced in the armature circuit, $w_r L_{AF} i_f$, is commonly referred to as the counter or back emf. It also represents the open-circuit armature voltage.
- There are several other forms in which the field and armature voltage equations are often expressed. For example: $$L_{AF} = \frac{N_a N_f}{\Re}$$ N_a = equivalent turns of armature windings
N_f = equivalent turns of field windings

$$L_{AF} i_f = N_a \frac{N_f i_f}{\Re} = N_a \Phi_f$$ Φ = field flux per pole

$$L_{AF} i_f = k_v$$ ⟸ frequently used

  – Even though a permanent magnet dc machine has no field circuit, the constant field flux produced by the permanent magnet is analogous to a dc machine with a constant k_v.

- The expression for the electromagnetic torque is:

$$T_e\left(i_1, i_2, \theta_r\right) = -i_1 i_2 L_{sr} \sin\theta_r$$

$$\theta_r = -\frac{\pi}{2}$$

$$T_e = L_{AF} i_f i_a = k_v i_a$$

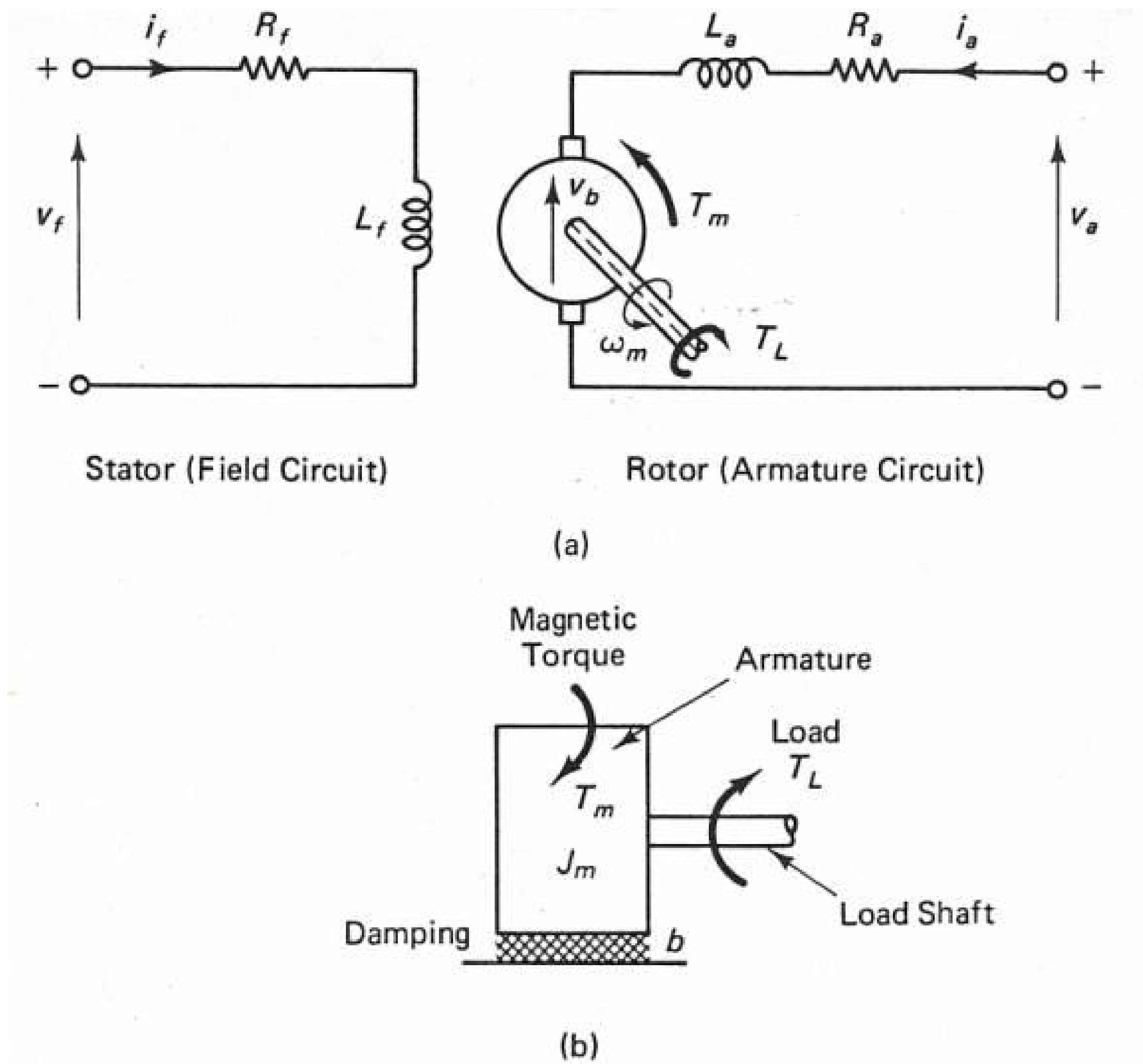
To account for rotational losses, sometimes k_v is multiplied by a factor < 1

- The field winding produces a stationary mmf and, owing to commutation, the armature winding also produces a stationary mmf which is displaced ½π electrical degrees from the mmf produced by the field winding. It follows then that the interaction of the two mmf's produces the electromagnetic torque.
- The torque and rotor speed are related by:

$$T_e = J\frac{d\omega_r}{dt} + B_m \omega_r + T_L$$

- J is the inertia of the rotor and, in some cases, the connected mechanical load ($kg\text{-}m^2$)
- A positive electromagnetic torque T_e acts to turn the rotor in the direction of increasing θ_r
- The load torque T_L is positive for a torque, on the shaft of the rotor, which opposes a positive electromagnetic torque T_e
- The constant B_m is a damping coefficient associated with the mechanical rotational system of the machine (N-m-s) and it is generally small and often neglected

Physical Modeling



Note change in variable names from previous equations:

$R_f = r_f$

$L_f = L_{FF}$

$L_a = L_{AA}$

$R_a = r_a$

$\omega_m = \omega_r$

$b = B_m$

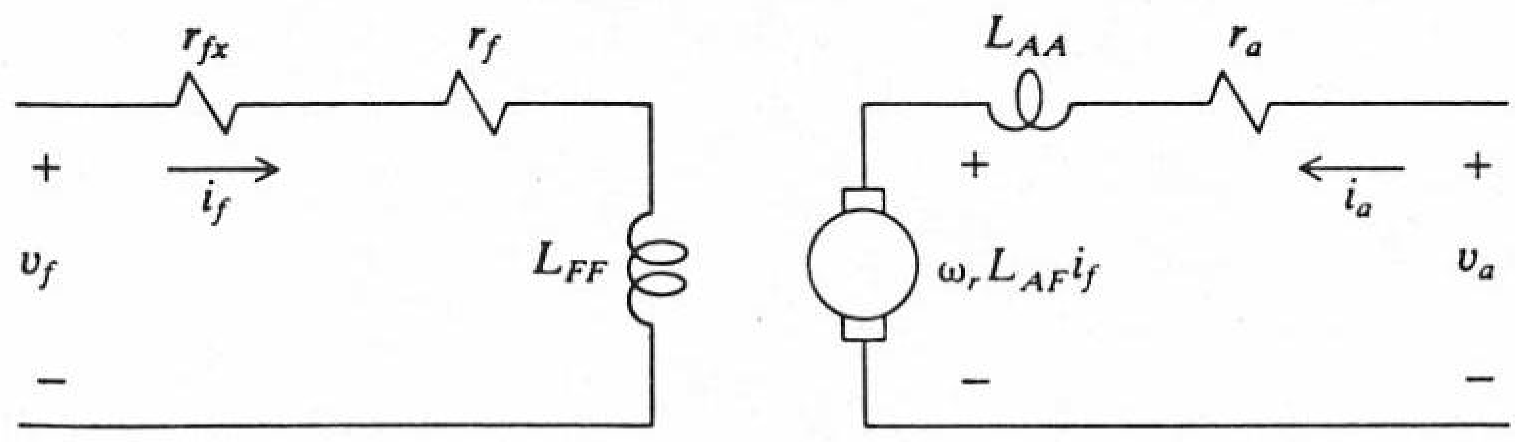
$T_m = T_e$

$J_m = J$

Basic Types of Direct-Current Machines

- The field and armature windings may be excited from separate sources or from the same source with the windings connected differently to form the basic types of dc machines.
  - Separate Winding Excitation (includes permanent magnet)
  - Shunt-Connected DC Machine
  - Series-Connected DC Machine
  - Compound-Connected DC Machine
- Here we give equivalent circuits for each of these machines along with an analysis and discussion of their steady-state operating characteristics.

- Separate Winding Excitation



Equivalent Circuit for Separate Field and Armature Excitation

– When the field and armature windings are supplied from separate voltage sources, the device may operate as either a motor or a generator.

– An external resistance r_{fx}, often referred to as a field rheostat, is connected in series with the field winding. It is used to adjust the field current if the field voltage is supplied from a constant source.

– The steady-state voltage equations are:

$$V_f = \left(r_{fx} + r_f\right) I_f = R_f I_f$$

$$V_a = r_a I_a + \omega_r L_{AF} I_f$$

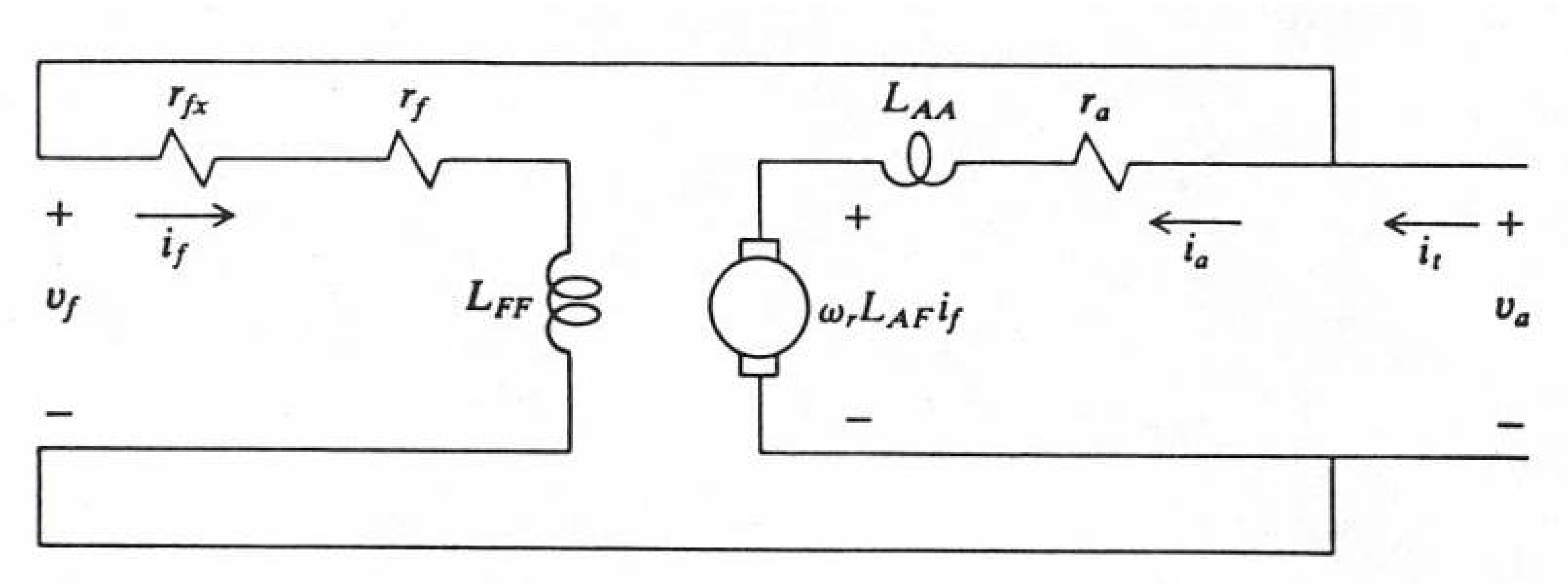
– Note: capital letters are used to denote steady-state voltages and currents.

- In the steady-state, if B_m is assumed to be zero, then:

$$T_e = J\frac{d\omega_r}{dt} + B_m\omega_r + T_L = T_L$$

- Analysis of steady-state performance is straightforward.
- A permanent-magnet dc machine fits into this class of dc machines. The field flux is established in these devices by a permanent magnet. The voltage equation for the field winding is eliminated and $L_{AF}i_f$ is replaced by constant k_v, which can be measured if not given.
- Most small, hand-held, fractional-horsepower dc motors are of this type, and speed control is achieved by controlling the amplitude of the applied armature voltage.

- ## Shunt-Connected DC Machine



Equivalent Circuit of a Shunt-Connected DC Machine

– With the connection shown, the machine may operate either as a motor or a generator.

– Since the field winding is connected between the armature terminals, this winding arrangement is commonly referred to as a shunt-connected dc machine or simply a shunt machine.

– During steady-state operation:

$$V_a = r_a I_a + \omega_r L_{AF} I_f$$
$$V_a = I_f R_f$$

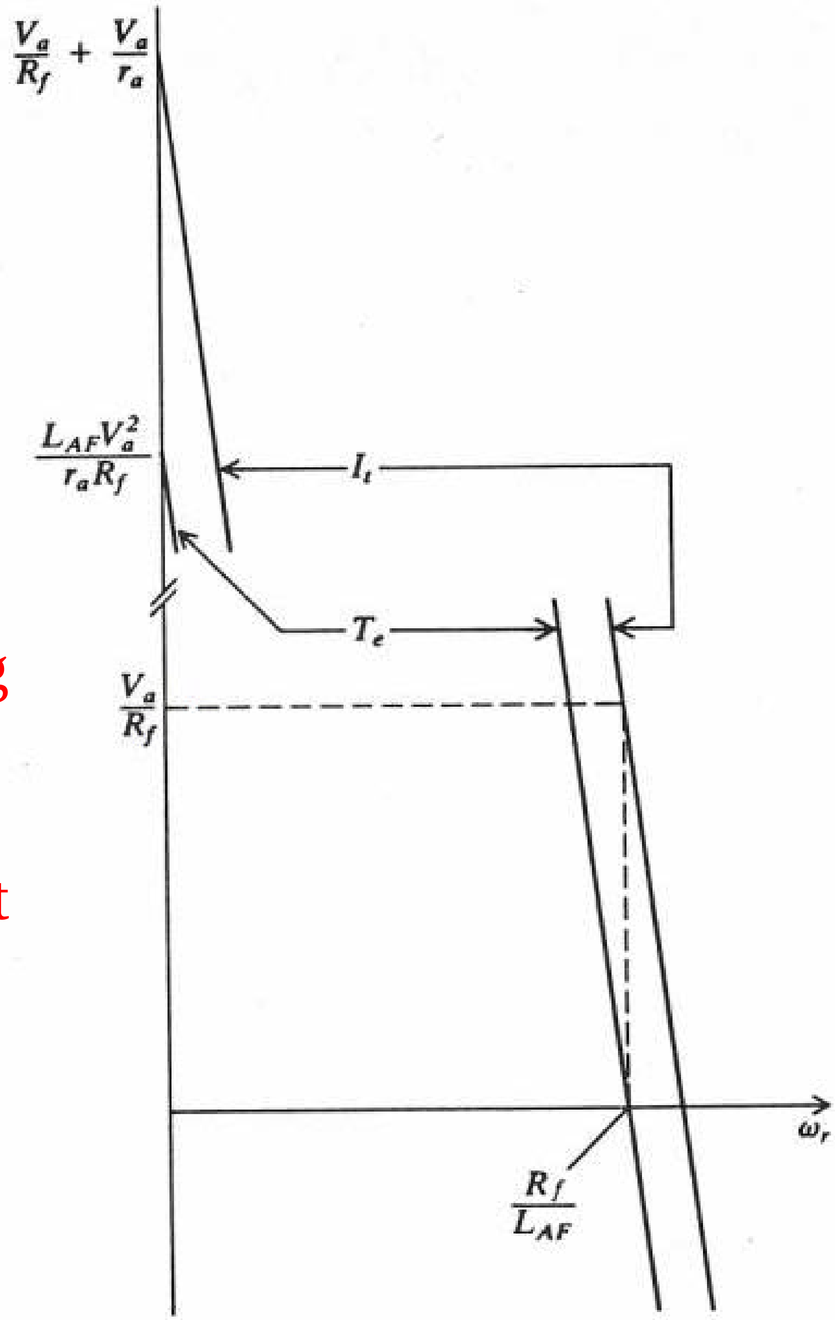
– The total current is: $I_t = I_f + I_a$

– The steady-state electromagnetic torque is found:

$$\left.\begin{aligned} V_a &= r_a I_a + \omega_r L_{AF} I_f \\ V_f &= I_f R_f \end{aligned}\right\} \Rightarrow T_e = L_{AF} i_f i_a$$

$$T_e = \frac{L_{AF} V_a^2}{r_a R_f}\left(1 - \frac{L_{AF}}{R_f}\omega_r\right)$$

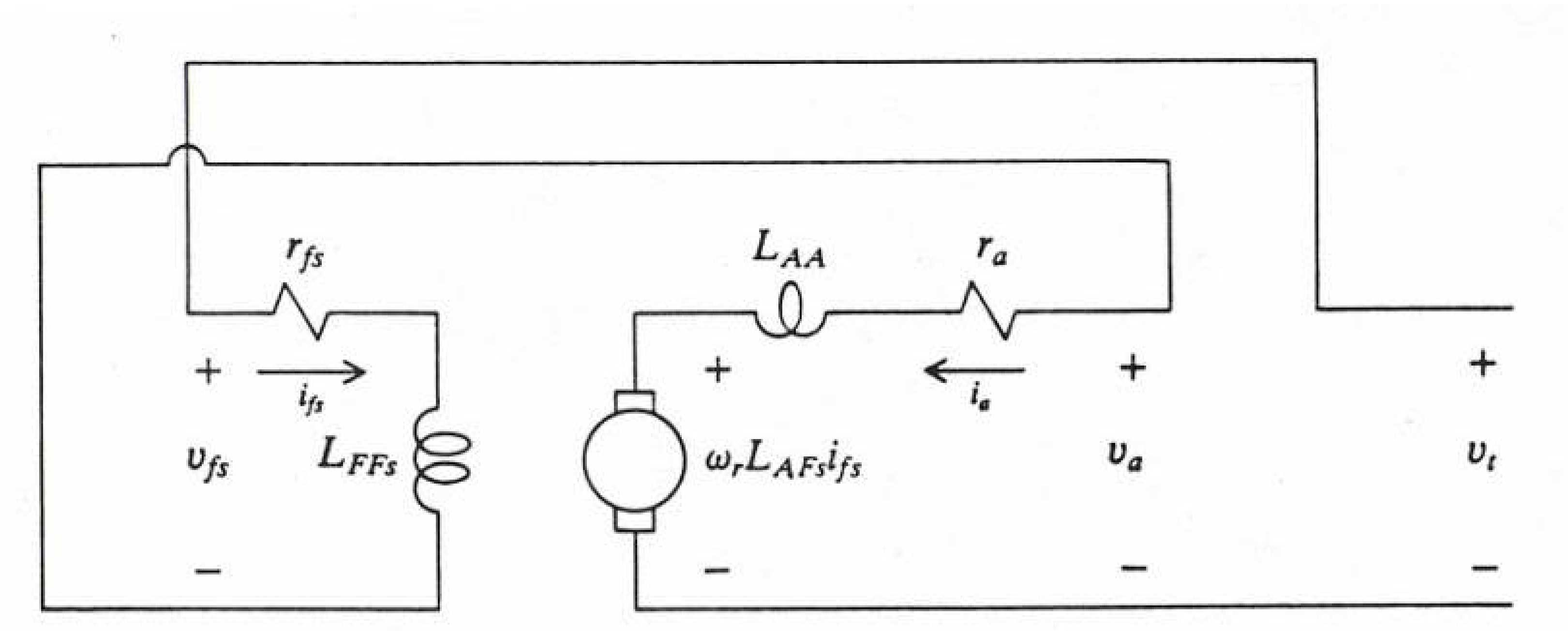
Steady-State Operating Characteristics of a Shunt-Connected DC Machine with Constant Source Voltage



- Several features of the steady-state operating characteristics of a shunt-connected dc machine with a constant voltage source warrant discussion:
  - At stall ($\omega_r = 0$), the steady-state armature current I_a is limited only by the armature resistance. In the case of small, permanent-magnet motors, the armature resistance is quite large and the starting armature current, which results when rated voltage is applied, is generally not damaging. However, large-horsepower machines are designed with a small armature resistance. Therefore, an excessively high armature current will occur during the starting period if rated voltage is applied to the armature terminals.

- To prevent high starting current, resistance may be inserted into the armature circuit at stall and decreased either manually or automatically to zero as the machine accelerates to normal operating speed.
- Other features of the shunt machine with a small armature resistance are the steep torque versus speed characteristics. In other words, the speed of the shunt machine does not change appreciably as the load torque is varied from zero to rated.

- ## Series-Connected DC Machine



Equivalent Circuit of a Series-Connected DC Machine

– When the field is connected in series with the armature circuit, the machine is referred to as a series-connected dc machine or a series machine.
– It is important to mention the physical difference between the field winding of a shunt machine and that of a series machine:
  - For a shunt-connected field winding, a large number of turns of small-diameter wire are used, making the resistance of the field winding quite large.
  - Since a series-connected field winding is in series with the armature, it is designed so as to minimize the voltage drop across it; thus, the winding is wound with a few turns of low-resistance wire.
– A series machine does not have wide application. However, a series field is often used in conjunction with a shunt field to form the more common compound-connected dc machine.

– For the series machine:

$$v_t = v_{fs} + v_a$$

$$i_a = i_{fs}$$

Subscript *s* stands for series field.

– The steady-state performance of the series-connected dc machine may be described by:

$$V_t = \left(r_a + r_{fs} + L_{AFs}\omega_r\right) I_a$$

$$T_e = L_{AF} i_f i_a = L_{AFs} I_a^2$$

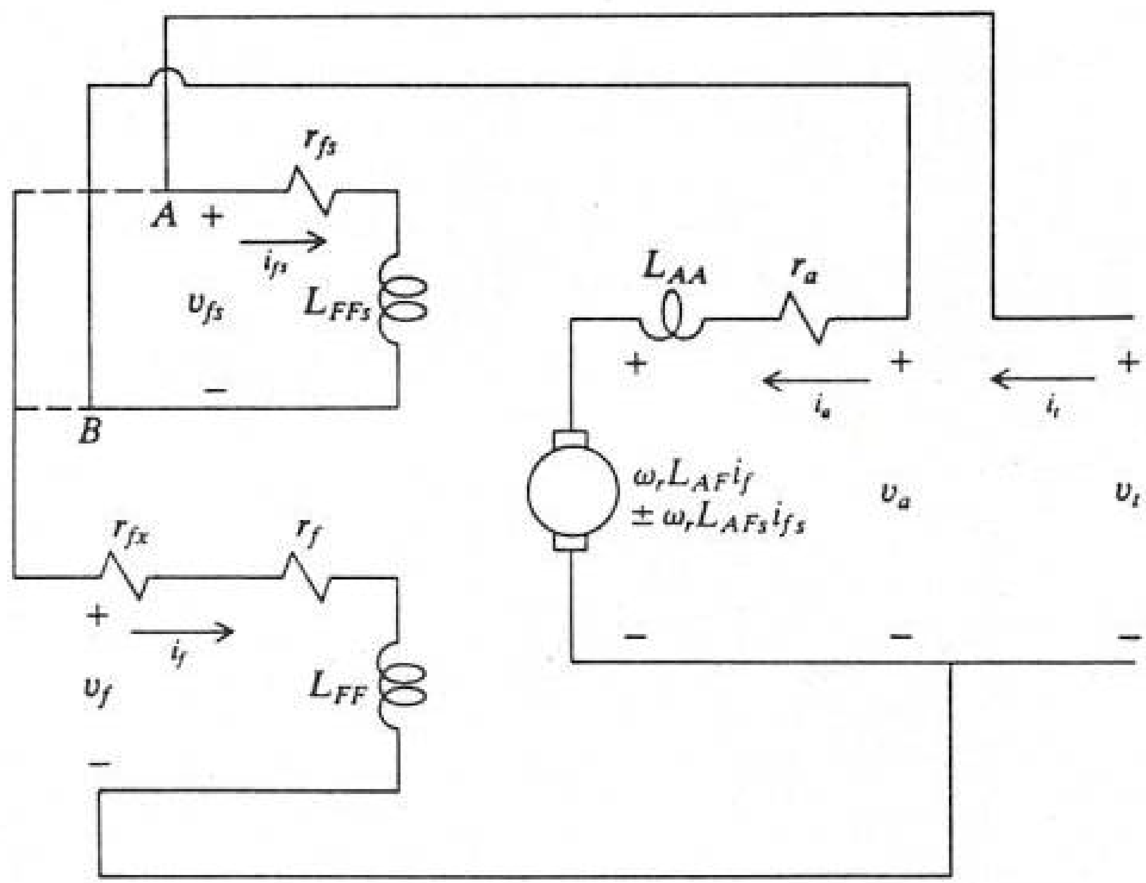
$$= \frac{L_{AFs} V_t^2}{\left(r_a + r_{fs} + L_{AFs}\omega_r\right)^2}$$

– Stall torque:

$$T_e = \frac{L_{AFs} V_t^2}{\left(r_a + r_{fs}\right)^2}$$

- The stall torque is quite high since it is proportional to the square of the armature current for a linear magnetic system. However, saturation of the magnetic system due to large armature currents will cause the torque to be less than calculated.
- At high rotor speeds, the torque decreases less rapidly with increasing speed. If the load torque is small, the series motor may accelerate to speeds large enough to cause damage to the machine.
- Consequently, the series motor is used in applications where a high starting torque is required and an appreciable load torque exists under normal operation.

- Compound-Connected DC Machine



Equivalent Circuit of a Compound-Connected DC Machine

- A compound-connected or compound dc machine is equipped with both a shunt and a series field winding.
- In most compound machines, the shunt field dominates the operating characteristics while the series field, which consists of a few turns of low-resistance wire, has a secondary influence. It may be connected so as to aid or oppose the flux produced by the shunt field.
- Depending upon the strength of the series field, this type of connection can produce a “flat” terminal voltage versus load current characteristic, whereupon a near-constant terminal voltage is achieved from no load to full load. In this case, the machine is said to be “flat-compounded.”
- An “over-compounded” machine occurs when the strength of the series field causes the terminal voltage at full load to be larger than at no load.

– The meaning of an "under-compounded" machine is obvious.

– In the case of compound dc motors, the series field is often connected to oppose the flux produced by the shunt field (differential compounding). If properly designed, this type of connection can provide a near-constant speed from no-load to full-load torque.

– The voltage equations for a compound dc machine may be written as:

$$\begin{bmatrix} v_f \\ v_t \end{bmatrix} = \begin{bmatrix} R_f + DL_{FF} & \pm DL_{FS} & 0 \\ \omega_r L_{AF} \pm DL_{FS} & \pm\omega_r L_{AFs} + r_{fs} + DL_{FFs} & r_a + DL_{AA} \end{bmatrix} \begin{bmatrix} i_f \\ i_{fs} \\ i_a \end{bmatrix}$$

L_{FS} is the mutual inductance between the shunt and series fields and the plus and minus signs are used so that either a cumulative or a differential connection may be described.

– The shunt field may be connected ahead of the series field (long-shunt connection) or behind the series field (short-shunt connection), as shown by A and B, respectively, in the figure. The long-shunt connection is commonly used. In this case:

$$v_t = v_f = v_{fs} + v_a$$

$$i_t = i_f + i_{fs} \qquad \text{where } i_{fs} = i_a$$

– The steady-state performance of a long-shunt-connected compound machine may be described by the following equations:

$$V_t = \left[\frac{r_a + r_{fs} \pm L_{AFs}\omega_r}{1 - \dfrac{L_{AF}}{R_f}\omega_r} \right] I_a$$

– The torque for the long-shunt connection may be obtained by employing $T_e = L_{AF} i_f i_a$ for each field winding.

– In particular:

$$T_e = L_{AF} I_f I_a \pm L_{AFs} I_{fs} I_a$$

$$= \frac{L_{AF} V_t^2 \left[1 - \frac{L_{AF}}{R_f}\omega_r\right]}{R_f \left(r_a + r_{fs} \pm L_{AFs}\omega_r\right)} \pm \frac{L_{AFs} V_t^2 \left[1 - \frac{L_{AF}}{R_f}\omega_r\right]^2}{\left(r_a + r_{fs} \pm L_{AFs}\omega_r\right)^2}$$

- Example Problem
  - A permanent-magnet dc motor is rated at 6V with the following parameters: $r_a = 7\ \Omega$, $L_{AA} = 120$ mH, $k_T = 2$ oz-in/A, $J = 150\ \mu$ oz-in-s^2. According to the motor information sheet, the no-load speed is approximately 3350 rpm and the no-load current is approximately 0.15 A. Interpret this information.
  - This permanent-magnet dc machine is operating with rated applied voltage and a load torque T_L of 0.5 oz-in. Determine the percent efficiency, i.e., (power output / power input)(100).

Time-Domain Block Diagrams, State Equations, and Transfer Functions

- Let's consider the time-domain block diagrams and state equations for the shunt-connected dc machine and the permanent-magnet dc machine.
- Shunt-Connected DC machine
  – The field and armature voltage equations and the relationship between torque and rotor speed are given as:

$$\begin{bmatrix} v_f \\ v_a \end{bmatrix} = \begin{bmatrix} r_f + DL_{FF} & 0 \\ \omega_r L_{AF} & r_a + DL_{AA} \end{bmatrix} \begin{bmatrix} i_f \\ i_a \end{bmatrix}$$

$$T_e = J\frac{d\omega_r}{dt} + B_m\omega_r + T_L$$

– These may be written as:

$$v_f = R_f\left[1+\frac{L_{FF}}{R_f}D\right]i_f = R_f\left[1+\tau_f D\right]i_f$$

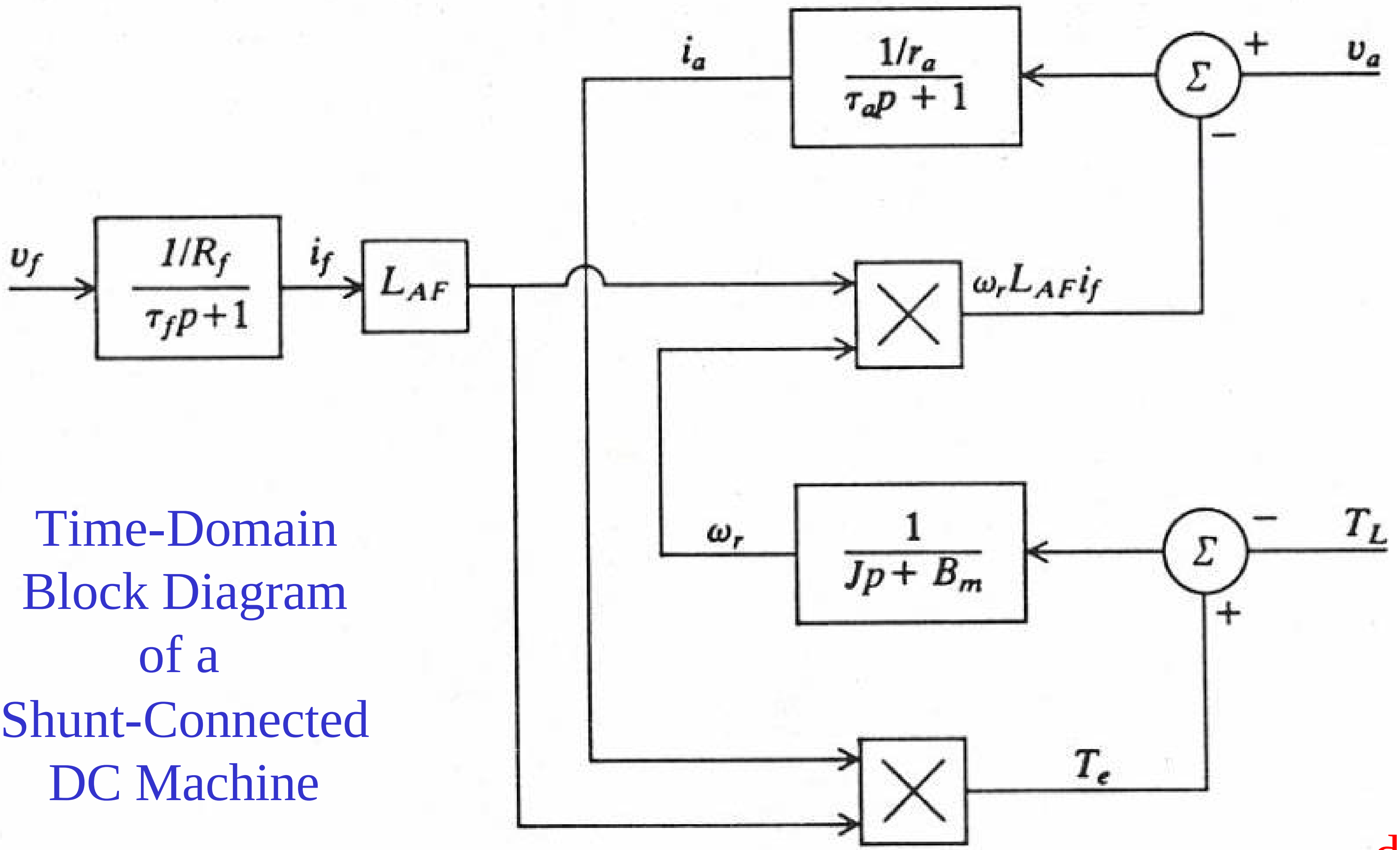
$$v_a = r_a\left[1+\frac{L_{AA}}{r_a}D\right]i_a + \omega_r L_{AF} i_f = r_a\left[1+\tau_a D\right]i_a + \omega_r L_{AF} i_f$$

$$T_e - T_L = \left(B_m + JD\right)\omega_r$$

– $\tau_f = L_{FF}/R_f$ is the field time constant and $\tau_a = L_{AA}/r_a$ is the armature time constant.

$$i_f = \frac{\frac{1}{R_f}}{\tau_f D + 1} v_f \qquad i_a = \frac{\frac{1}{r_a}}{\tau_a D + 1}\left(v_a - \omega_r L_{AF} i_f\right)$$

$$\omega_r = \frac{1}{JD + B_m}\left(T_e - T_L\right) = \frac{1}{JD + B_m}\left(L_{AF} i_f i_a - T_L\right)$$



Time-Domain
Block Diagram
of a
Shunt-Connected
DC Machine

$$p \triangleq D \triangleq \frac{d}{dt}$$

– The state variables of a system are defined as a minimal set of variables such that knowledge of these variables at any initial time t_0 plus information on the input excitation subsequently applied as sufficient to determine the state of the system at any time $t > t_0$.

– In the case of dc machines, the field current i_f, the armature current i_a, the rotor speed ω_r, and the rotor position θ_r are the state variables. However, since θ_r can be established from ω_r by using

$$\theta_r = \text{ angular displacement}$$

$$\omega_r = \text{ angular velocity}$$

$$\theta_r = \int_0^t \omega_r(\xi)\,d\xi + \theta_r(0)$$

and since θ_r is considered a state variable only when the shaft position is a controlled variable, we omit θ_r here.

– The state-variable equations (nonlinear) are:

$$\frac{di_f}{dt} = -\frac{R_f}{L_{FF}} i_f + \frac{1}{L_{FF}} v_f$$

$$\frac{di_a}{dt} = -\frac{r_a}{L_{AA}} i_a - \frac{L_{AF}}{L_{AA}} i_f \omega_r + \frac{1}{L_{AA}} v_a$$

$$\frac{d\omega_r}{dt} = -\frac{B_m}{J} \omega_r + \frac{L_{AF}}{J} i_f i_a - \frac{1}{J} T_L$$

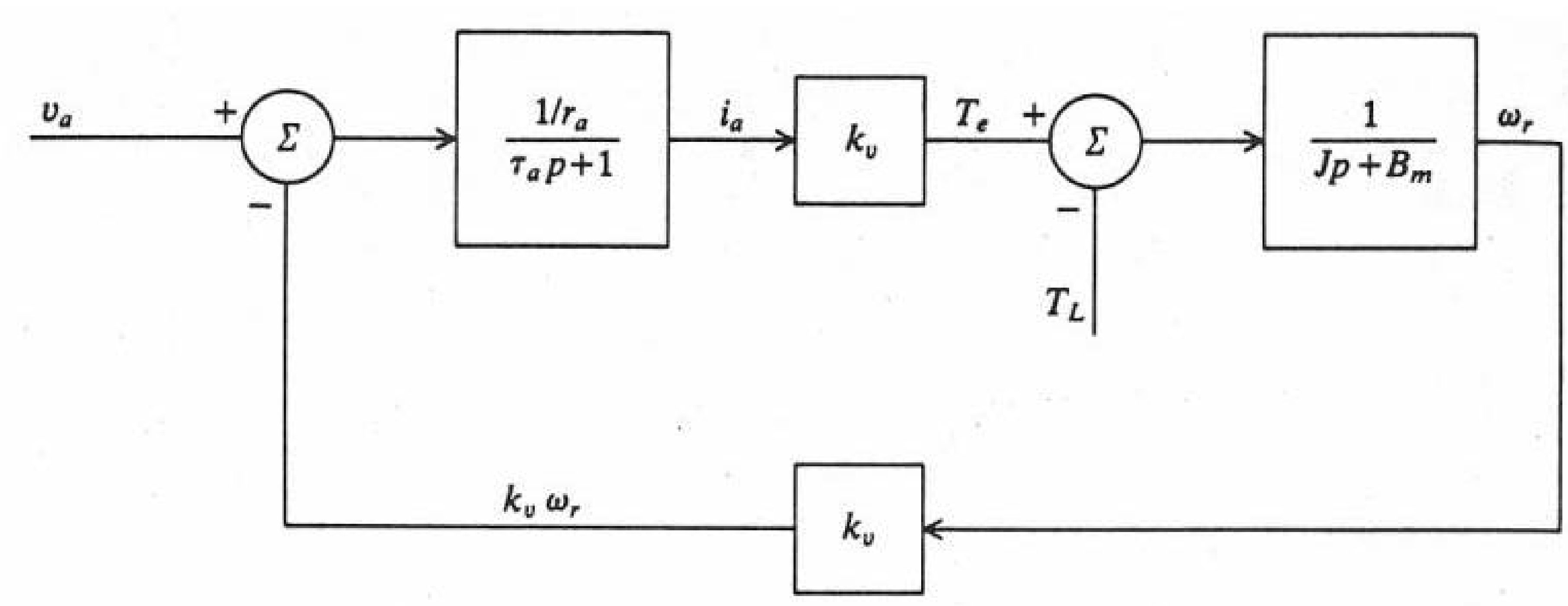
– In matrix form:

$$\frac{d}{dt}\begin{bmatrix} i_f \\ i_a \\ \omega_r \end{bmatrix} = \begin{bmatrix} -\frac{R_f}{L_{ff}} & 0 & 0 \\ 0 & -\frac{r_a}{L_{AA}} & 0 \\ 0 & 0 & -\frac{B_m}{J} \end{bmatrix} \begin{bmatrix} i_f \\ i_a \\ \omega_r \end{bmatrix} + \begin{bmatrix} 0 \\ -\frac{L_{AF}}{L_{AA}} i_f \omega_r \\ \frac{L_{AF}}{J} i_f i_a \end{bmatrix} + \begin{bmatrix} \frac{1}{L_{FF}} & 0 & 0 \\ 0 & \frac{1}{L_{AA}} & 0 \\ 0 & 0 & -\frac{1}{J} \end{bmatrix} \begin{bmatrix} v_f \\ v_a \\ T_L \end{bmatrix}$$

- Permanent-Magnet DC Machine
  - The equations which describe the operation of a permanent-magnet dc machine are identical to those of a shunt-connected dc machine with the field current held constant. So the work here applies to both.
  - For the permanent-magnet machine, $L_{AF}i_f$ is replaced by k_v, which is a constant determined by the strength of the magnet, the reluctance of the iron, and the number of turns of the armature winding.
  - The time-domain block diagram may be developed by using the following equations:

$$v_a = r_a\left[1+\frac{L_{AA}}{r_a}D\right]i_a + \omega_r L_{AF} i_f = r_a\left[1+\tau_a D\right]i_a + \omega_r k_v$$

$$T_e - T_L = \left(B_m + JD\right)\omega_r$$



Time-Domain
Block Diagram
of a
Permanent-Magnet
DC Machine

$$p \triangleq D \triangleq \frac{d}{dt}$$

– Since k_v is constant, the state variables are now i_a and ω_r. The state-variable equations are:

$$\frac{di_a}{dt} = -\frac{r_a}{L_{AA}} i_a - \frac{k_v}{L_{AA}} \omega_r + \frac{1}{L_{AA}} v_a$$

$$\frac{d\omega_r}{dt} = -\frac{B_m}{J} \omega_r + \frac{k_v}{J} i_a - \frac{1}{J} T_L$$

– In matrix form:

$$\frac{d}{dt}\begin{bmatrix} i_a \\ \omega_r \end{bmatrix} = \begin{bmatrix} -\frac{r_a}{L_{AA}} & -\frac{k_v}{L_{AA}} \\ \frac{k_v}{J} & -\frac{B_m}{J} \end{bmatrix} \begin{bmatrix} i_a \\ \omega_r \end{bmatrix} + \begin{bmatrix} \frac{1}{L_{AA}} & 0 \\ 0 & -\frac{1}{J} \end{bmatrix} \begin{bmatrix} v_a \\ T_L \end{bmatrix}$$

- Transfer Functions for a Permanent-Magnet DC Motor
  - Once the permanent-magnet dc motor is portrayed in block diagram from, it is often advantageous, for control design purposes, to express transfer functions between state and input variables.
  - Here the state variables are i_a and ω_r and the input variables are v_a and T_L.
  - From the block diagram, we have the relations:

$$i_a = \frac{\frac{1}{r_a}}{\tau_a D + 1}\left(v_a - k_v \omega_r\right)$$

$$\omega_r = \frac{1}{JD + B_m}\left(k_v i_a - T_L\right)$$

– After substitution and considerable work, we find:

$$\omega_r = \frac{\left(\frac{1}{k_v \tau_a \tau_m}\right) v_a - \left(\frac{1}{J}\right)\left(D + \frac{1}{\tau_a}\right) T_L}{D^2 + \left(\frac{1}{\tau_a} + \frac{B_m}{J}\right) D + \left(\frac{1}{\tau_a}\right)\left(\frac{1}{\tau_m} + \frac{B_m}{J}\right)} \qquad \text{where} \quad \tau_m = \frac{J r_a}{k_v^2}$$

– τ_m is called the inertia time constant

$$i_a = \frac{\left(\frac{1}{\tau_a r_a}\right)\left(D + \frac{B_m}{J}\right) v_a - \left(\frac{1}{k_v \tau_a \tau_m}\right) T_L}{D^2 + \left(\frac{1}{\tau_a} + \frac{B_m}{J}\right) D + \left(\frac{1}{\tau_a}\right)\left(\frac{1}{\tau_m} + \frac{B_m}{J}\right)}$$

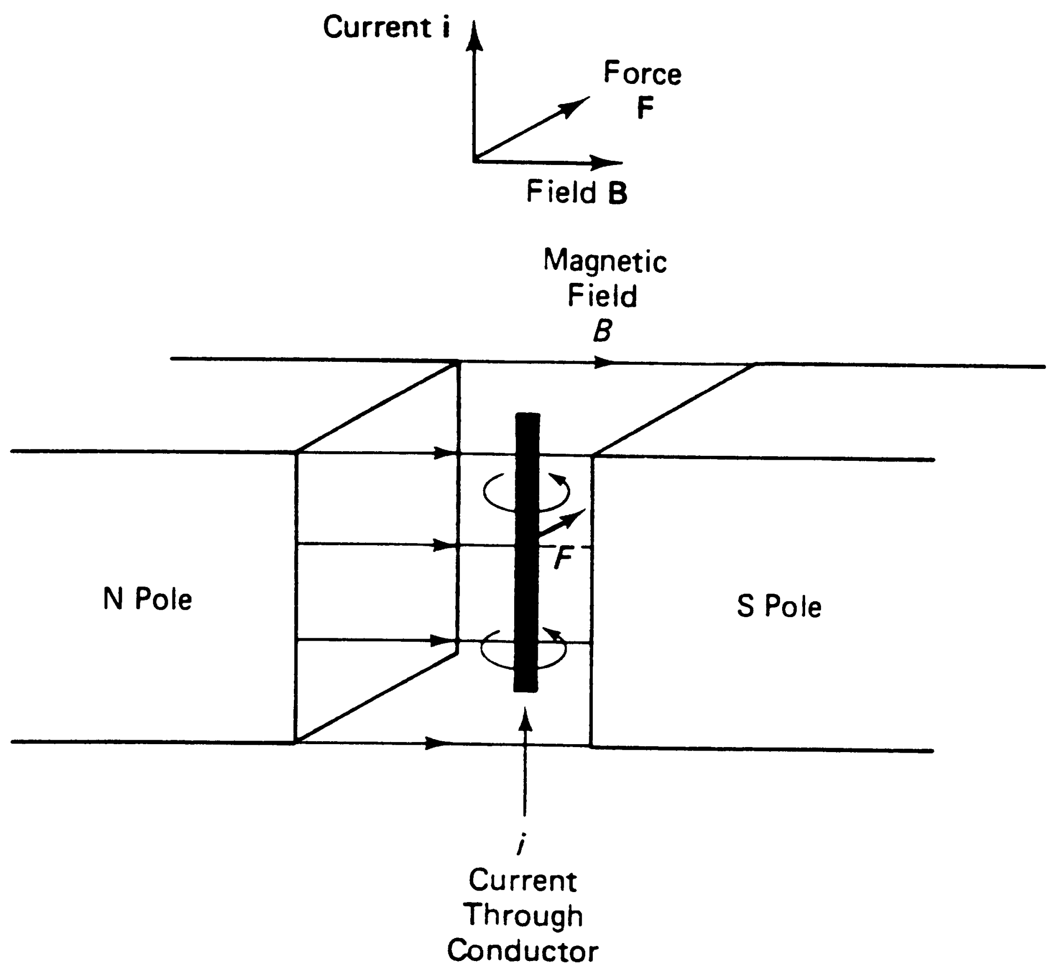
Summary

- The dc machine is unique in that it exerts a torque on the rotating member as a result of the interaction of two stationary, orthogonal magnetic systems.
- One is produced by current flowing in the windings of the stationary member (field) and the other is caused by the current flowing in the windings of the rotating member (armature).
- The permanent-magnet dc motor is still used quite widely in low-power control systems.
- However, brushless dc motors are rapidly replacing the permanent-magnet dc motor; the equations which describe these two devices are very similar.

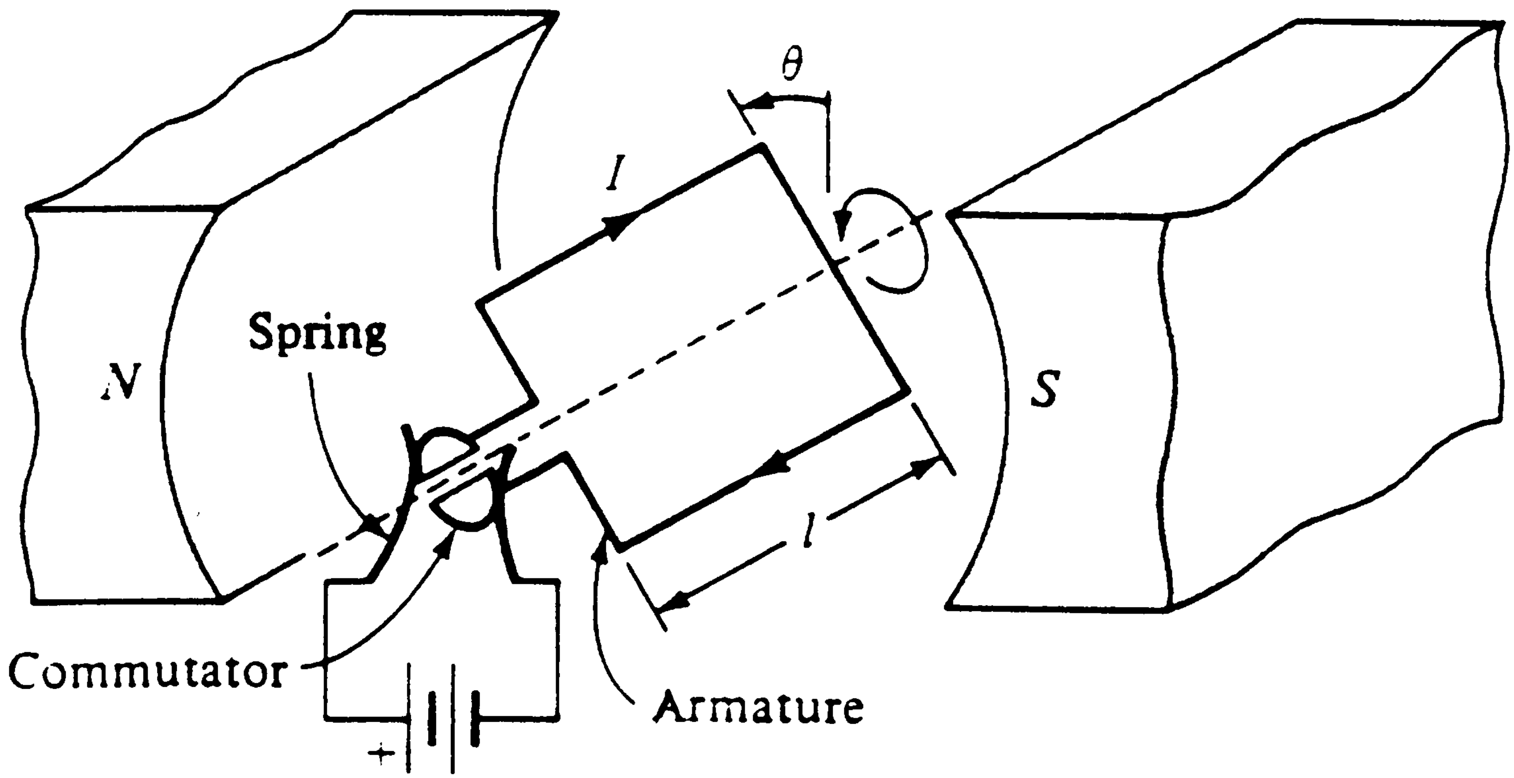
Elementary Approach to Permanent-Magnet DC Motor Modeling

$$\vec{F} = \oint i\overrightarrow{d\ell} \times \vec{B} = Bi\ell$$

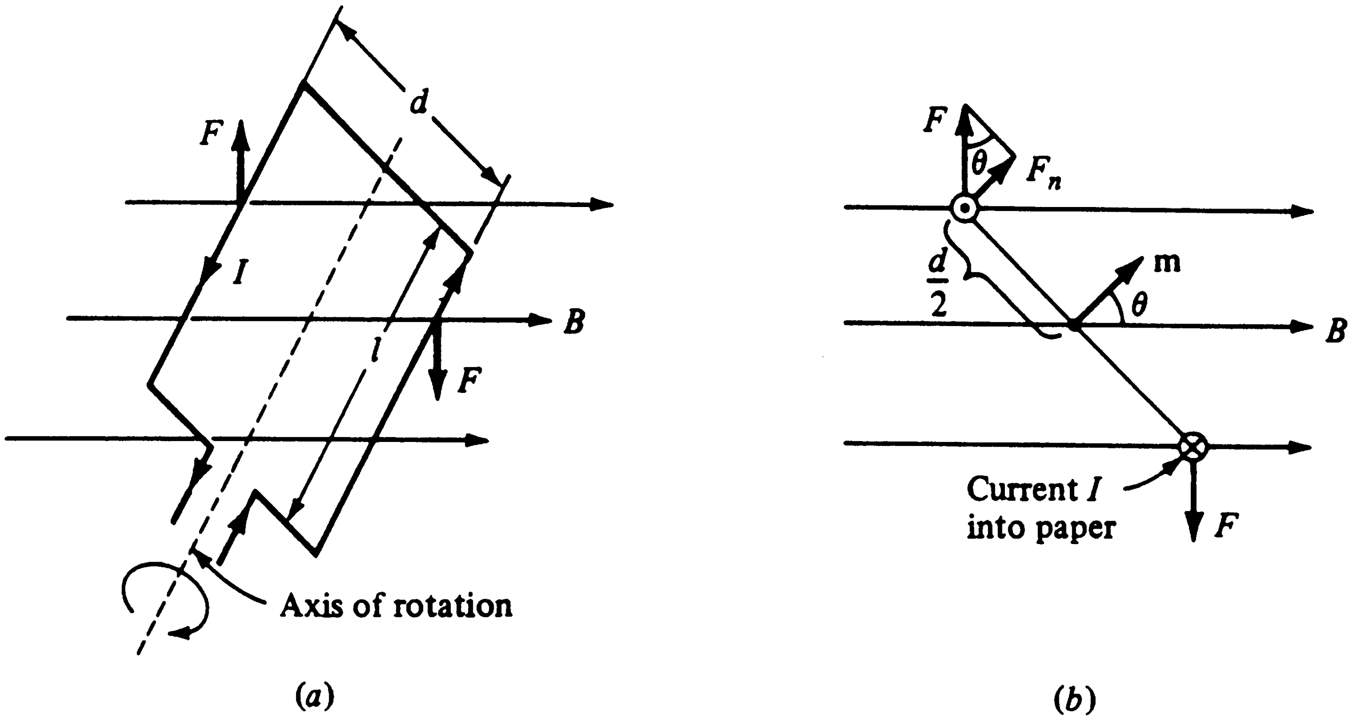
$$V_b = \int \vec{v} \times \vec{B} \bullet \overrightarrow{d\ell} = B\ell v$$



Elements of a Simple DC Motor



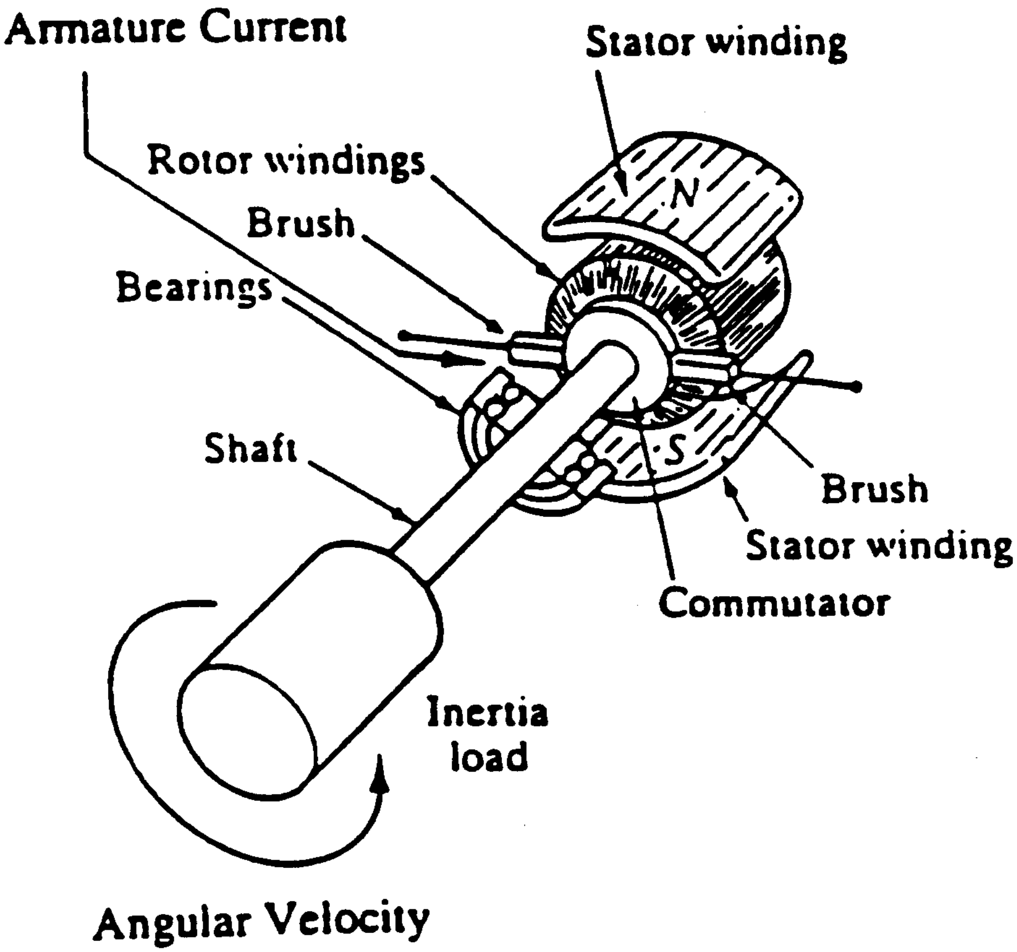
Torque of a DC Motor



$$T = 2F_n \left(\frac{d}{2} \right) N = \left(iB\ell \sin\theta \right) dN = iABN\sin\theta = mBN\sin\theta$$

$$\vec{T} = N\left[\vec{m} \times \vec{B} \right]$$

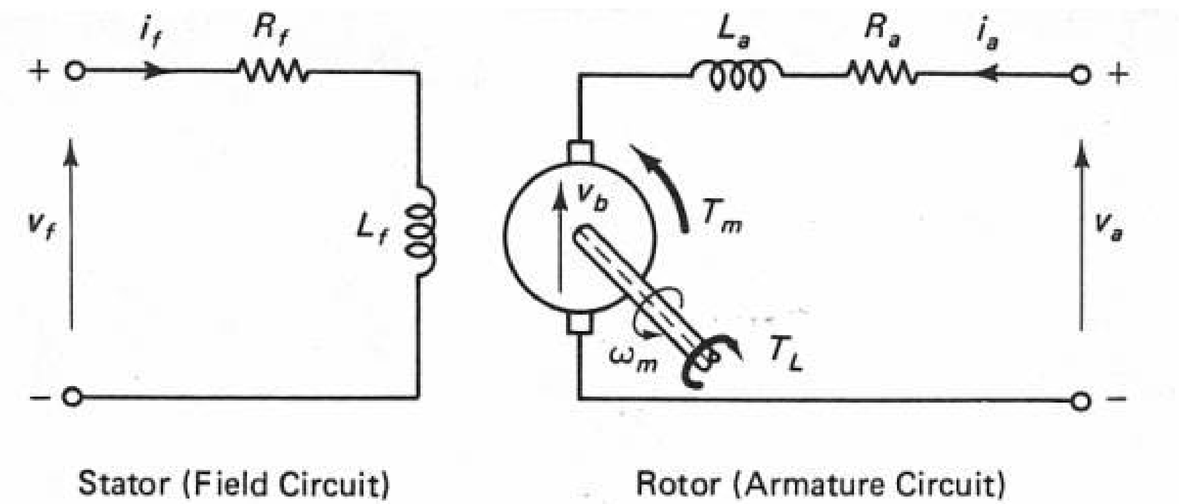
Schematic of a Brushed DC Motor



- <u>Modeling Assumptions</u>
  - The copper armature windings in the motor are treated as a resistance and inductance in series. The distributed inductance and resistance is lumped into two characteristic quantities, L and R.
  - The commutation of the motor is neglected. The system is treated as a single electrical network which is continuously energized.
  - The compliance of the shaft connecting the load to the motor is negligible. The shaft is treated as a rigid member. Similarly, the coupling between the tachometer and motor is also considered to be rigid.
  - The total inertia J is a single lumped inertia, equal to the sum of the inertias of the rotor, the tachometer, and the driven load.

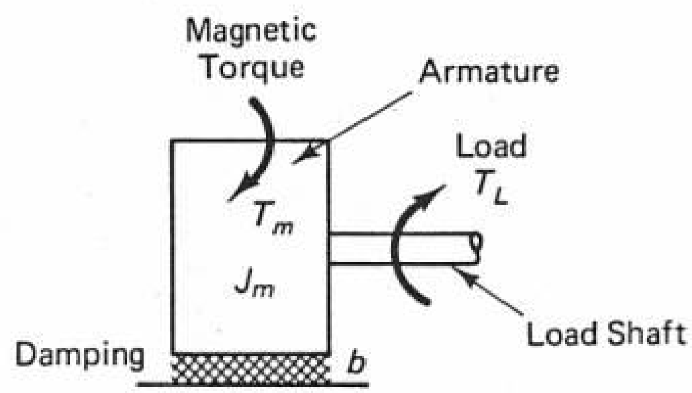
– There exists motion only about the axis of rotation of the motor, i.e., a one-degree-of-freedom system.
– The parameters of the system are constant, i.e., they do not change over time.
– The damping in the mechanical system is modeled as viscous damping *B*, i.e., all stiction and dry friction are neglected.
– Neglect noise on either the sensor or command signal.
– The amplifier dynamics are assumed to be fast relative to the motor dynamics. The unit is modeled by its DC gain, K_{amp}.
– The tachometer dynamics are assumed to be fast relative to the motor dynamics. The unit is modeled by its DC gain, K_{tach}.

Physical Modeling



(a)

For a permanent-magnet DC motor, i_f = constant.



(b)

- <u>Mathematical Modeling</u>
- The steps in mathematical modeling are as follows:
  - Define System, System Boundary, System Inputs and Outputs
  - Define Through and Across Variables
  - Write Physical Relations for Each Element
  - Write System Relations of Equilibrium and/or Compatibility
  - Combine System Relations and Physical Relations to Generate the Mathematical Model for the System

Physical Relations

$$V_L = L\frac{di_L}{dt} \qquad V_R = Ri_R \qquad T_B = B\omega$$

$$T_J = J\alpha = J\dot{\omega} \qquad J \equiv J_{motor} + J_{tachometer} + J_{load}$$

$$T_m = K_t i_m \qquad V_b = K_b \omega$$

$$P_{out} = T_m\omega = K_t i_m \omega \qquad P_{in} = V_b i_m = K_b \omega i_m$$

$$\frac{P_{out}}{P_{in}} = \frac{K_t}{K_b}$$

$$P_{out} = P_{in}$$

$$K_t = K_b \equiv K_m$$

$$K_t(oz-in/A) = 1.3524K_b(V/krpm)$$

$$K_t(Nm/A) = 9.5493\times10^{-3}K_b(V/krpm)$$

$$K_t(Nm/A) = K_b(V-s/rad)$$

System Relations + Equations of Motion

$$V_{in} - V_R - V_L - V_b = 0$$

$$T_m - T_B - T_J = 0$$

$$i_R = i_L = i_m \equiv i$$

$$V_{in} - Ri - L\frac{di}{dt} - K_b\omega = 0 \qquad J\dot{\omega} + B\omega - K_t i = 0$$

$$\begin{bmatrix} \dot{\omega} \\ \dot{i} \end{bmatrix} = \begin{bmatrix} -B/J & K_t/J \\ -K_b/L & -R/L \end{bmatrix} \begin{bmatrix} \omega \\ i \end{bmatrix} + \begin{bmatrix} 0 \\ 1/L \end{bmatrix} V_{in}$$

Steady-State Conditions

$$V_{in} - Ri - L\frac{di}{dt} - K_b\omega = 0$$

$$V_{in} - R\left(\frac{T}{K_t}\right) - K_b\omega = 0$$

$$T = \frac{K_t}{R}V_{in} - \frac{K_t K_b}{R}\omega$$

$$T_s = \frac{K_t}{R}V_{in}$$ Stall Torque

$$\omega_0 = \frac{V_{in}}{K_b}$$ No-Load Speed

Transfer Functions

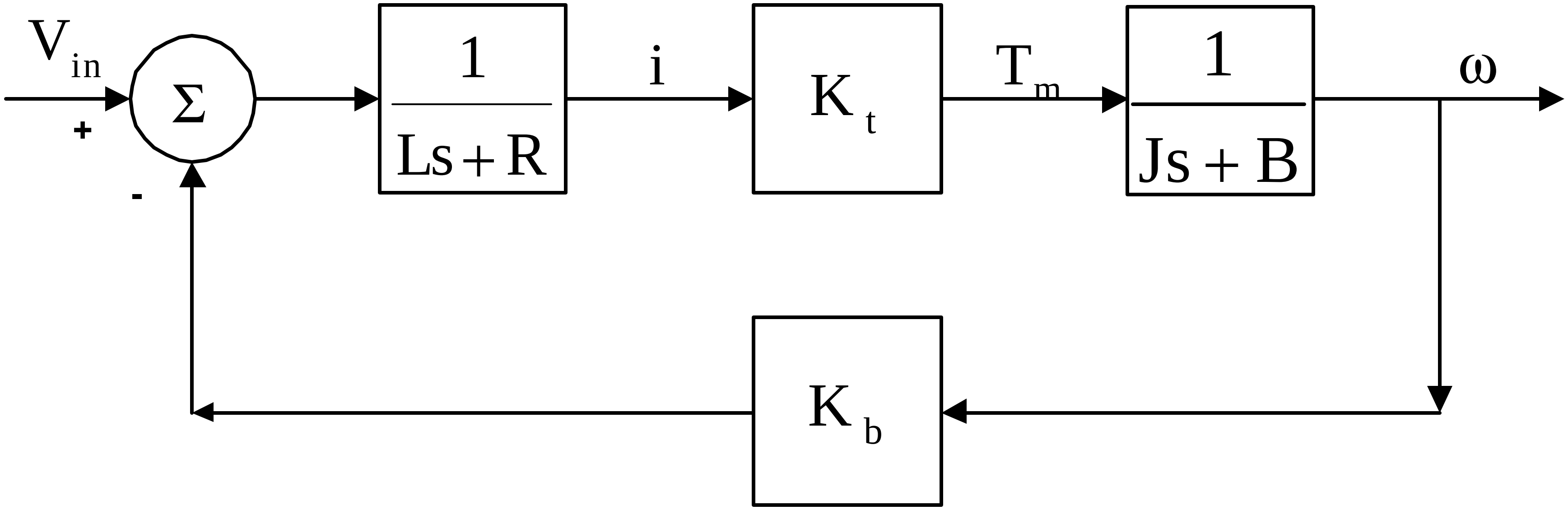
$$V_{in} - Ri - L\frac{di}{dt} - K_b\omega = 0 \qquad J\dot{\omega} + B\omega - K_t i = 0$$

$$V_{in}(s) - (Ls+R)I(s) - K_b\Omega(s) = 0 \qquad (Js+B)\Omega(s) - K_t I(s) = 0$$

$$\frac{\Omega(s)}{V_{in}(s)} = \frac{K_t}{(Js+B)(Ls+R)+K_tK_b} = \frac{K_t}{JLs^2+(BL+JR)s+(BR+K_tK_b)}$$

$$= \frac{\dfrac{K_t}{JL}}{s^2+\left(\dfrac{B}{J}+\dfrac{R}{L}\right)s+\left(\dfrac{BR}{JL}+\dfrac{K_tK_b}{JL}\right)}$$

Block Diagram



Simplification

$$\tau_m = \frac{J}{B} \ll \tau_e = \frac{L}{R}$$

$$V_{in} - Ri - K_b\omega = 0 \qquad\qquad J\dot{\omega} + B\omega - K_t i = 0$$

$$J\dot{\omega} + B\omega = K_t i = K_t\left(\frac{1}{R}\left(V_{in} - K_b\omega\right)\right) = \frac{K_t}{R}\left(V_{in} - K_b\omega\right)$$

$$\dot{\omega} + \left(\frac{K_t K_b}{RJ} + \frac{B}{J}\right)\omega = \frac{K_t}{RJ}V_{in}$$

$$\dot{\omega} + \left(\frac{1}{\tau_{motor}} + \frac{1}{\tau_m}\right)\omega = \frac{K_t}{RJ}V_{in}$$

$$\dot{\omega} + \left(\frac{1}{\tau_{motor}}\right)\omega = \frac{K_t}{RJ}V_{in} \qquad \text{since } \tau_m \gg \tau_{motor}$$

Control of DC Motors

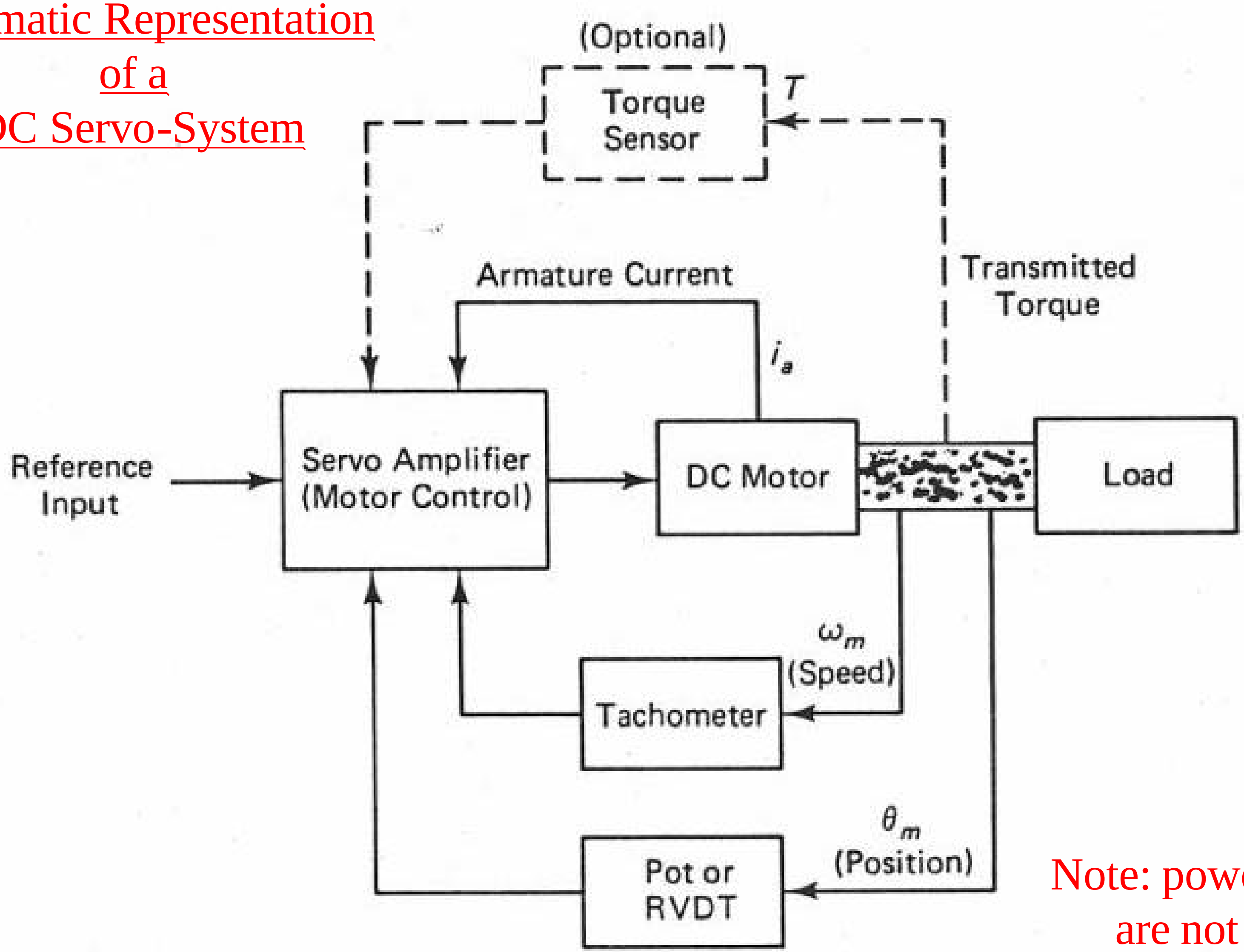
- DC Motors can be operated over a wide range of speeds and torques and are particularly suited as variable-drive actuators.
- The function of a conventional servo system that uses a DC motor as the actuator is almost exclusively motion control (position and speed control). There are applications that require torque control and they usually require more sophisticated control techniques.
- Two methods of control of a DC motor are:
  - Armature Control
  - Field Control

- Armature Control
  - Here the field current in the stator circuit is kept constant. The input voltage v_a to the rotor circuit is varied in order to achieve a desired performance. The motor torque can be kept constant simply by keeping the armature current constant because the field current is virtually constant in the case of armature control. Since v_a directly determines the motor back emf after allowance is made for the impedance drop due to resistance and inductance of the armature circuit, it follows that armature control is particularly suitable for speed manipulation over a wide range of speeds.

- Field Control
  - Here the armature voltage (and current) is kept constant. The input voltage v_f to the field circuit is varied. Since i_a is kept more or less constant, the torque will vary in proportion to the field current i_f. Since the armature voltage is kept constant, the back emf will remain virtually unchanged. Hence the speed will be inversely proportional to i_f. Therefore, by increasing the field voltage, the motor torque can be increased while the motor speed is decreased, so that the output power will remain more or less constant in field control. Field control is particularly suitable for constant-power drives under varying torque-speed conditions, e.g., tape-transport mechanisms.

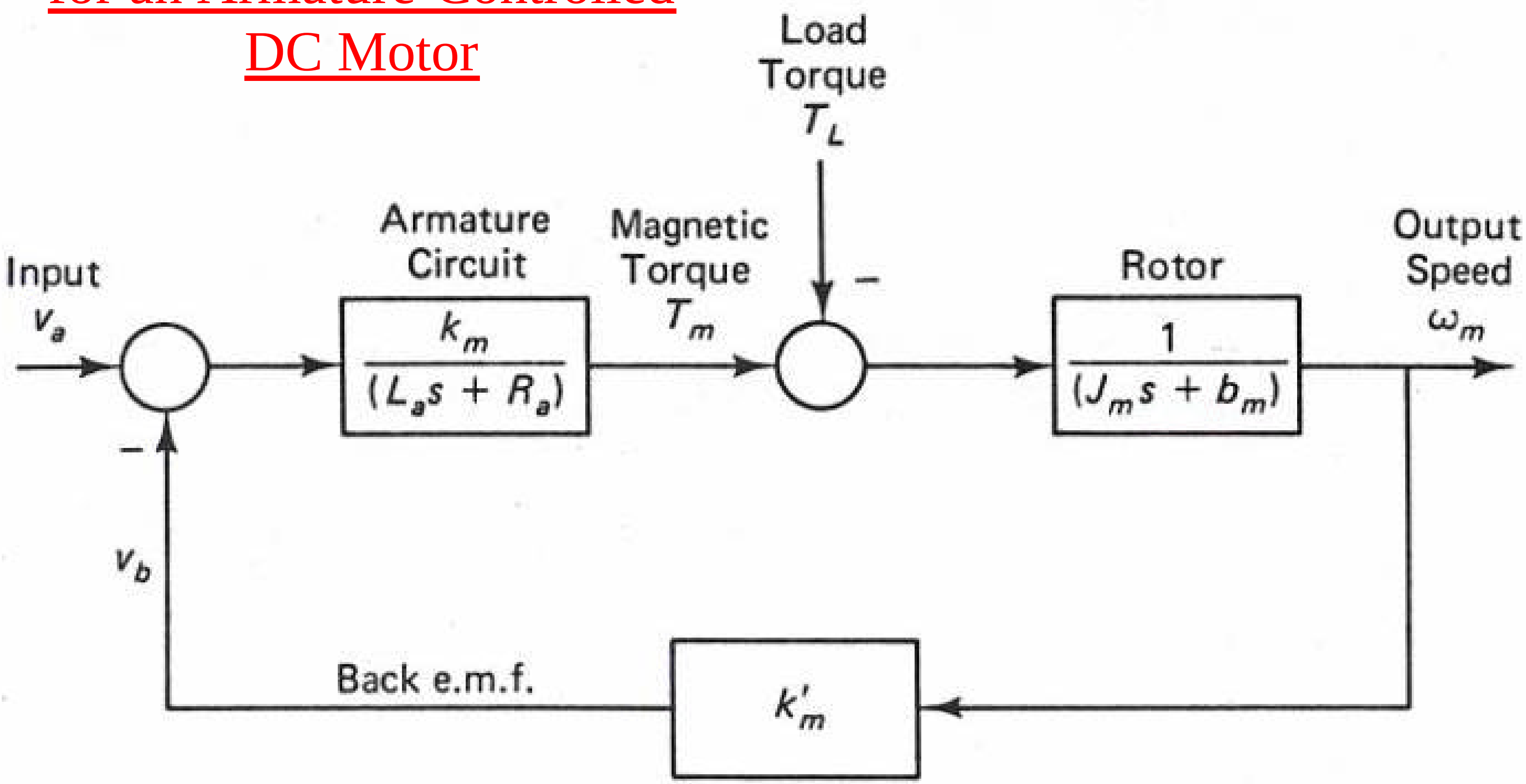
- DC Servomotors
  - DC servo-systems normally employ both velocity feedback and position feedback for accurate position control and also for accurate velocity control. Motion control requires indirect control of motor torque. In applications where torque itself is a primary output (e.g., metal-forming operations, tactile operations) and in situations where small motion errors could produce large unwanted forces (e.g., in parts assembly), direct control of motor torque would be necessary. For precise torque control, direct measurement of torque (e.g., strain gage sensors) would be required.

Schematic Representation of a DC Servo-System

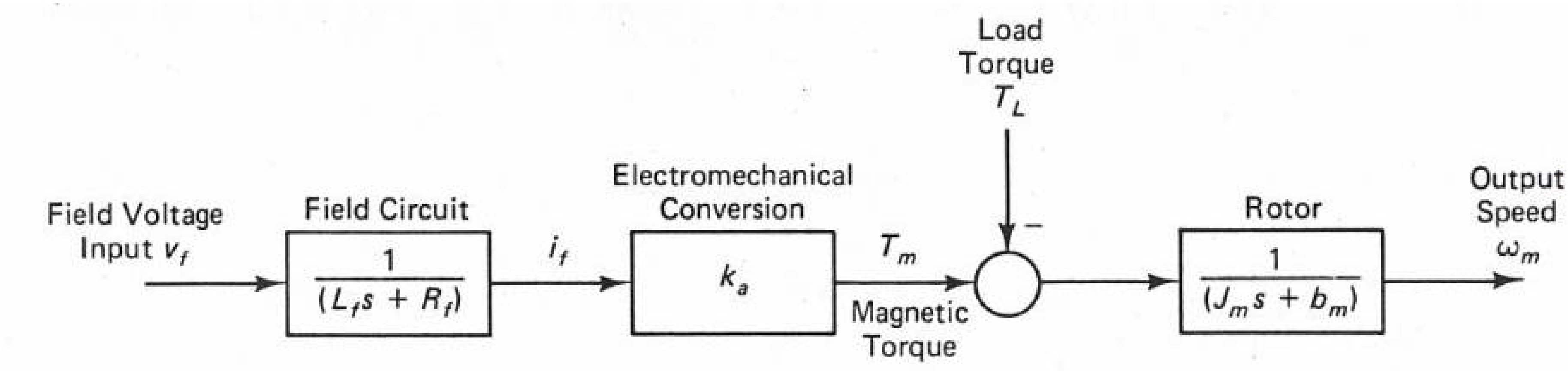


Note: power supplies are not shown

Open-Loop Block Diagram for an Armature-Controlled DC Motor

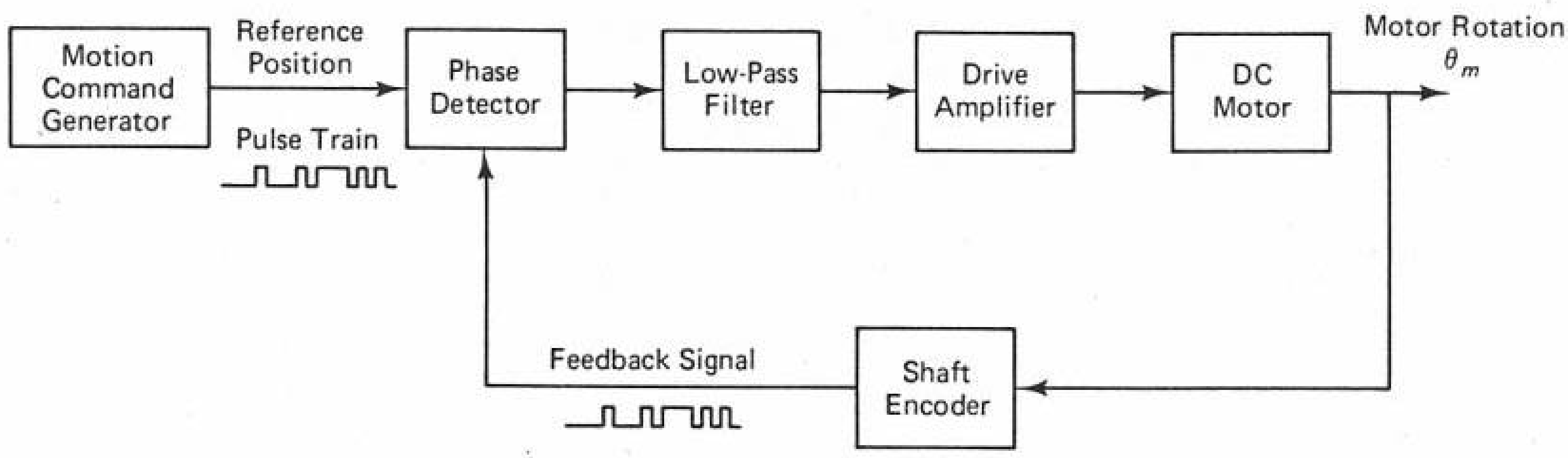


Open-Loop Block Diagram for a Field-Controlled DC Motor



- <u>Phase-Locked Control</u>
  - Phase-locked control is a modern approach to controlling DC motors.
  - This is a phase-control method. The objective is to maintain a fixed phase difference (ideally zero) between the reference signal and the position signal. Under these conditions, the two signals are phase-locked. Any deviation from the locked conditions will generate an error signal that will bring the motor motion back in phase with the reference command. In this manner, deviations due to external load changes on the motor are also corrected.
  - How do you determine the phase difference between two pulse signals? One method is by detecting the edge transitions. An alternative method is to take the product of the two signals and then low-pass filter the result.

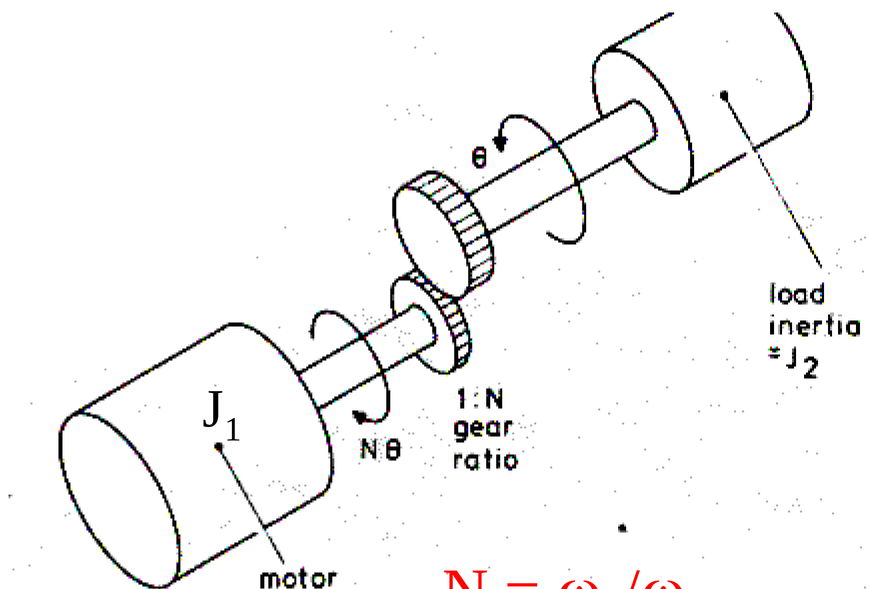
Schematic Diagram of a Phase-Locked Servo



Geared Systems, Optimum Gear Ratios, and Motor Selection

- Servomechanisms may be direct drive (motor coupled directly to load) or geared systems.
- For geared systems, choice of the motor also involves a choice of gear ratio.

Neglect backlash and elasticity in either gear teeth or shafts



$N = \omega_1/\omega_2$

- If we neglect frictional and other load effects, then the equation of motion for this system in terms of ω_2 is given by:

$$\left(J_2 + N^2 J_1\right)\frac{d\omega_2}{dt} = NT_m$$

- T_m is the electromagnetic torque from the motor.
- An important consequence of this result is that motor inertia is often the dominant inertia in a servo system. Consider the following numerical example:
  - Motor rotor inertia $J_1 = 1$, load inertia $J_2 = 100$, and gear ratio $N = 100$.
  - Physically the load appears 100 times "larger" than the motor, but because of the high (but not unusual) gear ratio, the motor's inertia effect is $N^2 J_1$ or 10,000, i.e., 100 times larger than the load.
  - Thus measures to "lighten" the load inertia are misplaced; we should really be striving for a lower inertia motor.

- When frictional and other load effects are negligible and inertia is dominant, an optimum gear ratio that maximizes load shaft acceleration for a given input torque exists and may be found as follows:

$$\left(J_2 + N^2 J_1\right)\frac{d\omega_2}{dt} = NT_m$$

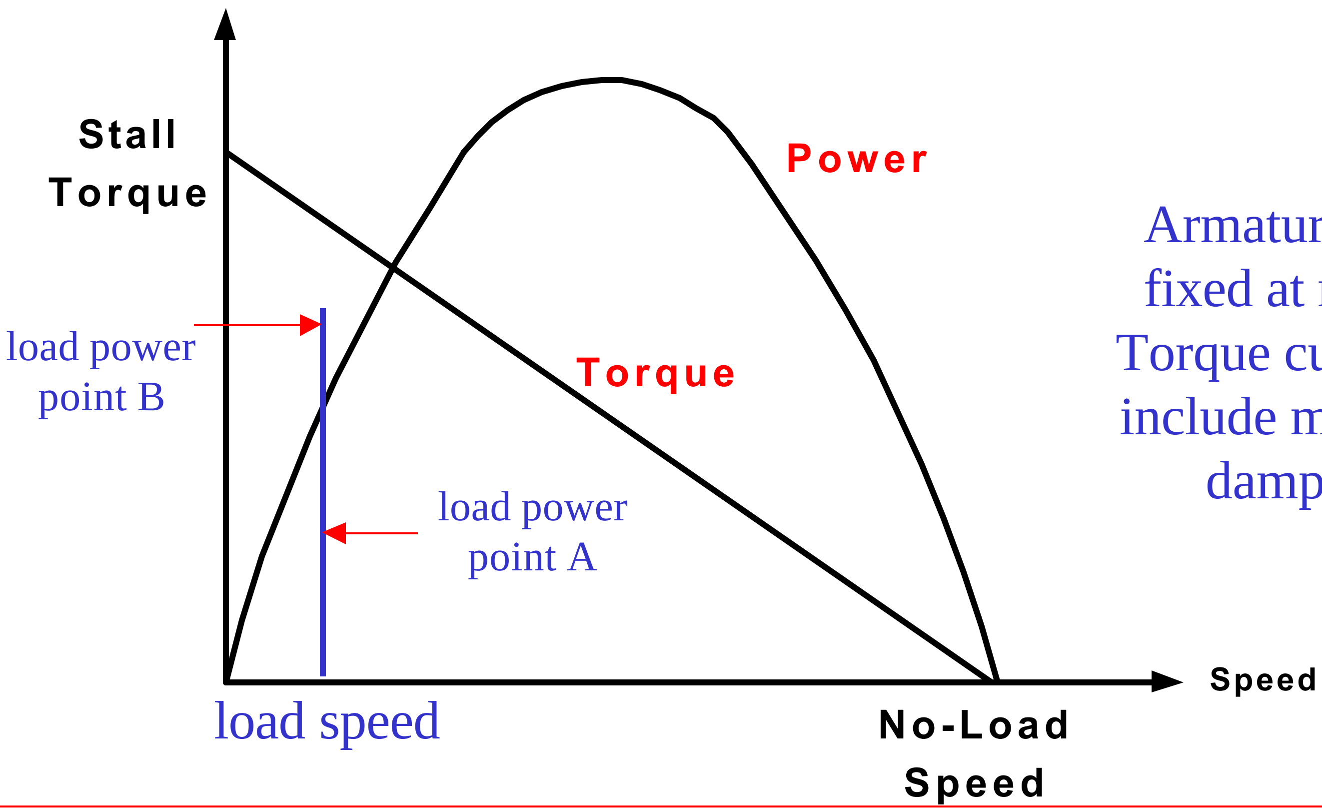
$$\frac{d\omega_2}{dt} = \frac{NT_m}{J_2 + N^2 J_1}$$

$$\frac{d}{dN}\left(\frac{d\omega_2}{dt}\right) = \frac{\left(J_2 + N^2 J_1\right)T_m - NT_m\left(2NJ_1\right)}{\left(J_2 + N^2 J_1\right)^2} = 0$$

$$N_{opt} = \sqrt{\frac{J_2}{J_1}}$$

Referred inertia of the motor rotor is equal to the actual load inertia

- When friction, load, and acceleration torques are all significant, gear ratio selection is less straightforward.
- Consider the linear speed/torque curve (typical of an armature-controlled dc motor with fixed field). Power is the product of speed and torque, hence the parabolic speed/power curve.



Armature voltage is fixed at rated value. Torque curve does not include motor viscous damping loss.

- One possible design criterion requires that we operate the motor at the maximum power point (which occurs at one half the no-load speed) when it is supplying the greatest demand of the load.
- One could compute the maximum power demand at maximum velocity; however, it is unlikely that this maximum speed will coincide with the peak power speed of the motor; thus a direct drive would not satisfy our requirement to operate at the peak power point.
- The figure shows two load power points at the same maximum load speed.
- If maximum load power corresponded to point A, a direct drive would be possible since the motor has more power than needed at this speed. The excess power simply means that more acceleration is available than is required.

- If instead our calculations had given point B, direct drive would not be possible since the load requires more power than the motor can supply at that speed. However, this load is less than the peak motor power, so suitable gearing can reconcile the supply/demand mismatch.
- Since one finds both direct-drive and geared systems in practical use, it is clear that the one scheme is not always preferable to the other.
- Direct drive is favored because of reduced backlash and compliance and longer motor life due to lower speed.
- The advantages of geared drives include smoother motor operation at higher speeds, possibly higher torsional natural frequency, and the lower cost of smaller motors. Of course, if a sufficiently large motor is simply not available, then a geared drive is a necessity.

Motor Selection Considerations

- Torque and speed are the two primary considerations in choosing a motor for a particular application. Motor manufacturers' data that are usually available to users include the following specifications:
- Mechanical Specifications
  – Mechanical time constant
  – Speed at rated load
  – Rated torque
  – Frictional torque
  – Dimensions and weight
  – No-load speed

- No-load acceleration
- Rated output power
- Damping constant
- Armature moment of inertia

- <u>Electrical Specifications</u>
  - Electrical time constant
  - Armature resistance and inductance
  - Compatible drive circuit specifications (voltage, current, etc.)
  - Input power
  - Field resistance and inductance
- <u>General Specifications</u>
  - Brush life and motor life

– Heat transfer characteristics
– Coupling methods
– Efficiency
– Mounting configuration
– Operating temperature and other environmental conditions

- It should be emphasized that there is no infallible guide to selecting the best motor. There are always several workable configurations. Constraints (e.g., space, positioning resolution) can often eliminate several designs.
- The engineer must make the motor drive system work both electrically and mechanically. Look at the motor-to-load interface before looking at the electrical drive-to-motor interface.
- In a typical motion control application the requirement will be to overcome some load frictional force and move a mass through a certain distance in a specified time.

- The designer should weigh the following:
  - Moment of Inertia
  - Torque
  - Power
  - Cost
- <u>Load Inertia</u>
  - For optimum system performance, the load moment of inertia should be similar to the motor inertia.
  - When gear reducers intervene between motor and load, the reflected load inertia is J_L/N^2, where N is the gear ratio.
  - If the motor inertia J_M is equal to the reflected load inertia, the fastest load acceleration will be achieved, or conversely, the torque to obtain a given acceleration will be minimized.

- Therefore, matched inertias are best for fast positioning.
- Peak power requirements are minimized by selecting the motor inertia so that the reflected load inertia is 2.5 times as large as the motor inertia. The torque will be increased but the maximum speed will be further reduced. A load inertia greater than 2.5 times the motor inertia is less than ideal, but it should not present any problems if the ratio is less than 5. A larger motor inertia implies that the same performance can be achieved at a lower cost by selecting a smaller motor.
- There is a wide range of motor inertias on the market today. Overlap is extensive. An engineer can virtually consider any type of motor including brushless and stepper, at the first stage of the design-inertia match.

- **Torque**
  - The motor must supply sufficient torque T_m to overcome the load friction and to accelerate the load over a distance (radians) *s* in time *t*.
  - Torque and acceleration at the motor are given by:

$$\left.\begin{aligned} \alpha_m &= N\alpha_L \\ T_L &= T_f + J_L\alpha_L \end{aligned}\right\} T_m = \frac{T_L}{N} + J_m\alpha_m \quad \Rightarrow \quad T_m = \frac{1}{N}\left[T_f + \alpha_L\left(J_L + N^2 J_m\right)\right]$$

  - For linear acceleration over distance *s* in time *t* :

$$\alpha_L = \frac{2s}{\tau^2}$$

  - For a damped ($\zeta = 0.7$) second-order response over distance *s* :

$$\left(\alpha_L\right)_{max} = \omega_n^2 s$$

- Allowances should always be made for variations in load and bearing behavior as well as motor production variations.
- An initial design should be planned without a gear reducer. In many cases direct drive is not possible because load torque requirements far exceed the torque delivered by a motor of reasonable size.
- Critical needs on space or weight can lead to gear reducers for otherwise perfectly matched motor/load systems.
- The problem with gear reducers is gear backlash. If gears mesh too tightly, there is severe sliding friction between teeth which can cause lockup. Thus, the teeth spacing, backlash, is a tradeoff between reducing the power loss within the gears (loose fit) or improving position accuracy (tight fit) of the load.

- Power
  - Besides maximum torque requirements, torque must be delivered over the load speed range. The product of torque and speed is power. Total power P is the sum of the power to overcome friction P_f and the power to accelerate the load P_a, the latter usually the dominant component:

$$P = P_f + P_a = T_f \omega + J\alpha\omega$$

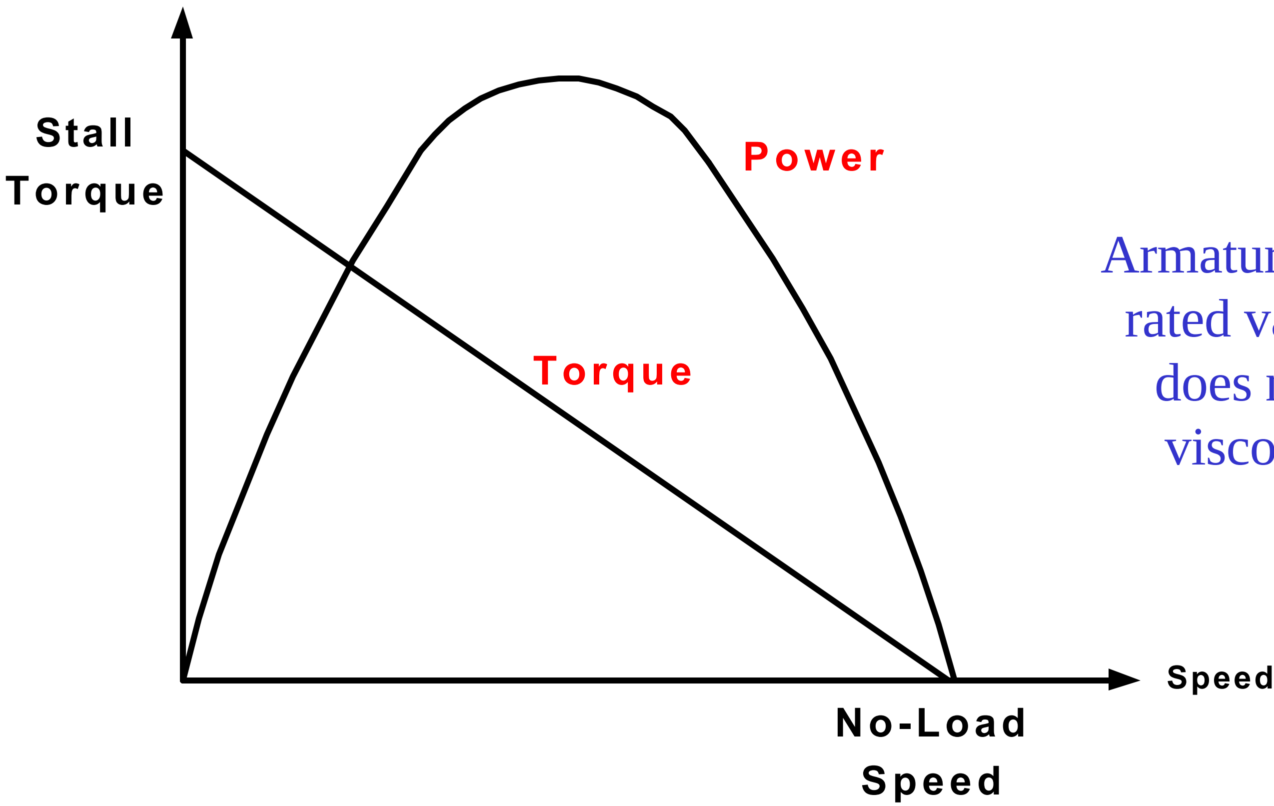
  - Peak power required during acceleration depends upon the velocity profile. If the load is linearly accelerated over distance *s* in time *t*, the maximum power is:

$$(P_a)_{max} = \frac{4Js^2}{\tau^3}$$

- If the load undergoes a damped (ζ = 0.7) second-order response over distance *s*, the maximum power is:

$$\left(P_a\right)_{max} = \left(0.146\right) J \omega_n^3 s^2$$

- It is interesting to note that to accelerate a load in one half the time will require eight times the power.
- The torque-speed curve for permanent-magnet DC motors is a linear line from stall torque to no-load speed. Therefore, the maximum power produced by the motor is the curve midpoint or one-fourth the stall torque and maximum speed product.



Armature voltage is fixed at rated value. Torque curve does not include motor viscous damping loss.

- The maximum speed for a permanent-magnet DC motor is 10,000 RPM, while for a brushless DC motor it is > 20,000 RPM.
- As a starting point, choose a motor with double the calculated power requirement.

• Cost

- Among several designs the single most important criterion is cost.
- Although it may be more prudent to choose the first workable design when only several units are involved, high-volume applications demand careful study of the economic tradeoffs.

– For example, permanent-magnet DC motors operate closed loop and the cost of an encoder can equal if not exceed the motor cost. In addition, stepper and brushless motors have electronic expenses greater than brushed motor electronic expenses.

Dynamic System Response

- Solution of Linear, Constant-Coefficient, Ordinary Differential Equations
  - Classical Operator Method
  - Laplace Transform Method
- Laplace Transform Properties
- 1st-Order Dynamic System Time and Frequency Response
- 2nd-Order Dynamic System Time and Frequency Response

Laplace Transform Methods

- A basic mathematical model used in many areas of engineering is the linear ordinary differential equation with constant coefficients:

$$a_n \frac{d^n q_o}{dt^n} + a_{n-1} \frac{d^{n-1} q_o}{dt^{n-1}} + \cdots + a_1 \frac{dq_o}{dt} + a_0 q_o =$$
$$b_m \frac{d^m q_i}{dt^m} + b_{m-1} \frac{d^{m-1} q_i}{dt^{m-1}} + \cdots + b_1 \frac{dq_i}{dt} + b_0 q_i$$

- q_o is the output (response) variable of the system
- q_i is the input (excitation) variable of the system
- a_n and b_m are the physical parameters of the system

- Straightforward analytical solutions are available no matter how high the order *n* of the equation.
- Review of the classical operator method for solving linear differential equations with constant coefficients will be useful. When the input $q_i(t)$ is specified, the right hand side of the equation becomes a known function of time, $f(t)$.
- The classical operator method of solution is a three-step procedure:
  - Find the complimentary (homogeneous) solution q_{oc} for the equation with $f(t) = 0$.
  - Find the particular solution q_{op} with $f(t)$ present.
  - Get the complete solution $q_o = q_{oc} + q_{op}$ and evaluate the constants of integration by applying known initial conditions.

- *Step 1*
  - To find q_{oc}, rewrite the differential equation using the differential operator notation $D = d/dt$, treat the equation as if it were algebraic, and write the system characteristic equation as:

$$a_n D^n + a_{n-1} D^{n-1} + \cdots + a_1 D + a_0 = 0$$

  - Treat this equation as an algebraic equation in the unknown *D* and solve for the *n* roots (eigenvalues) $s_1, s_2, \ldots, s_n$. Since root finding is a rapid computerized operation, we assume all the roots are available and now we state rules that allow one to immediately write down q_{oc}:
  - Real, unrepeated root s_1:

$$q_{oc} = c_1 e^{s_1 t}$$

  - Real root s_2 repeated *m* times:

$$q_{oc} = c_0 e^{s_2 t} + c_1 t e^{s_2 t} + c_2 t^2 e^{s_2 t} + \cdots + c_m t^m e^{s_2 t}$$

– When the *a's* are real numbers, then any complex roots that might appear always come in pairs $a \pm ib$:

$$q_{oc} = ce^{at}\sin(bt + \phi)$$

– For repeated root pairs $a \pm ib$, $a \pm ib$, and so forth, we have:

$$q_{oc} = c_0 e^{at}\sin(bt + \phi_0) + c_1 t e^{at}\sin(bt + \phi_1) + \cdots$$

– The *c's* and *f's* are constants of integration whose numerical values cannot be found until the last step.

- *Step 2*
  - The particular solution q_{op} takes into account the "forcing function" *f(t)* and methods for getting the particular solution depend on the form of *f(t)*.
  - The method of undetermined coefficients provides a simple method of getting particular solutions for most *f(t)'s* of practical interest.
  - To check whether this approach will work, differentiate *f(t)* over and over. If repeated differentiation ultimately leads to zeros, or else to repetition of a finite number of different time functions, then the method will work.
  - The particular solution will then be a sum of terms made up of each different type of function found in *f(t)* and all its derivatives, each term multiplied by an unknown constant (undetermined coefficient).

- If *f(t)* or one of its derivatives contains a term identical to a term in q_{oc}, the corresponding terms should be multiplied by t.
- This particular solution is then substituted into the differential equation making it an identity. Gather like terms on each side, equate their coefficients, and obtain a set of simultaneous algebraic equations that can be solved for all the undetermined coefficients.

Step 3

- We now have q_{oc} (with n unknown constants) and q_{op} (with no unknown constants). The complete solution $q_o = q_{oc} + q_{op}$. The initial conditions are then applied to find the n unknown constants.

- Certain advanced analysis methods are most easily developed through the use of the Laplace Transform.
- A transformation is a technique in which a function is transformed from dependence on one variable to dependence on another variable. Here we will transform relationships specified in the time domain into a new domain wherein the axioms of algebra can be applied rather than the axioms of differential or difference equations.
- The transformations used are the Laplace transformation (differential equations) and the Z transformation (difference equations).
- The Laplace transformation results in functions of the time variable t being transformed into functions of the frequency-related variable s.

- The Z transformation is a direct outgrowth of the Laplace transformation and the use of a modulated train of impulses to represent a sampled function mathematically.
- The Z transformation allows us to apply the frequency-domain analysis and design techniques of continuous control theory to discrete control systems.
- One use of the Laplace Transform is as an alternative method for solving linear differential equations with constant coefficients. Although this method will not solve any equations that cannot be solved also by the classical operator method, it presents certain advantages:
  - Separate steps to find the complementary solution, particular solution, and constants of integration are not used. The complete solution, including initial conditions, is obtained at once.

- There is never any question about which initial conditions are needed. In the classical operator method, the initial conditions are evaluated at $t = 0^+$, a time just after the input is applied. For some kinds of systems and inputs, these initial conditions are not the same as those before the input is applied, so extra work is required to find them. The Laplace Transform method uses the conditions before the input is applied; these are generally physically known and are often zero, simplifying the work.
- For inputs that cannot be described by a single formula for their entire course, but must be defined over segments of time, the classical operator method requires a piecewise solution with tedious matching of final conditions of one piece with initial conditions of the next. The Laplace Transform method handles such discontinuous inputs very neatly.
- The Laplace Transform method allows the use of graphical techniques for predicting system performance without actually solving system differential equations.

- All theorems and techniques of the Laplace Transform derive from the fundamental definition for the direct Laplace Transform *F(s)* of the time function *f(t)*:

$$L\left[f(t)\right] = F(s) = \int_0^{\infty} f(t) e^{-st} dt \quad t > 0$$

$$s = \text{ a complex variable } = \sigma + i\omega$$

- This integral cannot be evaluated for all *f(t)'s*, but when it can, it establishes a unique pair of functions, *f(t)* in the time domain and its companion *F(s)* in the *s* domain. Comprehensive tables of Laplace Transform pairs are available. Signals we can physically generate always have corresponding Laplace transforms. When we use the Laplace Transform to solve differential equations, we must transform entire equations, not just isolated *f(t)* functions, so several theorems are necessary for this.

- Linearity Theorem:

$$L\left[a_1 f_1(t) + a_2 f_2(t)\right] = L\left[a_1 f_1(t)\right] + L\left[a_2 f_2(t)\right] = a_1 F_1(s) + a_2 F_2(s)$$

- Differentiation Theorem:

$$L\left[\frac{df}{dt}\right] = sF(s) - f(0)$$

$$L\left[\frac{d^2 f}{dt^2}\right] = s^2 F(s) - sf(0) - \frac{df}{dt}(0)$$

$$L\left[\frac{d^n f}{dt^n}\right] = s^n F(s) - s^{n-1} f(0) - s^{n-2}\frac{df}{dt}(0) - \cdots - \frac{d^{n-1} f}{dt^{n-1}}(0)$$

  – *f(0), (df/dt)(0),* etc., are initial values of *f(t)* and its derivatives evaluated numerically at a time instant *before* the driving input is applied.

- Integration Theorem:

$$L\left[\int f(t)dt\right] = \frac{F(s)}{s} + \frac{f^{(-1)}(0)}{s}$$

$$L\left[f^{(-n)}(t)\right] = \frac{F(s)}{s^n} + \sum_{k=1}^{n} \frac{f^{(-k)}(0)}{s^{n-k+1}}$$

where $f^{(-n)}(t) = \int \cdots \int f(t)(dt)^n$ and $f^{(-0)}(t) = f(t)$

  – Again, the initial values of *f(t)* and its integrals are evaluated numerically at a time instant *before* the driving input is applied.

- Delay Theorem:
  - The Laplace Transform provides a theorem useful for the dynamic system element called *dead time (transport lag)* and for dealing efficiently with discontinuous inputs.

$$\left.\begin{aligned} u(t) &= 1.0 \Rightarrow t > 0 \\ u(t) &= 0 \Rightarrow t < 0 \\ u(t-a) &= 1.0 \Rightarrow t > a \\ u(t-a) &= 0 \Rightarrow t < a \end{aligned}\right\} \quad L\left[f(t-a)u(t-a)\right] = e^{-as}F(s)$$

- Final Value Theorem:
  - If we know $Q_0(s)$, $q_0(\infty)$ can be found quickly without doing the complete inverse transform by use of the *final value theorem*.

$$\lim_{t\to\infty} f(t) = \lim_{s\to 0} sF(s)$$

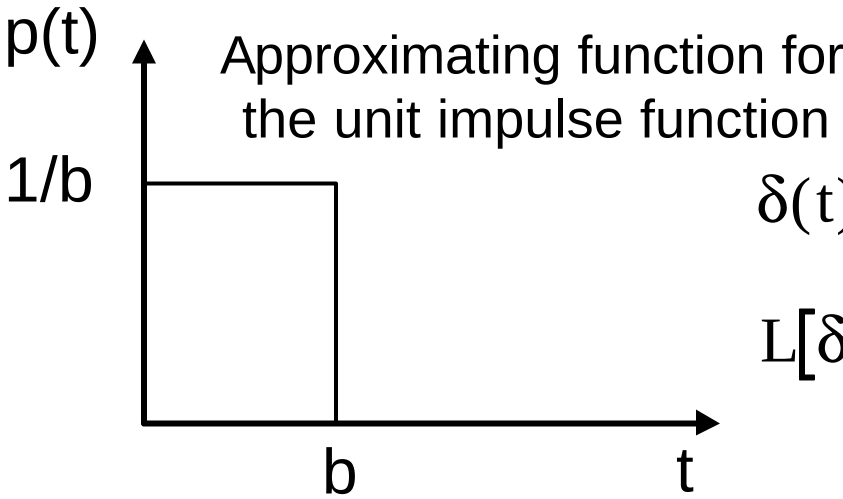
  - This is true if the system and input are such that the output approaches a constant value as t approaches ∞.
- Initial Value Theorem:
  - This theorem is helpful for finding the value of *f(t)* just after the input has been applied, i.e., at $t = 0^+$. In getting the *F(s)* needed to apply this theorem, our usual definition of initial conditions as those *before* the input is applied must be used.

$$\lim_{t\to 0} f(t) = \lim_{s\to\infty} sF(s)$$

- Impulse Function $\delta(t)$

$$\delta(t) = 0 \Rightarrow t \neq 0$$

$$\int_{-\varepsilon}^{+\varepsilon} \delta(t)\,dt = 1 \Rightarrow \varepsilon > 0$$



$$\delta(t) = \lim_{b \to 0} p(t)$$

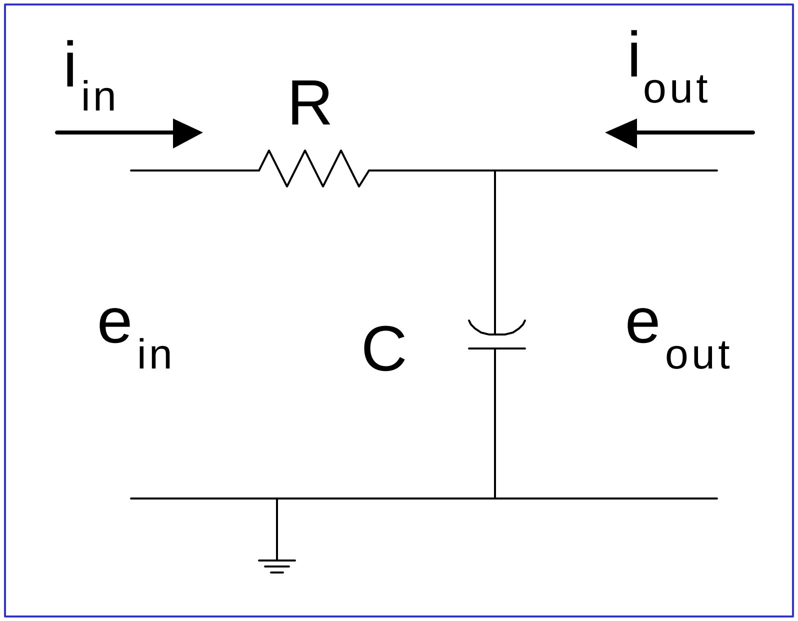
$$L[\delta(t)] = L\left[\frac{du}{dt}\right] = sU(s) = s\frac{1}{s} = 1.0$$

- The step function is the integral of the impulse function, or conversely, the impulse function is the derivative of the step function.
- When we multiply the impulse function by some number, we increase the "strength of the impulse", but "strength" now means area, not height as it does for "ordinary" functions.

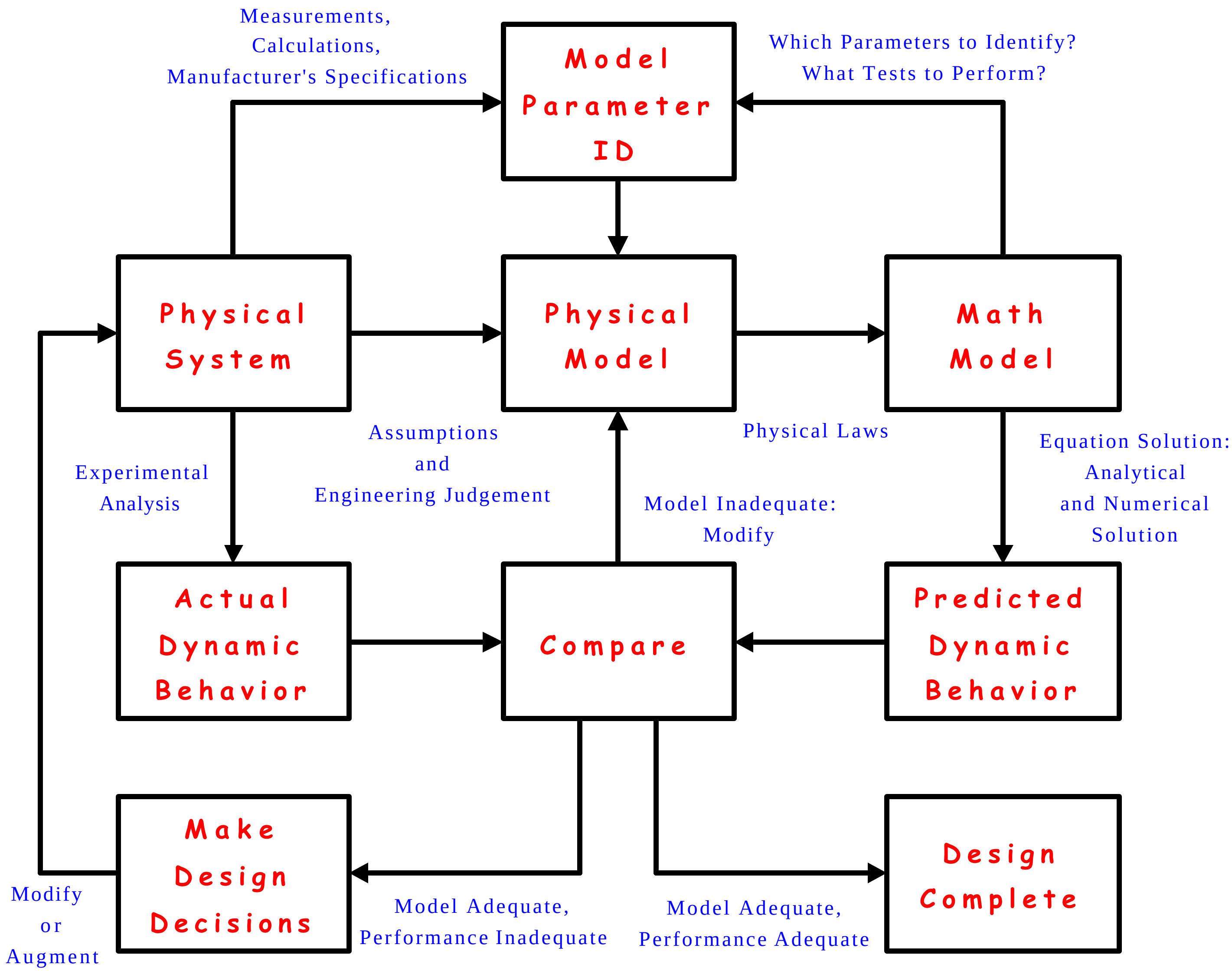
- An impulse that has an infinite magnitude and zero duration is mathematical fiction and does not occur in physical systems.
- If, however, the magnitude of a pulse input to a system is very large and its duration is very short compared to the system time constants, then we can approximate the pulse input by an impulse function.
- The impulse input supplies energy to the system in an infinitesimal time.

- Inverse Laplace Transformation
  - A convenient method for obtaining the inverse Laplace transform is to use a table of Laplace transforms. In this case, the Laplace transform must be in a form immediately recognizable in such a table.
  - If a particular transform F(s) cannot be found in a table, then we may expand it into partial fractions and write F(s) in terms of simple functions of s for which inverse Laplace transforms are already known.
  - These methods for finding inverse Laplace transforms are based on the fact that the unique correspondence of a time function and its inverse Laplace transform holds for any continuous time function.

Time Response & Frequency Response
1st-Order Dynamic System
Example: RC Low-Pass Filter



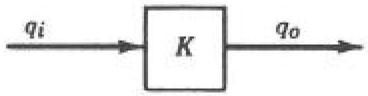
Dynamic System Investigation of the RC Low-Pass Filter



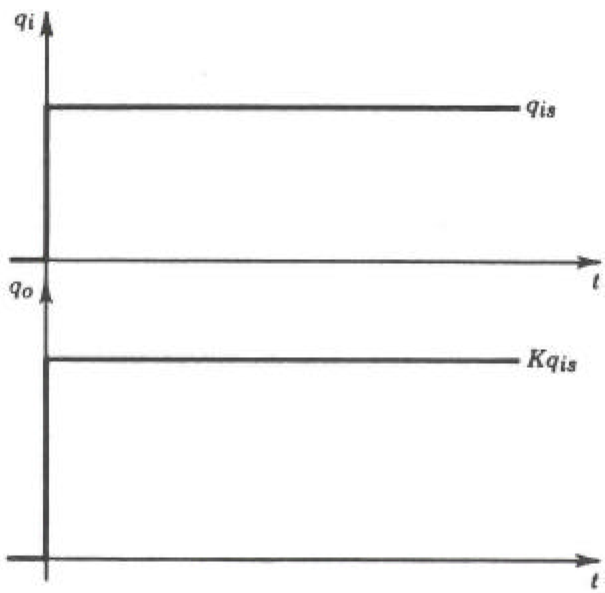
Dynamic System Investigation

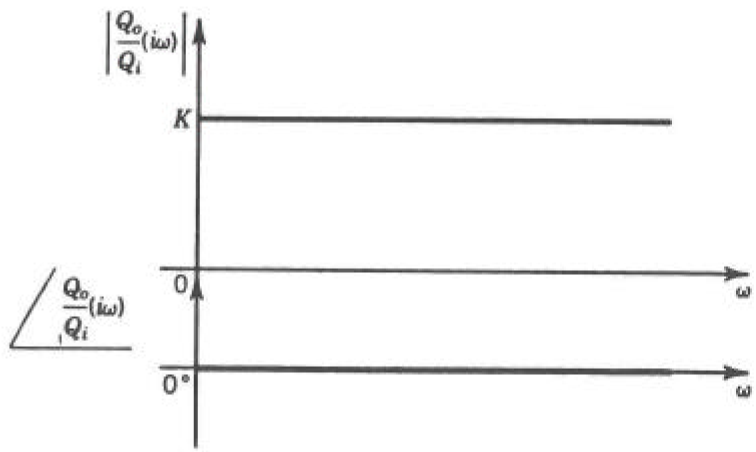
Zero-Order Dynamic System Model

$$q_o = K\, q_i$$

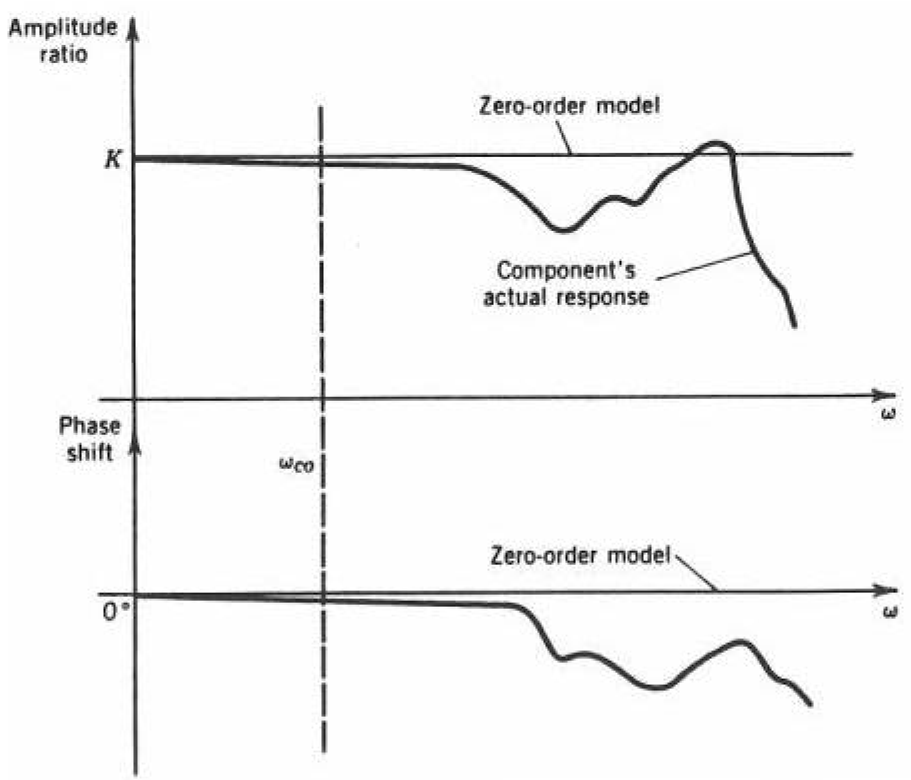


$$\frac{Q_o}{Q_i}(s) = K \qquad \frac{Q_o}{Q_i}(i\omega) = K\angle 0^\circ$$





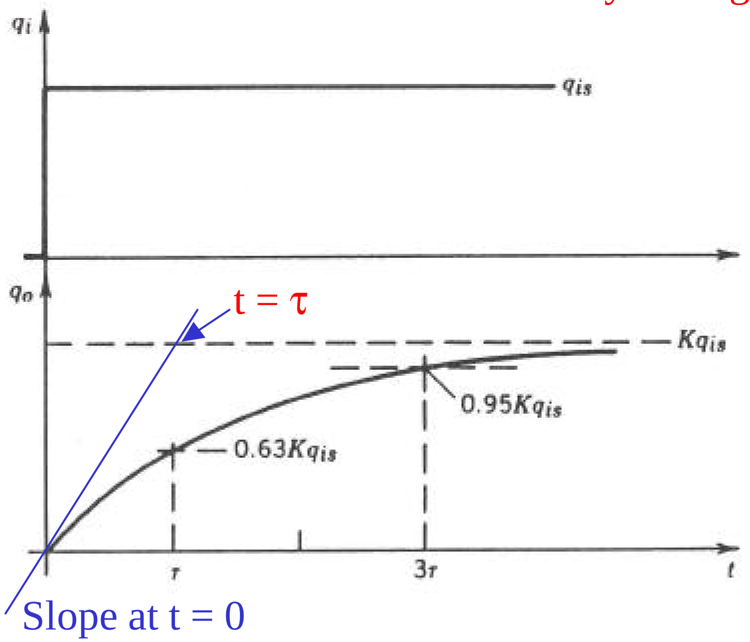
Validation of a Zero-Order Dynamic System Model



1^{st}-Order Dynamic System Model

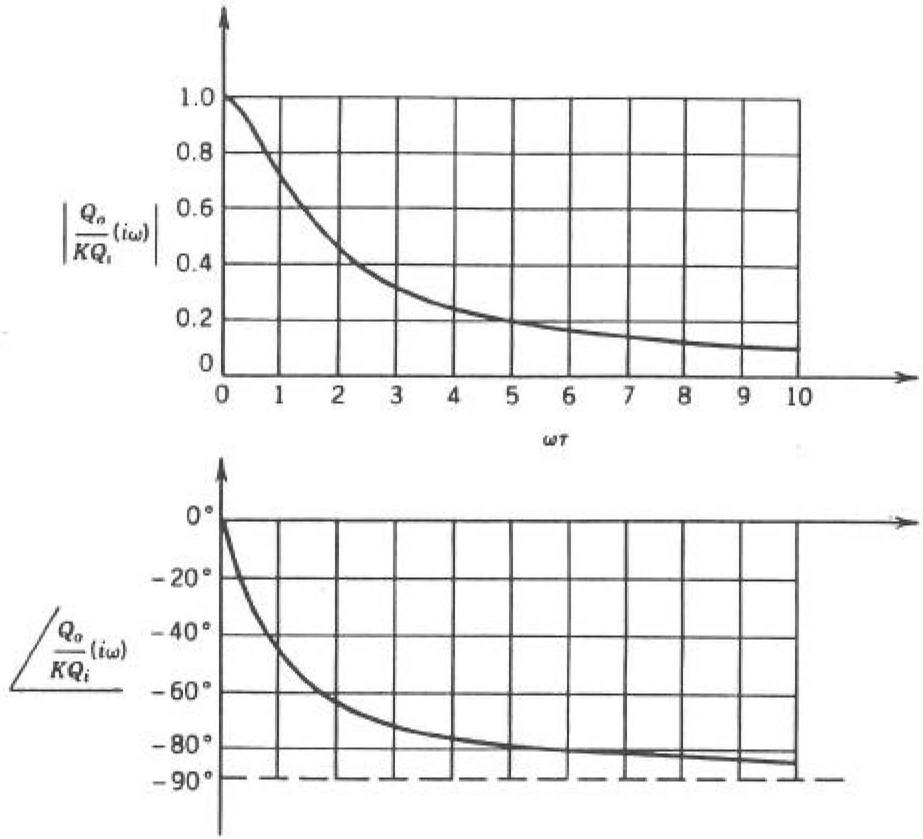
$$\tau \frac{dq_o}{dt} + q_o = Kq_i$$

τ = time constant
K = steady-state gain



$$q_o = Kq_{is}\left(1 - e^{-(t/\tau)}\right)$$

$$\dot{q}_o = \frac{Kq_{is}}{\tau} e^{-\frac{t}{\tau}} \quad \Rightarrow \quad \dot{q}_o\Big|_{t=0} = \frac{Kq_{is}}{\tau}$$



$$\frac{Q_o}{Q_i}(s) = \frac{K}{\tau s + 1} \qquad \frac{Q_o}{Q_i}(i\omega) = \frac{K}{\sqrt{(\omega\tau)^2 + 1}} \angle -\tan^{-1}\omega\tau$$

- How would you determine if an experimentally-determined step response of a system could be represented by a first-order system step response?

$$q_o(t) = Kq_{is}\left(1 - e^{-\frac{t}{\tau}}\right)$$

$$\frac{q_o(t) - Kq_{is}}{Kq_{is}} = -e^{-\frac{t}{\tau}}$$

$$1 - \frac{q_o(t)}{Kq_{is}} = e^{-\frac{t}{\tau}}$$

$$\log_{10}\left[1 - \frac{q_o(t)}{Kq_{is}}\right] = -\frac{t}{\tau}\log_{10} e = -0.4343\frac{t}{\tau}$$

Straight-Line Plot:

$$\log_{10}\left[1 - \frac{q_o(t)}{Kq_{is}}\right] \text{ vs. } t$$

Slope = $-0.4343/\tau$

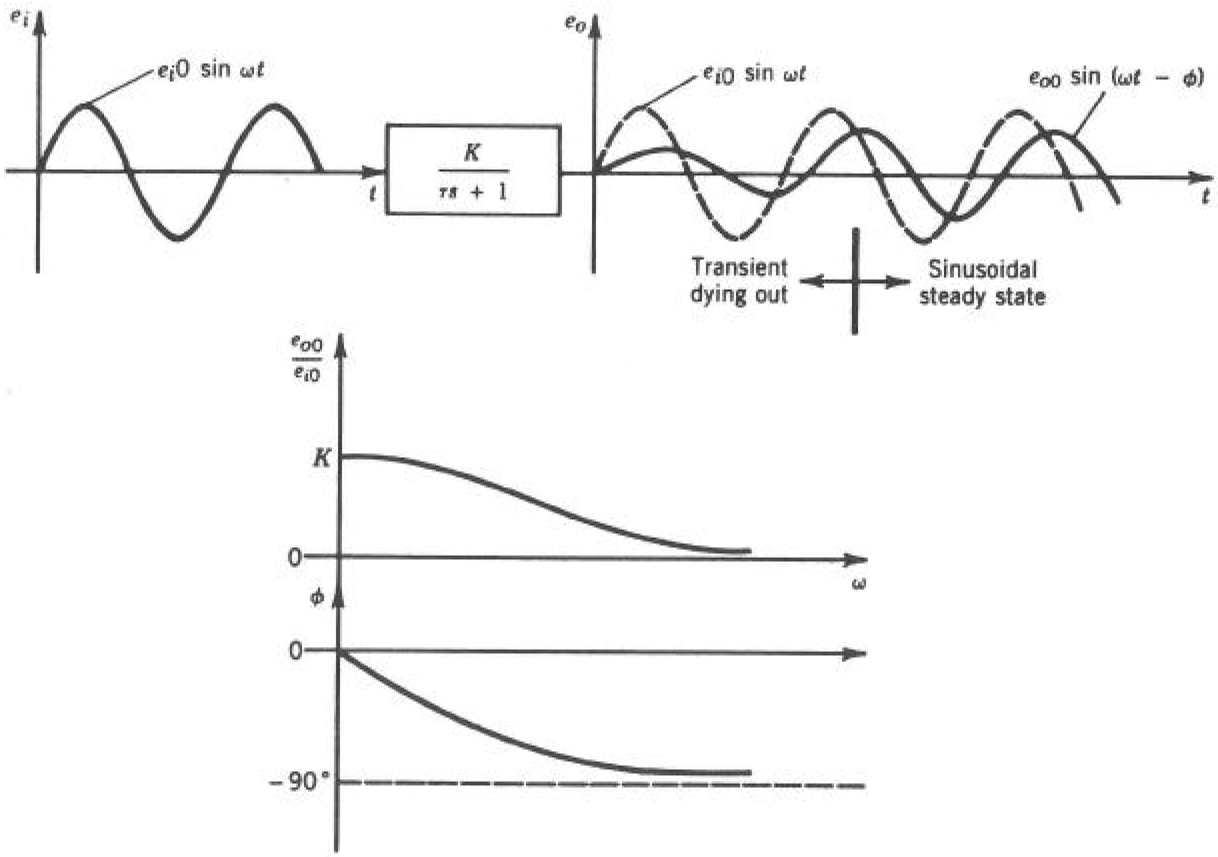
- This approach gives a more accurate value of τ since the best line through all the data points is used rather than just two points, as in the 63.2% method. Furthermore, if the data points fall nearly on a straight line, we are assured that the instrument is behaving as a first-order type. If the data deviate considerably from a straight line, we know the system is not truly first order and a τ value obtained by the 63.2% method would be quite misleading.
- An even stronger verification (or refutation) of first-order dynamic characteristics is available from frequency-response testing. If the system is truly first-order, the amplitude ratio follows the typical low- and high-frequency asymptotes (slope 0 and –20 dB/decade) and the phase angle approaches -90° asymptotically.

– If these characteristics are present, the numerical value of τ is found by determining ω (rad/sec) at the breakpoint and using $\tau = 1/\omega_{break}$. Deviations from the above amplitude and/or phase characteristics indicate non-first-order behavior.

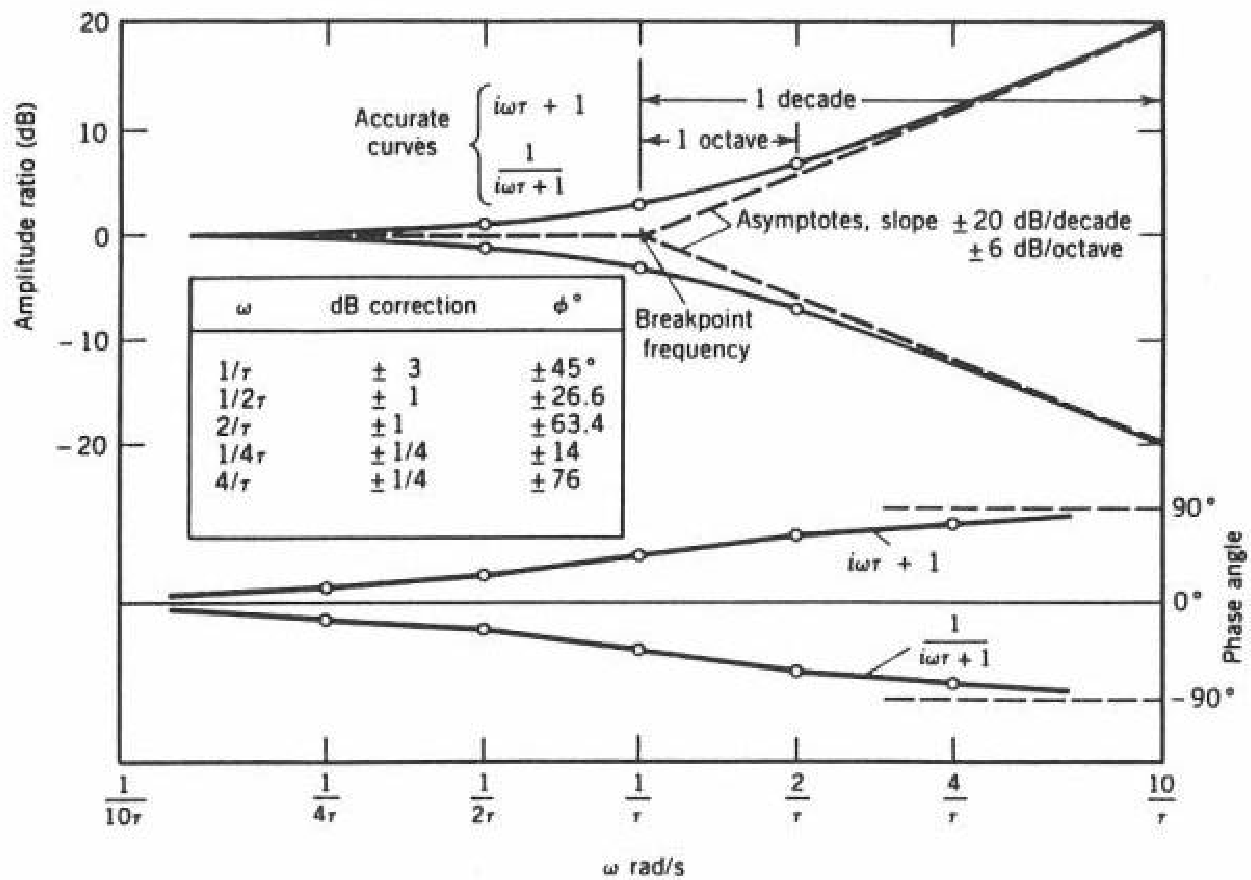
- What is the relationship between the unit-step response and the unit-ramp response and between the unit-impulse response and the unit-step response?
  - For a linear time-invariant system, the response to the derivative of an input signal can be obtained by differentiating the response of the system to the original signal.
  - For a linear time-invariant system, the response to the integral of an input signal can be obtained by integrating the response of the system to the original signal and by determining the integration constants from the zero-output initial condition.

- Unit-Step Input is the derivative of the Unit-Ramp Input.
- Unit-Impulse Input is the derivative of the Unit-Step Input.
- Once you know the unit-step response, take the derivative to get the unit-impulse response and integrate to get the unit-ramp response.

System Frequency Response



Bode Plotting of 1^{st}-Order Frequency Response

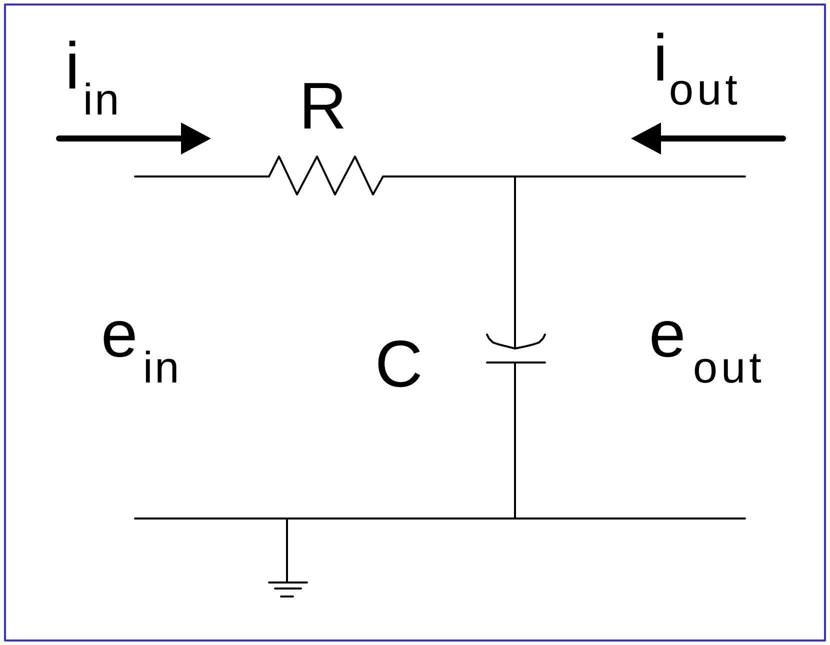


dB = 20 $\log_{10}$ (amplitude ratio)

decade = 10 to 1 frequency change

octave = 2 to 1 frequency change

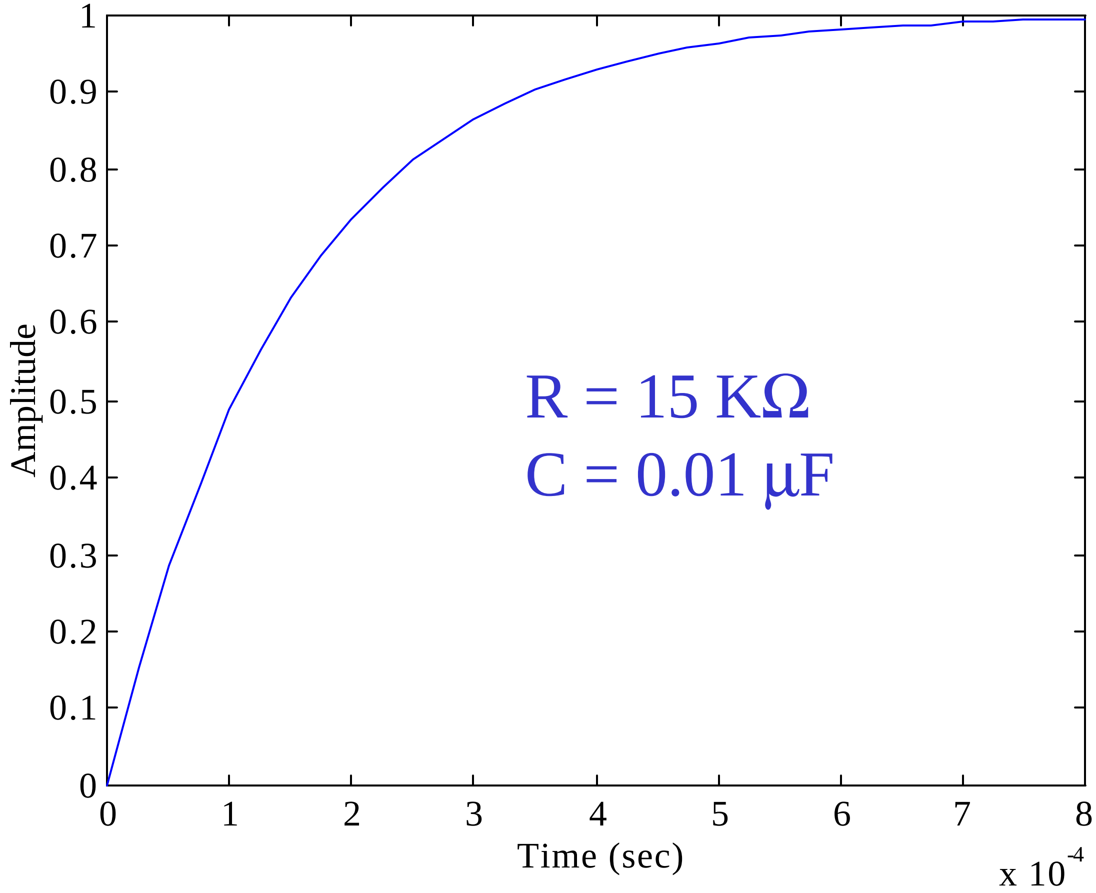
Analog Electronics:
RC Low-Pass Filter Time Response & Frequency Response



$$\begin{bmatrix} e_{in} \\ i_{in} \end{bmatrix} = \begin{bmatrix} RCs+1 & -R \\ Cs & -1 \end{bmatrix} \begin{bmatrix} e_{out} \\ i_{out} \end{bmatrix}$$

$$\frac{e_{out}}{e_{in}} = \frac{1}{RCs+1} = \frac{1}{\tau s+1} \qquad \text{when } i_{out} = 0$$

Time Response to Unit Step Input



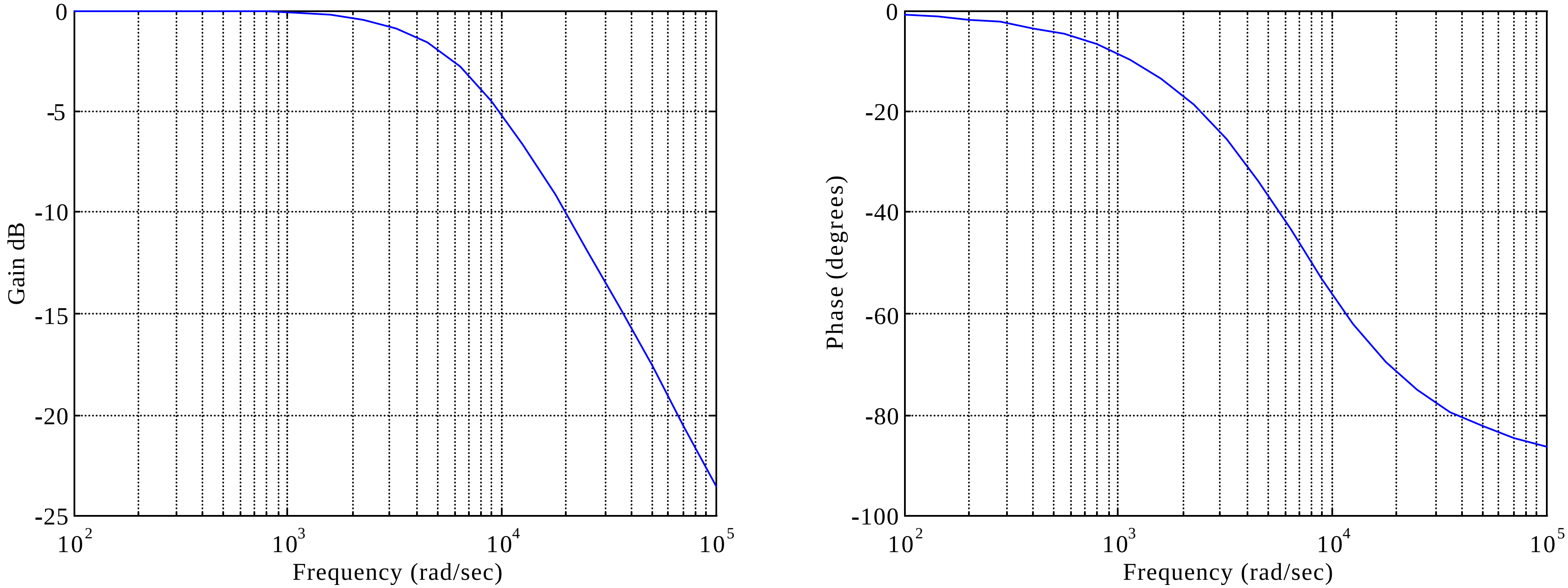
Time Constant $\tau = RC$

- Time Constant τ
  - Time it takes the step response to reach 63% of the steady-state value
- Rise Time $T_r = 2.2\,\tau$
  - Time it takes the step response to go from 10% to 90% of the steady-state value
- Delay Time $T_d = 0.69\,\tau$
  - Time it takes the step response to reach 50% of the steady-state value

Frequency Response

R = 15 KΩ

C = 0.01 μF



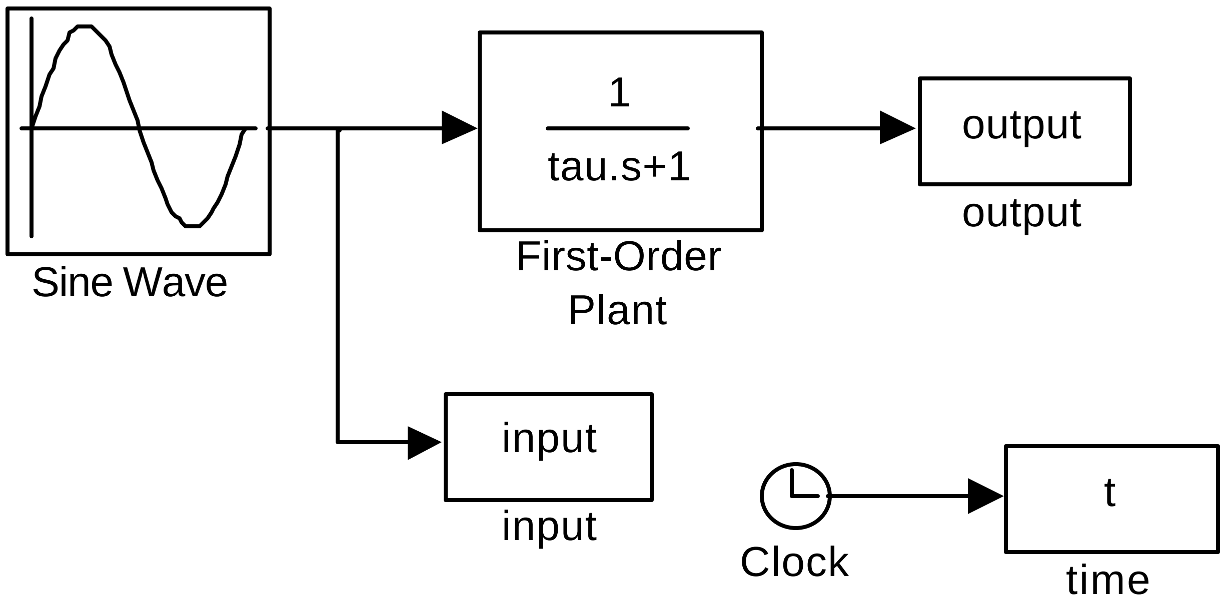
Bandwidth = 1/τ

$$\frac{e_{out}}{e_{in}}(i\omega) = \frac{K}{i\omega\tau + 1} = \frac{K\angle 0^\circ}{\sqrt{(\omega\tau)^2 + 1^2}\angle \tan^{-1}\omega\tau} = \frac{K}{\sqrt{(\omega\tau)^2 + 1^2}}\angle -\tan^{-1}\omega\tau$$

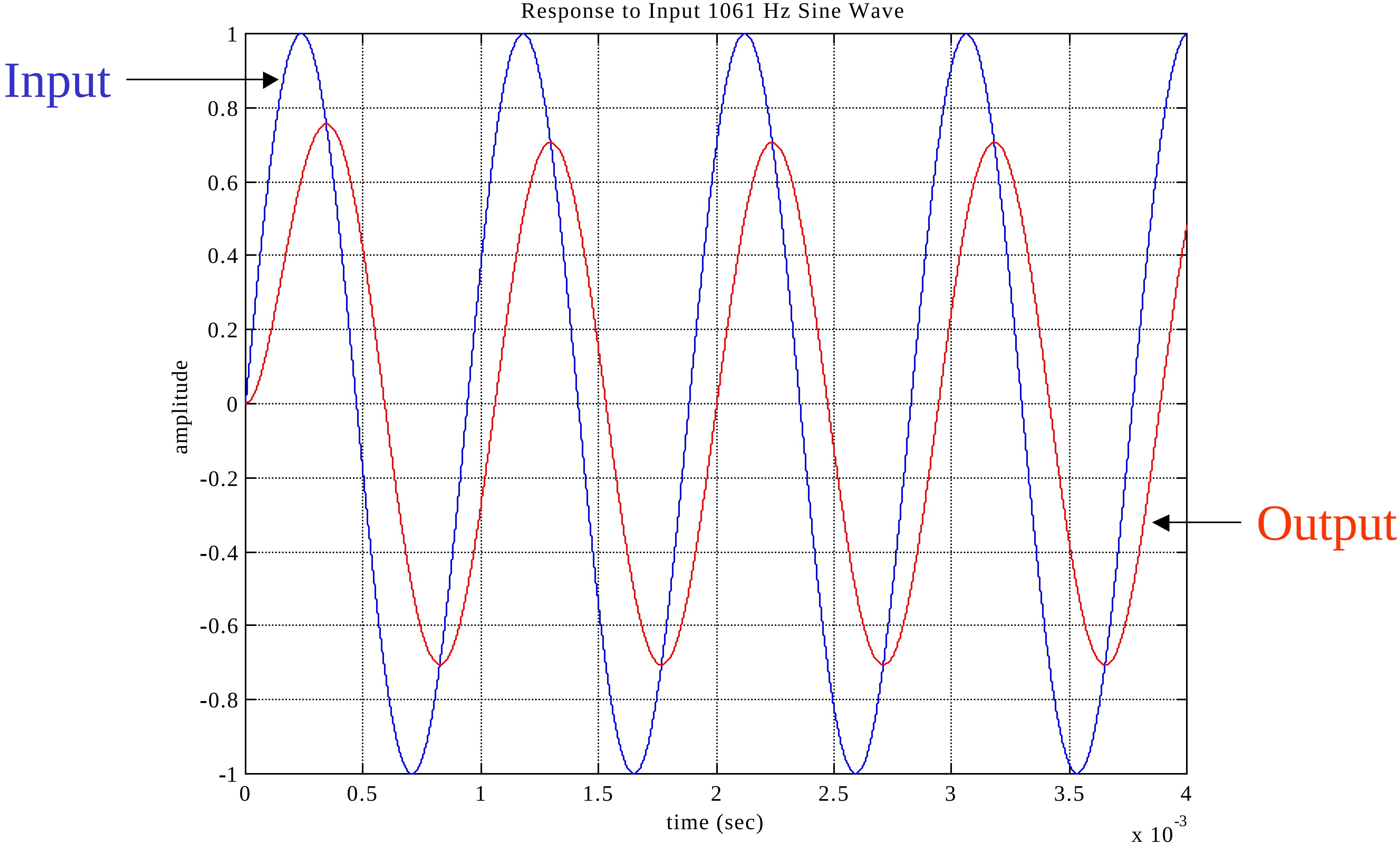
- Bandwidth
  - The bandwidth is the frequency where the amplitude ratio drops by a factor of 0.707 = -3dB of its gain at zero or low-frequency.
  - For a 1^{st} -order system, the bandwidth is equal to $1/\tau$.
  - The larger (smaller) the bandwidth, the faster (slower) the step response.
  - Bandwidth is a direct measure of system susceptibility to noise, as well as an indicator of the system speed of response.

MatLab / Simulink Diagram
Frequency Response for 1061 Hz Sine Input

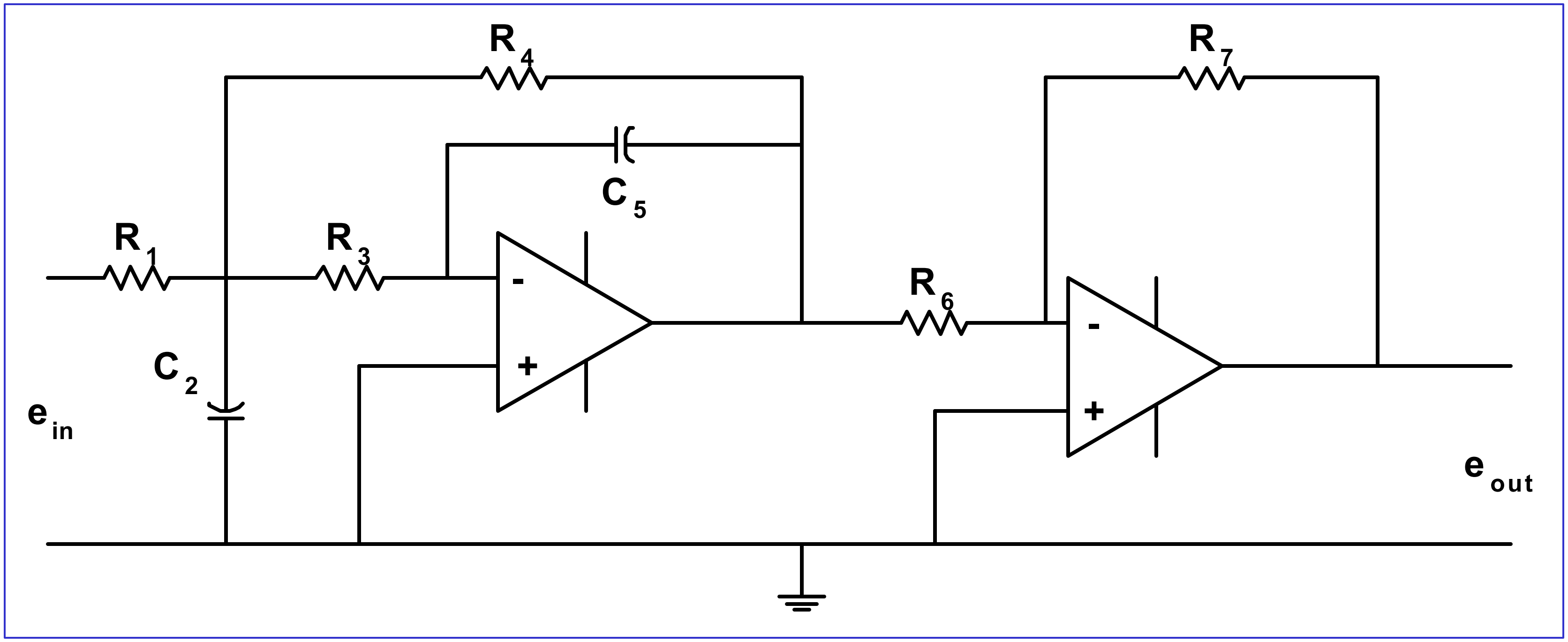
$\tau = 1.5\text{E-}4$ sec



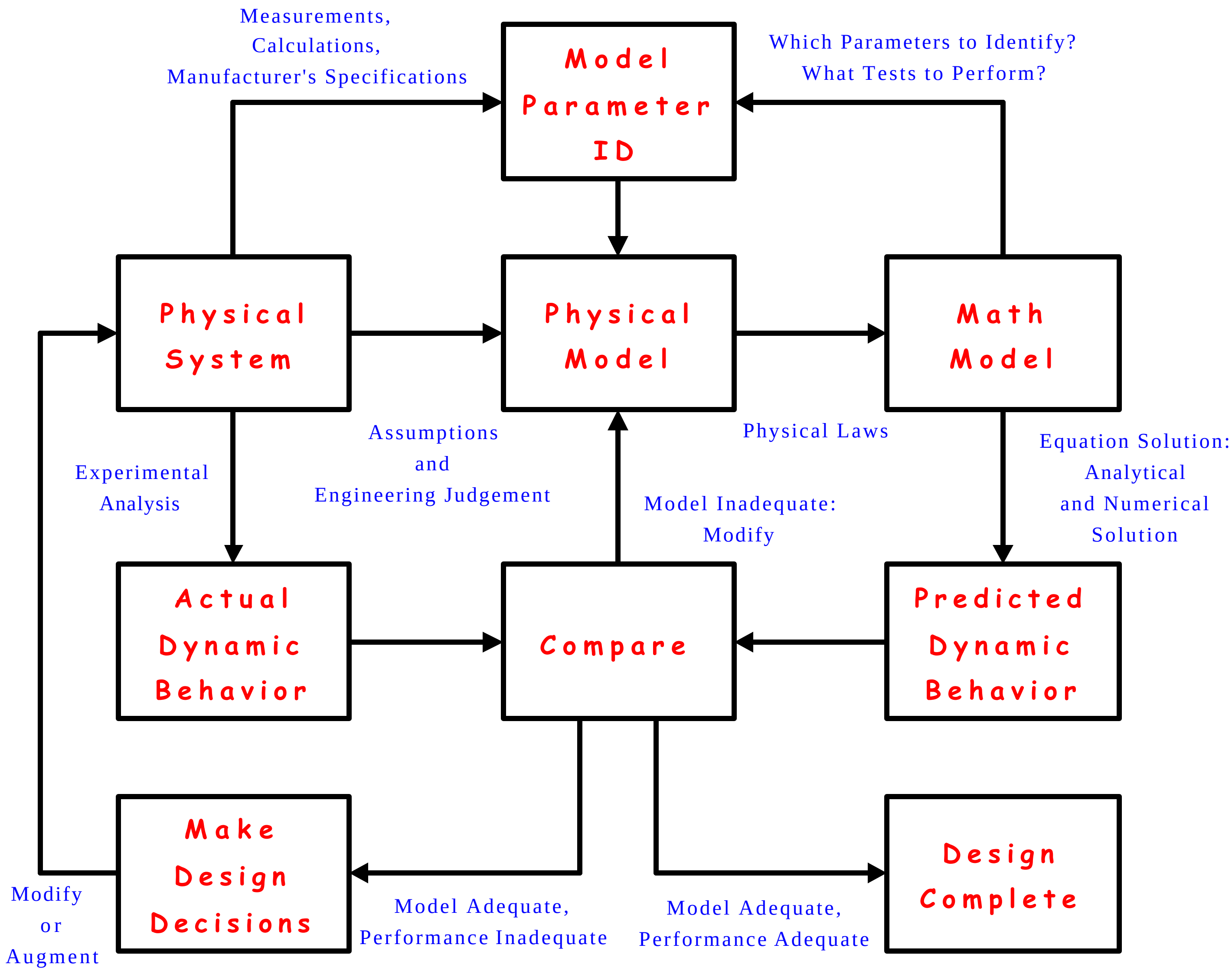
Amplitude Ratio = 0.707 = -3 dB   Phase Angle = -45°



Time Response & Frequency Response
2^{nd}-Order Dynamic System
Example: 2-Pole, Low-Pass, Active Filter



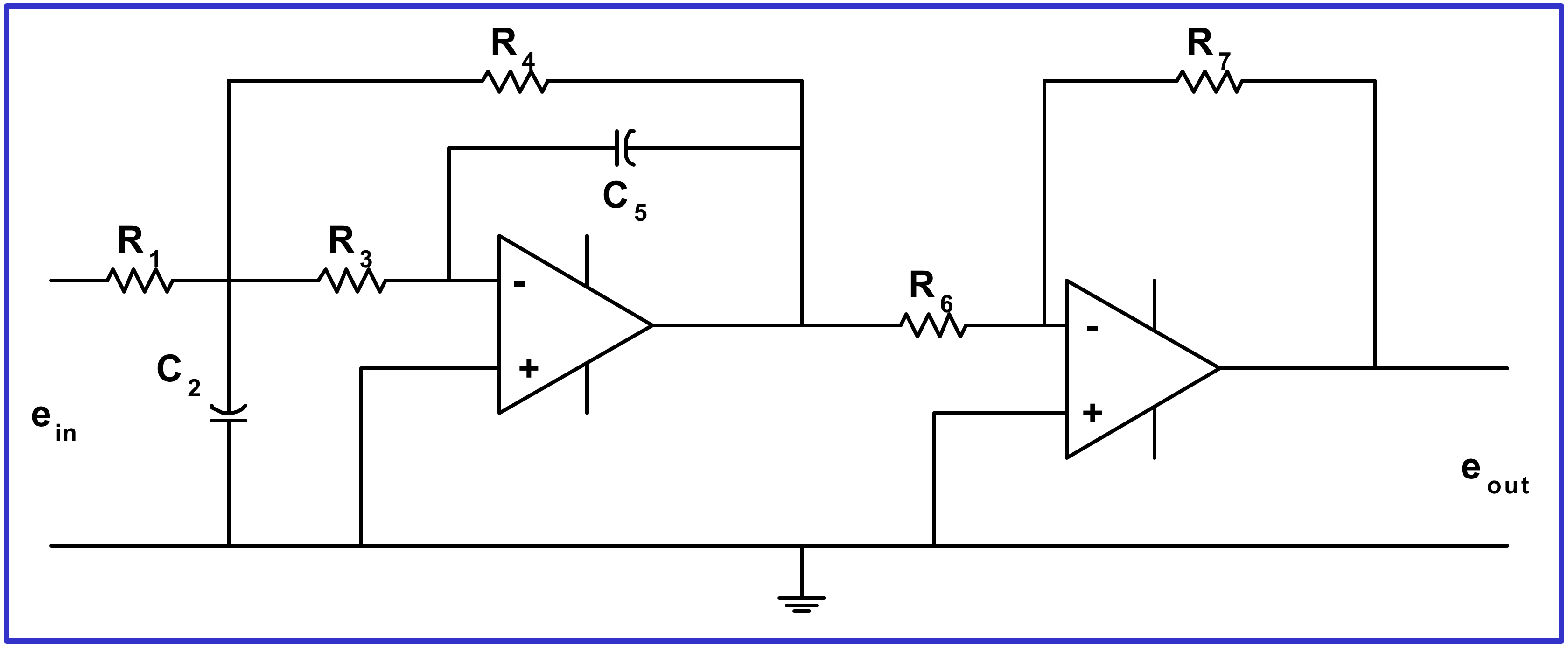
Dynamic System Investigation
of the Two-Pole, Low-Pass, Active Filter



Dynamic System Investigation

Physical Model Ideal Transfer Function

$$\frac{e_{out}}{e_{in}}(s) = \frac{\left(\frac{R_7}{R_6}\right)\left(\frac{1}{R_1R_3C_2C_5}\right)}{s^2 + \left(\frac{1}{R_3C_2} + \frac{1}{R_1C_2} + \frac{1}{R_4C_2}\right)s + \frac{1}{R_3R_4C_2C_5}}$$



2nd-Order Dynamic System Model

$$a_2 \frac{d^2 q_0}{dt^2} + a_1 \frac{dq_0}{dt} + a_0 q_0 = b_0 q_i$$

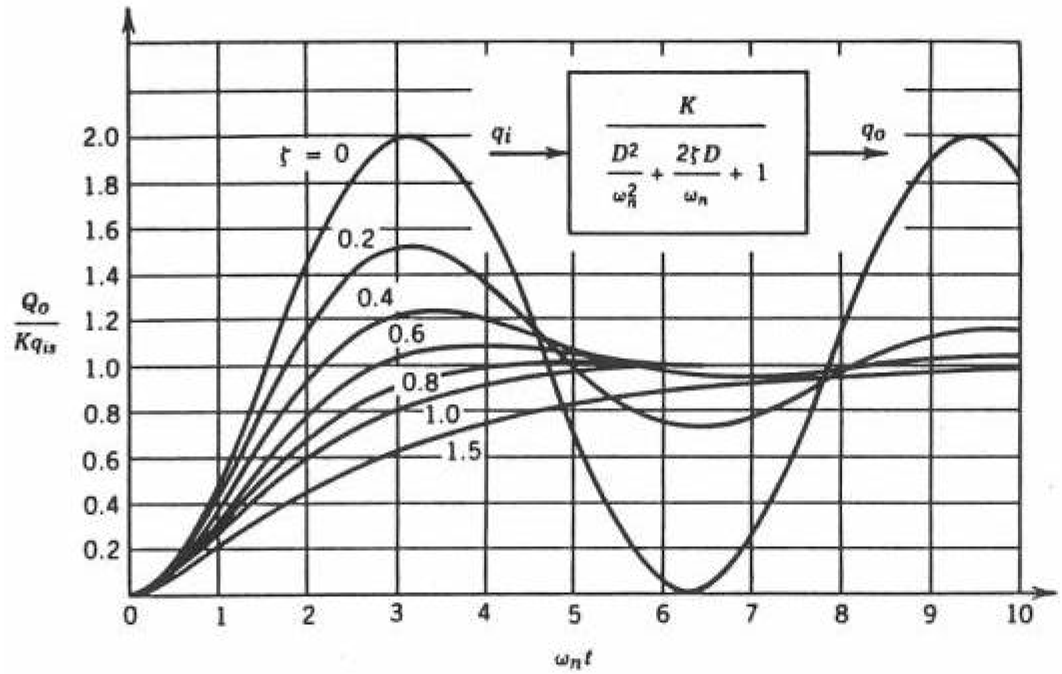
$$\frac{1}{\omega_n^2}\frac{d^2 q_0}{dt^2} + \frac{2\zeta}{\omega_n}\frac{dq_0}{dt} + q_0 = K q_i$$

$$\omega_n \triangleq \sqrt{\frac{a_0}{a_2}} = \text{undamped natural frequency}$$

$$\zeta \triangleq \frac{a_1}{2\sqrt{a_2 a_0}} = \text{damping ratio}$$

$$K \triangleq \frac{b_0}{a_0} = \text{steady-state gain}$$

Step Response of a 2nd-Order System



$$\frac{1}{\omega_n^2}\frac{d^2 q_0}{dt^2}+\frac{2\varsigma}{\omega_n}\zeta\frac{dq_0}{dt}+q_0 = Kq_i$$

Step Response of a 2nd-Order System

Underdamped

$$q_o = Kq_{is}\left[1-\frac{1}{\sqrt{1-\zeta^2}}e^{-\zeta\omega_n t}\sin\left(\omega_n\sqrt{1-\zeta^2}\,t+\sin^{-1}\sqrt{1-\zeta^2}\right)\right] \qquad \zeta<1$$

Critically Damped

$$q_o = Kq_{is}\left[1-\left(1+\omega_n t\right)e^{-\omega_n t}\right] \qquad \zeta=1$$

Over-damped

$$q_o = Kq_{is}\left[1-\frac{\zeta+\sqrt{\zeta^2-1}}{2\sqrt{\zeta^2-1}}e^{\left(-\zeta+\sqrt{\zeta^2-1}\right)\omega_n t}+\frac{\zeta-\sqrt{\zeta^2-1}}{2\sqrt{\zeta^2-1}}e^{\left(-\zeta-\sqrt{\zeta^2-1}\right)\omega_n t}\right] \qquad \zeta>1$$

Frequency Response of a 2^{nd}-Order System

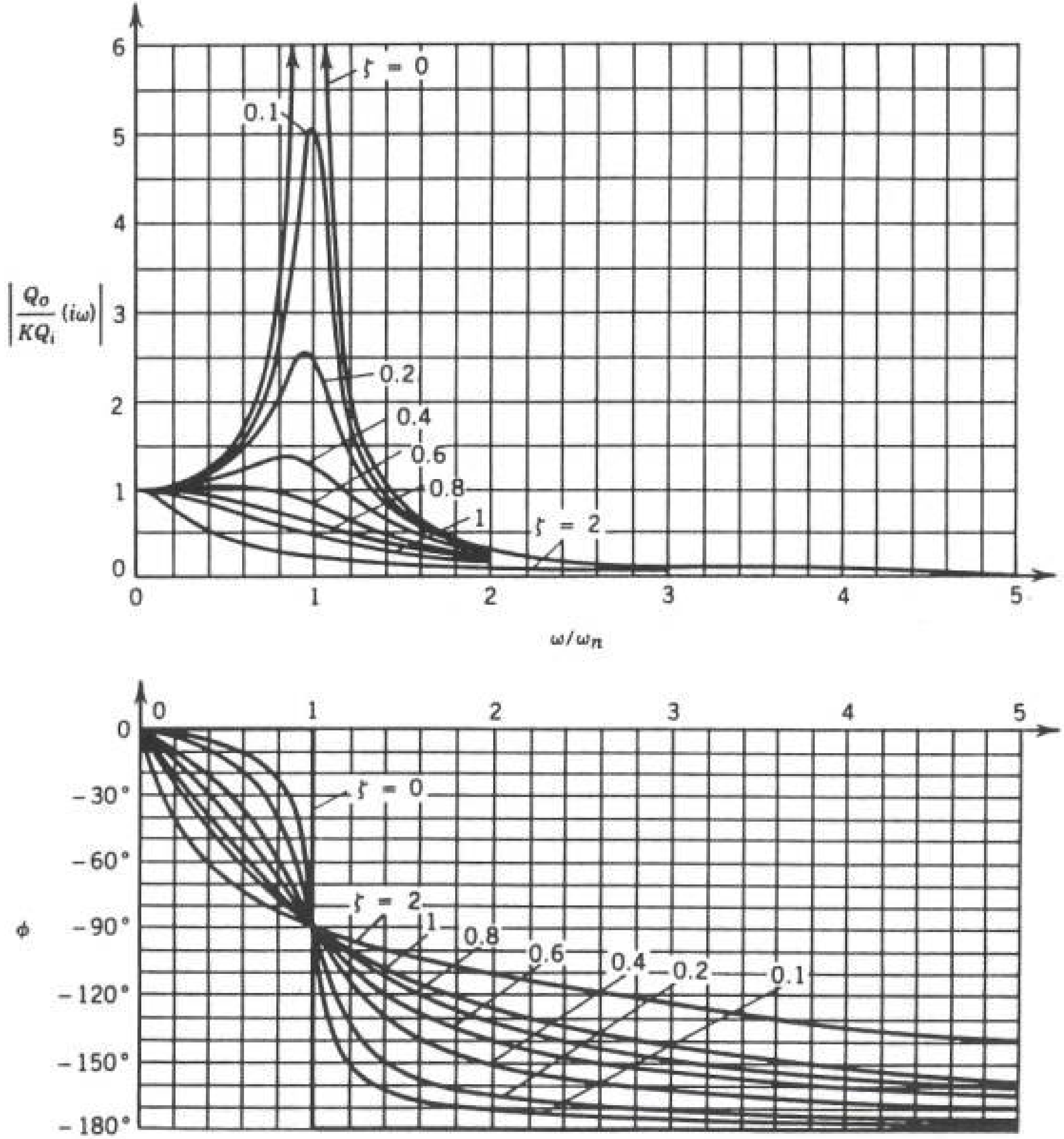
Laplace Transfer Function

$$\frac{Q_o}{Q_i}(s) = \frac{K}{\frac{s^2}{\omega_n^2} + \frac{2\zeta s}{\omega_n} + 1}$$

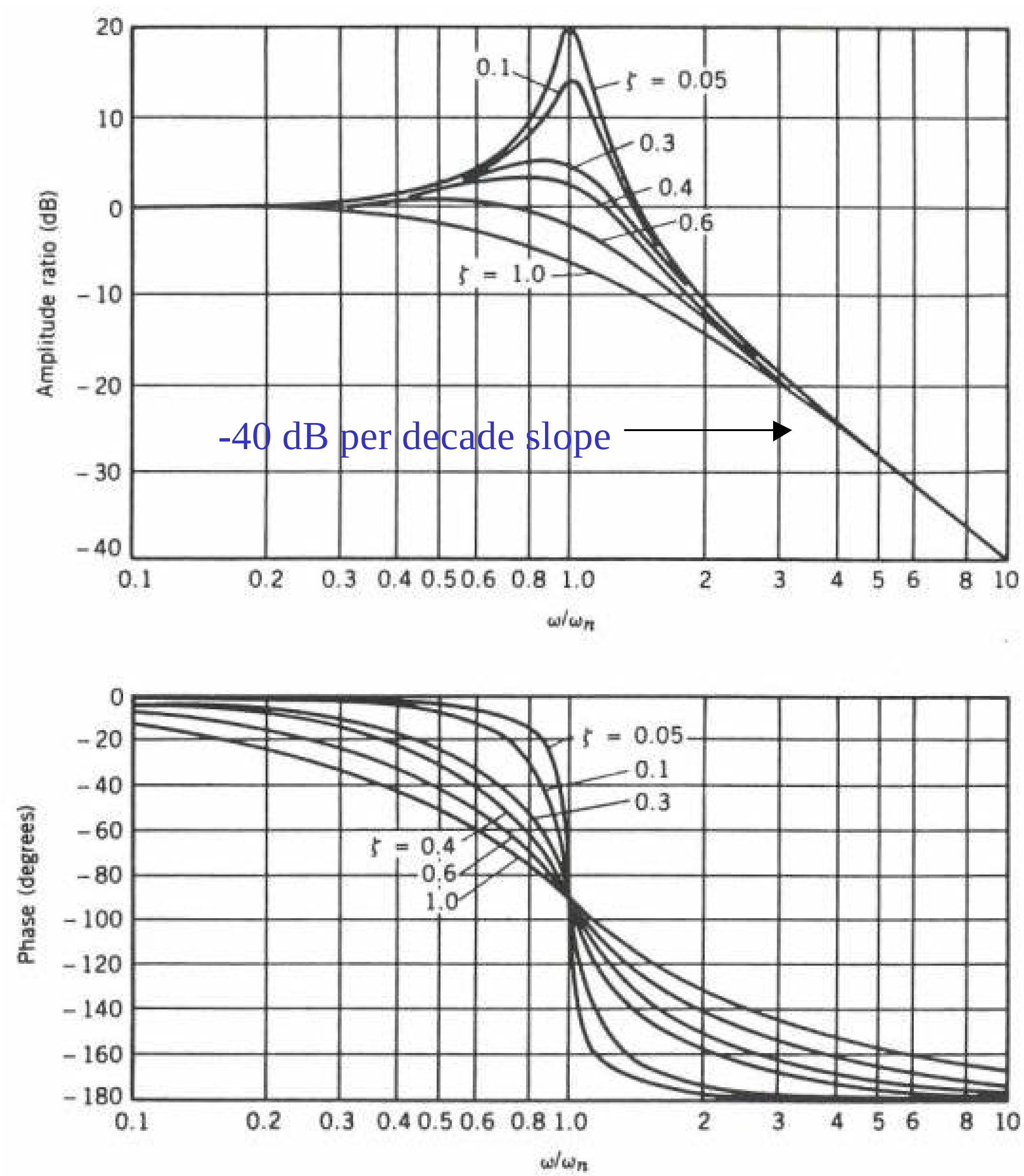
Sinusoidal Transfer Function

$$\frac{Q_o}{Q_i}(i\omega) = \frac{K}{\sqrt{\left[1-\left(\frac{\omega}{\omega_n}\right)^2\right]^2 + \frac{4\zeta^2\omega^2}{\omega_n^2}}} \angle \tan^{-1}\frac{2\zeta}{\left(\frac{\omega}{\omega_n} - \frac{\omega_n}{\omega}\right)}$$

Frequency Response of a 2nd-Order System



Frequency Response of a 2nd-Order System



Some Observations

- When a physical system exhibits a natural oscillatory behavior, a 1^{st}-order model (or even a cascade of several 1^{st}-order models) cannot provide the desired response. The simplest model that does possess that possibility is the 2^{nd}-order dynamic system model.
- This system is very important in control design.
  - System specifications are often given assuming that the system is 2^{nd} order.
  - For higher-order systems, we can often use dominant pole techniques to approximate the system with a 2^{nd}-order transfer function.

- Damping ratio ζ clearly controls oscillation; $\zeta < 1$ is required for oscillatory behavior.
- The undamped case ($\zeta = 0$) is not physically realizable (total absence of energy loss effects) but gives us, mathematically, a sustained oscillation at frequency ω_n.
- Natural oscillations of damped systems are at the damped natural frequency ω_d, and not at ω_n.

$$\omega_d = \omega_n \sqrt{1-\zeta^2}$$

- In hardware design, an optimum value of $\zeta = 0.64$ is often used to give maximum response speed without excessive oscillation.

- Undamped natural frequency ω_n is the major factor in response speed. For a given ζ response speed is directly proportional to ω_n.
- Thus, when 2nd-order components are used in feedback system design, large values of ω_n (small lags) are desirable since they allow the use of larger loop gain before stability limits are encountered.
- For frequency response, a resonant peak occurs for $\zeta < 0.707$. The peak frequency is ω_p and the peak amplitude ratio depends only on ζ.

$$\omega_p = \omega_n \sqrt{1-2\zeta^2} \qquad \text{peak amplitude ratio} = \frac{K}{2\zeta\sqrt{1-\zeta^2}}$$

- Bandwidth
  - The bandwidth is the frequency where the amplitude ratio drops by a factor of 0.707 = -3dB of its gain at zero or low-frequency.
  - For a 1st -order system, the bandwidth is equal to 1/ τ.
  - The larger (smaller) the bandwidth, the faster (slower) the step response.
  - Bandwidth is a direct measure of system susceptibility to noise, as well as an indicator of the system speed of response.
  - For a 2nd-order system:

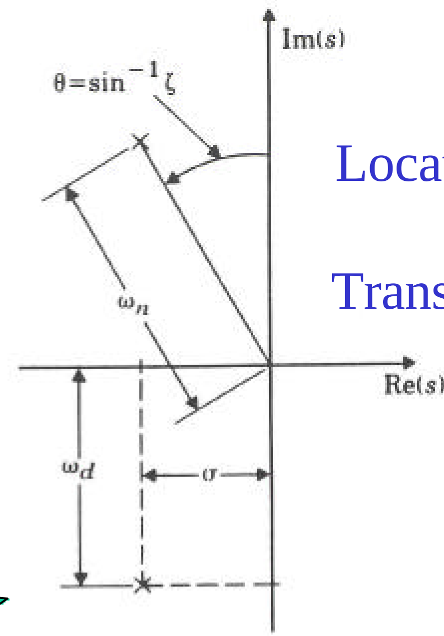
$$\mathrm{BW} = \omega_n \sqrt{1 - 2\zeta^2 + \sqrt{2 - 4\zeta^2 + 4\zeta^4}}$$

– As ζ varies from 0 to 1, BW varies from $1.55\omega_n$ to $0.64\omega_n$. For a value of $\zeta = 0.707$, BW = ω_n. For most design considerations, we assume that the bandwidth of a 2nd-order all pole system can be approximated by ω_n.

$$G(s) = \frac{K\omega_n^2}{s^2 + 2\varsigma\omega_n s + \omega_n^2}$$

$$s_{1,2} = -\varsigma\omega_n \pm i\omega_n\sqrt{1-\varsigma^2}$$

$$s_{1,2} = -\sigma \pm i\omega_d$$



Location of Poles
Of
Transfer Function

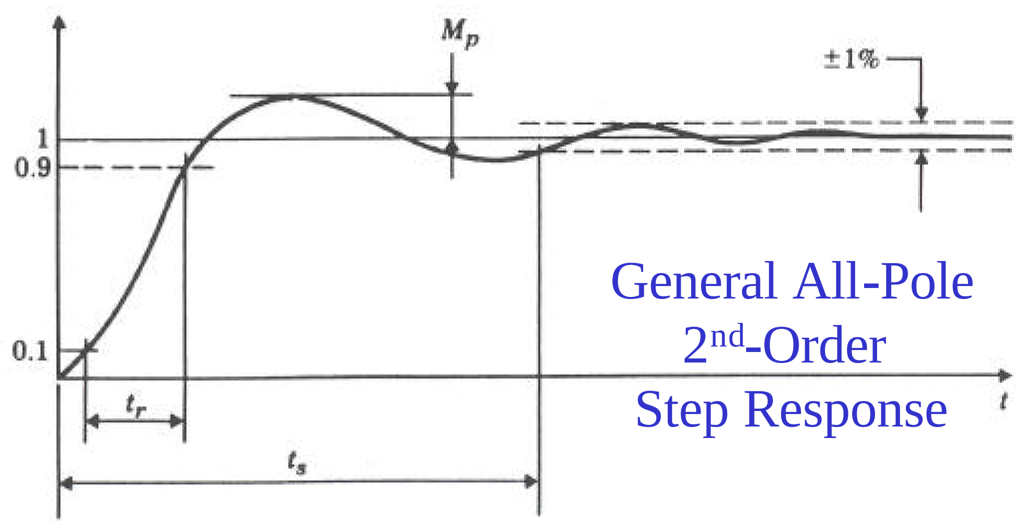
$$y(t) = 1 - e^{-\sigma t}\left(\cos\omega_d t + \frac{\sigma}{\omega_d}\sin\omega_d t\right)$$

$$t_r \approx \frac{1.8}{\omega_n} \quad \text{rise time}$$

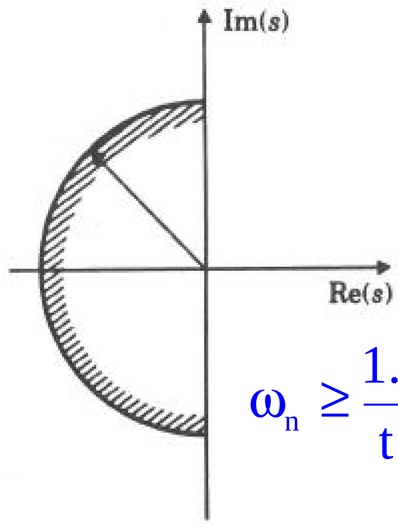
$$t_s \approx \frac{4.6}{\varsigma\omega_n} \quad \text{settling time}$$

$$M_p = e^{\frac{-\pi\varsigma}{\sqrt{1-\varsigma^2}}} \quad (0 \le \zeta < 1) \quad \text{overshoot}$$

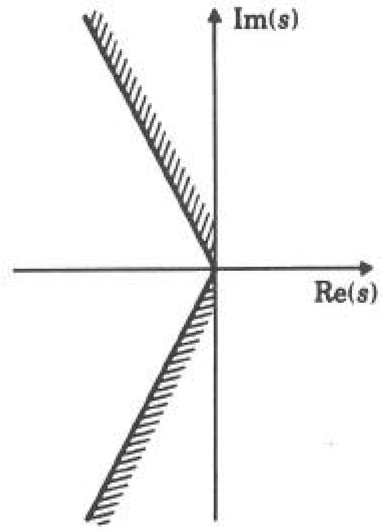
$$\approx \left(1 - \frac{\zeta}{0.6}\right) \quad (0 \le \zeta \le 0.6)$$

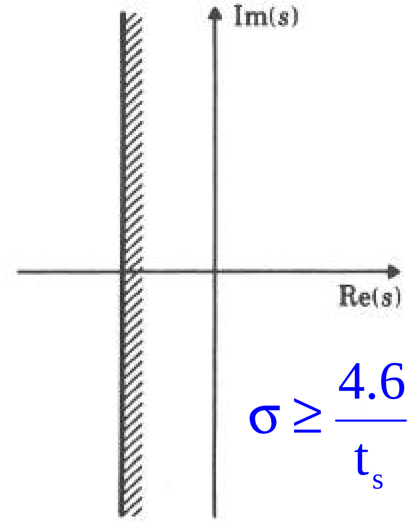


General All-Pole
2nd-Order
Step Response



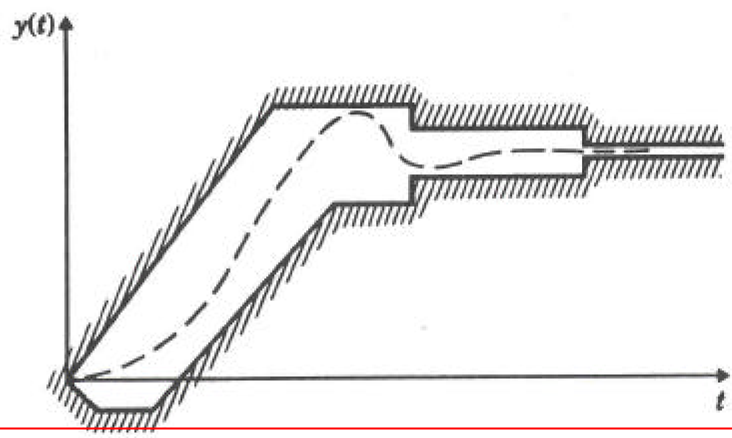
$$\omega_n \geq \frac{1.8}{t_r}$$

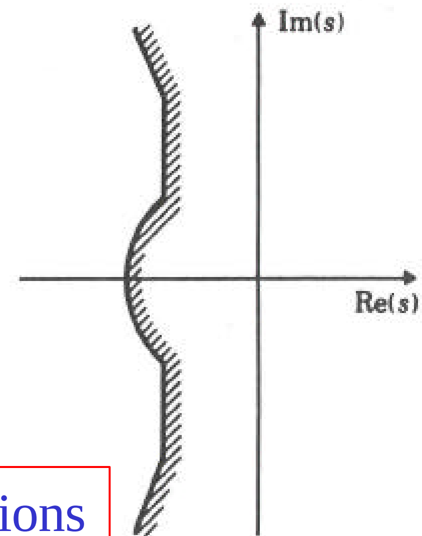




$$\sigma \geq \frac{4.6}{t_s}$$

$$\zeta \geq 0.6\left(1 - M_p\right) \qquad 0 \leq \zeta \leq 0.6$$





Time-Response Specifications vs. Pole-Location Specifications

Experimental Determination of ζ and ω_n

- ζ and ω_n can be obtained in a number of ways from step or frequency-response tests.
- For an underdamped second-order system, the values of ζ and ω_n may be found from the relations:

$$M_p = e^{-\frac{\pi\zeta}{\sqrt{1-\zeta^2}}} \Rightarrow \zeta = \sqrt{\frac{1}{\left(\frac{\pi}{\log_e\left(M_p\right)}\right)^2+1}}$$

$$T = \frac{2\pi}{\omega_d} \qquad \omega_d = \omega_n\sqrt{1-\zeta^2} \Rightarrow \omega_n = \frac{\omega_d}{\sqrt{1-\zeta^2}} = \frac{2\pi}{T\sqrt{1-\zeta^2}}$$

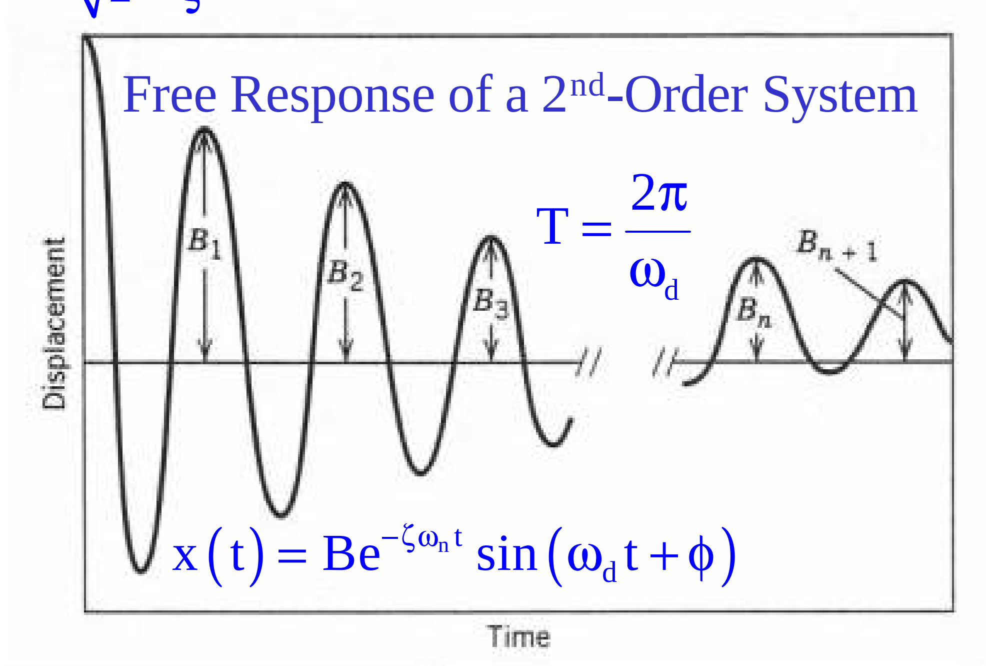
- Logarithmic Decrement δ is the natural logarithm of the ratio of two successive amplitudes.

$$\delta = \ln\left(\frac{x(t)}{x(t+T)}\right) = \ln\left(e^{\zeta\omega_n T}\right) = \zeta\omega_n T$$

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

$$\zeta = \frac{\delta}{\sqrt{(2\pi)^2 + \delta^2}}$$

$$\delta = \frac{1}{n}\ln\frac{B_1}{B_{n+1}}$$



- If several cycles of oscillation appear in the record, it is more accurate to determine the period T as the average of as many distinct cycles as are available rather than from a single cycle.
- If a system is strictly linear and second-order, the value of *n* is immaterial; the same value of ζ will be found for any number of cycles. Thus if ζ is calculated for, say, *n* = 1, 2, 4, and 6 and different numerical values of ζ are obtained, we know that the system is not following the postulated mathematical model.

- For overdamped systems ($\zeta > 1.0$), no oscillations exist, and the determination of ζ and ω_n becomes more difficult. Usually it is easier to express the system response in terms of two time constants.

– For the overdamped step response:

$$q_o = Kq_{is}\left[1-\frac{\zeta+\sqrt{\zeta^2-1}}{2\sqrt{\zeta^2-1}}e^{\left(-\zeta+\sqrt{\zeta^2-1}\right)\omega_n t} + \frac{\zeta-\sqrt{\zeta^2-1}}{2\sqrt{\zeta^2-1}}e^{\left(-\zeta-\sqrt{\zeta^2-1}\right)\omega_n t}\right] \qquad \zeta>1$$

$$\frac{q_o}{Kq_{is}} = \frac{\tau_1}{\tau_2-\tau_1}e^{-\frac{t}{\tau_1}} - \frac{\tau_2}{\tau_2-\tau_1}e^{-\frac{t}{\tau_2}} + 1$$

– where $\tau_1 \triangleq \dfrac{1}{\left(\zeta-\sqrt{\zeta^2-1}\right)\omega_n}$ $\qquad \tau_2 \triangleq \dfrac{1}{\left(\zeta+\sqrt{\zeta^2-1}\right)\omega_n}$

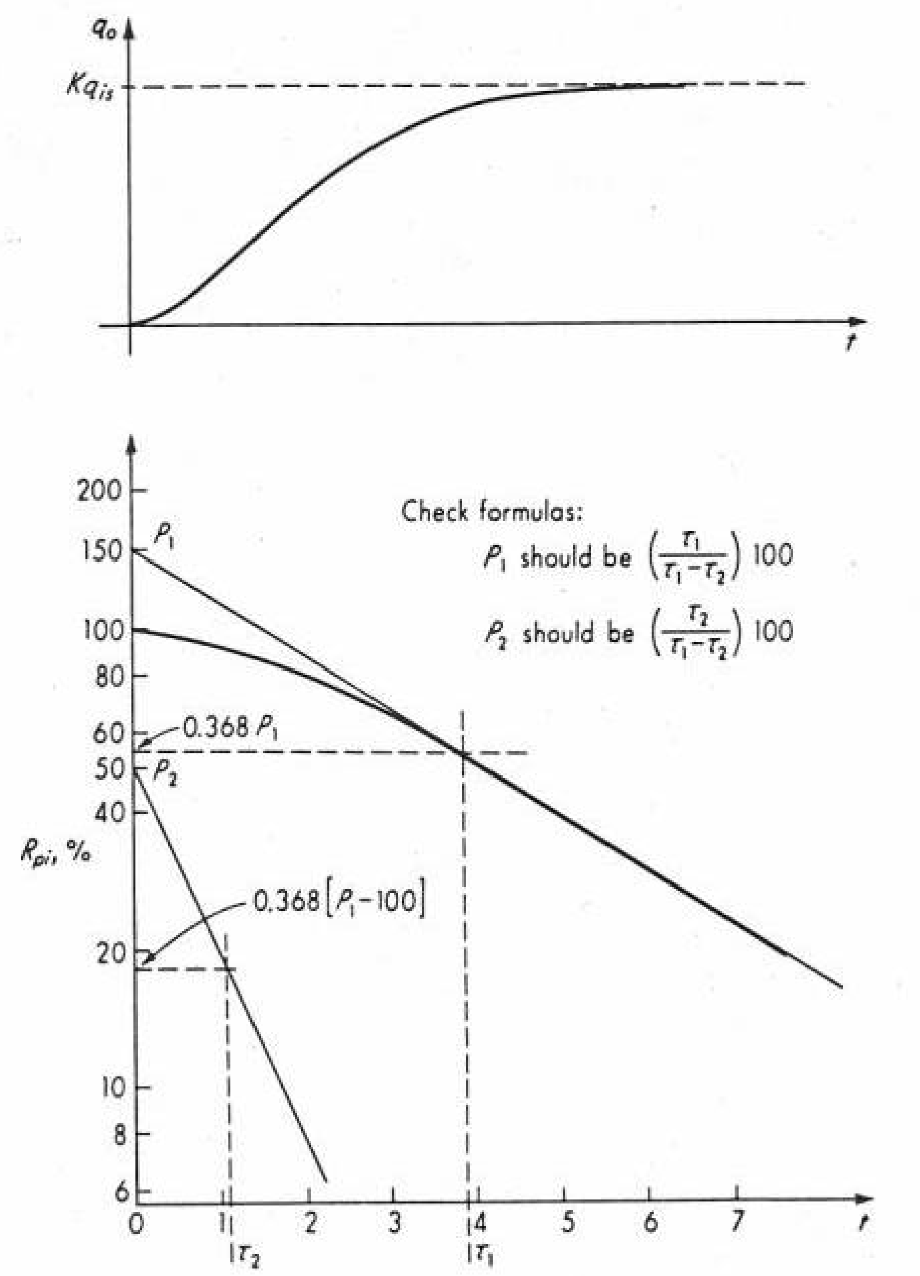
– To find τ_1 and τ_2 from a step-function response curve, we may proceed as follows:

- Define the percent incomplete response R_{pi} as:

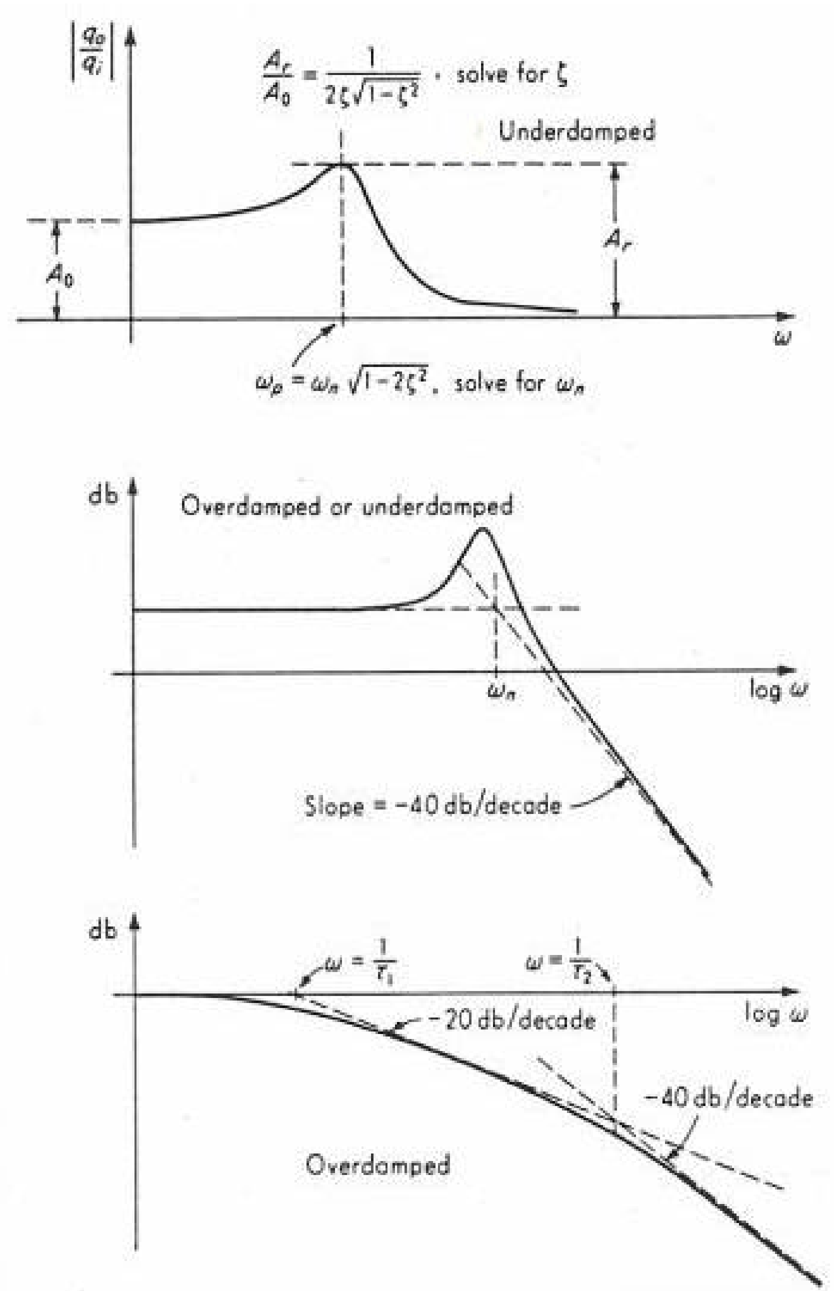
$$R_{pi} \triangleq \left(1 - \frac{q_o}{Kq_{is}}\right)100$$

- Plot R_{pi} on a logarithmic scale versus time *t* on a linear scale. This curve will approach a straight line for large *t* if the system is second-order. Extend this line back to t = 0, and note the value P_1 where this line intersects the R_{pi} scale. Now, τ_1 is the time at which the straight-line asymptote has the value $0.368P_1$.
- Now plot on the same graph a new curve which is the difference between the straight-line asymptote and R_{pi}. If this new curve is not a straight line, the system is not second-order. If it is a straight line, the time at which this line has the value $0.368(P_1-100)$ is numerically equal to τ_2.
- Frequency-response methods may also be used to find τ_1 and τ_2.

Step-Response Test for Overdamped Second-Order Systems



Frequency-Response Test of Second-Order Systems



MATLAB Tutorials

For

Mechatronics

Dr. Kevin Craig
Associate Professor of Mechanical Engineering
Department of Mechanical Engineering,
Aeronautical Engineering, and Mechanics
Rensselaer Polytechnic Institute
Troy, New York 12180
Phone: (518) 276-6626 E-mail: craigk@rpi.edu

(Written for: Matlab version 5.3, Simulink version 3.0, Control System Toolbox version 4.2)

July 21, 2000

ACKNOWLEDGEMENTS

This document would not have been possible without the assistance of several of my M.S. and Ph.D. graduate students over the past several years. They are:

Dale Lombardo
Michael Chen
Celal Tufekci
Jeongmin Lee
Shorya Awtar
Kevin Bennion
Sean Russell

I am very grateful for their assistance. They used the document in classes and research, offered valuable suggestions for improvement, and helped revise it as MatLab was updated.

Thank You!

Table of Contents

1. Before you begin

Prior to starting the tutorials contained in this booklet, take a little time to read through this introductory material.

The author would like to acknowledge that the format and some of the examples in this tutorial are based upon those contained in the MATLAB Manual from the Mathworks. The tutorials from the Mathworks are very well done, but are too long for use in Mechatronics.

1.1 What is in this tutorial?

Five short tutorials are contained in this booklet. They are :

- Basic MATLAB Tutorial
- Plotting Tutorial
- Transfer Function Tutorial
- Simulink Tutorial
- M-File Tutorial (optional)

These tutorial sections are in the order that they were intended to be used, (i.e., each tutorial builds upon the previous ones.)

1.2 Who can use this tutorial?

This tutorial was written for students and engineers in the field of Mechatronics. However, any college-level student with a level of understanding of computers and linear algebra should be able to use sections 2 and 3 of this tutorial. The last three sections of the tutorial are directed toward control system applications in MATLAB, and an understanding of the subject matter is assumed.

1.3 Why use MATLAB?

Many students will find that MATLAB is a very powerful numerical analysis tool. It can be used to evaluate complex functions, simulate dynamic systems, solve equations, and in many other applications. And now MATLAB version 5.3 can perform symbolic analysis with Symbolic toolbox (e.g., Mathematica, MathCad, Maple).

1.4 How the tutorial is structured?

This tutorial has, as much as possible, a consistent structure. Each section is intended to be an interactive tutorial. The reader should go through the tutorial while sitting at a computer terminal.

- Instructions and explanations are in this font.
- `MATLAB Commands and the expected responses are in this font and are indented.`

1.5 What is MATLAB?

MATLAB is a sophisticated mathematics and simulation environment that can be used to model and analyze dynamic systems. It handles continuous, discrete, linear, or nonlinear systems, and has extensive features for matrix manipulations. MATLAB is an open environment for which many specialized toolboxes have been developed:

- Control System toolbox
- Optimization toolbox
- System Identification toolbox
- Neural Network toolbox
- Signal Processing toolbox
- Robust Control toolbox
- Fuzzy Logic Toolbox
- and others

These toolboxes are designed to provide the user with a powerful set of analysis tools in each of their specific application areas. The success of the Control System toolbox has led to the development of Simulink. Simulink is graphical environment for modeling and simulating block diagrams and general nonlinear systems.

2. *Basic MATLAB Tutorial*

2.1 How MATLAB is structured

The workspace

The workspace is where all of the user's variables are stored.

The Command window

This window is a text window that appears once MATLAB is started. All user commands are issued from this window.

User Variables

The basic entity in MATLAB is the rectangular matrix (with real or complex entries). Each matrix must have a name, and the naming rules are similar to the rules for variable names in most computing languages. The exception to this is that unlike some languages, MATLAB is *case-sensitive*. In other words, variable names are sensitive to upper case and lower case. For example, if a matrix exists in the workspace named `stuff`, this matrix cannot be referred to as `STUFF`.

M-functions and Script files

M-functions and script files are often referred to under the larger category of m-files. M-files represent an important aspect of MATLAB that the user should be aware of. The full power and flexibility of MATLAB is based on these m-files.

M-files are simply text files with a ".m" extension. These files are written in the MATLAB programming language. This language is an extension of the same commands that one uses in the workspace with the addition of some program-flow-control commands.

MATLAB Data Files

MATLAB data files are binary files used to store workspace variables for later use.

Diary Files

Diary files save a record of a user's command window session in a text file (graphs are not saved). This can be extremely useful in tracking down mistakes when a long series of commands has been issued. Diary files also make it much easier to communicate your problems to a consultant (or MATLAB expert) when you ask for assistance.

MATLAB's Order of Operations

For mathematical expressions, MATLAB uses the standard order of operations: arithmetic, relational, and logical. However, when a command is issued, variable/function names have to run through a mini-gauntlet. First, the names are checked against the variables in the workspace, then the current disk directory is searched for an m-file of the same name, and finally, the entire MATLAB search path (type `help path` for more) is run through. MATLAB always uses the first occurrence. So be careful not to make any variables with the same name as a function you plan to use!

MATLAB Command Syntax

The general syntax of MATLAB commands is the following:

```
[output1,output2,...]=command_name(input1,input2,...)
```

where the command outputs are enclosed with square brackets and inputs within parentheses. If there is only one output, brackets are optional.

Handle Graphics

This is the MATLAB graphics system. It includes high-level commands for two-dimensional and three-dimensional data visualization, image processing, animation, and presentation graphics. It also includes low-level commands that allow you to fully customize the appearance of graphics as well as to build complete Graphical User Interfaces on your MATLAB applications.

Math Library

This is a vast collection of computational algorithms ranging from elementary functions like sum, sine, cosine, and complex arithmetic, to more sophisticated functions like matrix inverse, matrix eigenvalues, Bessel functions, and fast Fourier transforms.

MATLAB API

This is a library that allows you to write C and Fortran programs that interact with MATLAB. It includes facilities for calling routines from MATLAB (dynamic linking), calling MATLAB as a computational engine, and for reading and writing MAT-files.

Simulink

Simulink, a companion program to MATLAB, is an interactive system for simulating dynamic systems. It is a graphical mouse-driven program that allows you to model a system by drawing a block diagram on the screen and manipulating it dynamically. It can work with linear, nonlinear, continuous-time, discrete-time, multivariable, and multirate systems.

Editor/Debugger

The Editor/Debugger provides basic text editing operations as well as access to M-file debugging tools. The Editor/Debugger offers a graphical user interface. It supports automatic indenting and syntax highlighting; for details see the General Options section under View Menu. You can also use debugging commands in the Command Window.

2.2 Fundamentals

In this section, you will be introduced to several basic aspects of entering matrices and controlling the workspace.

Entering Simple Matrices

Matrices can be entered in several ways. Try the following and observe the resulting output.

Delimiters: The brackets [] indicate the beginning and end of a matrix. Spaces or commas are used to separate elements within a row, and semicolons are used to separate rows.

```
A=[1,2,3;4 5 6;7 8 9]
```

Alternatively, the matrix can be typed in as a matrix using the return key at the end of a row. The entry is not finished until the closed bracket is supplied.

```
A=[1 2 3
4 5 6
7 8 9]
```

The resulting output should be the same for the two lines above:

```
A=
  1 2 3
  4 5 6
  7 8 9
```

Enter the following lines and observe MATLAB's response...

```
A=[1,2,3;4,5;7,8,9]
```

What happened? Why?

```
A=[2 4 6;8 10 12;14 16 18];
```

Notice that on this last example there is a semicolon on the end of the line. This should have caused MATLAB to "not respond" with any output. Don't worry, MATLAB did record the matrix

A. The semicolon at the end of a line of input has the effect of preventing MATLAB from echoing the results to the screen. This is an important feature when very large matrices are being defined.

Matrix elements can be any MATLAB expression. For example :

```
x=[-1.3 sqrt(3) (1+2+3)*4/5]
```

results in

```
x=
  -1.3000 1.7321 4.8000
```

To define (or refer to) individual elements of a matrix, use the variable name and parentheses, try this:

```
x(5)=abs(x(1))
```

This produces

```
x=
  -1.3000  1.7321  4.8000  0.0000  1.3000
```

Note that the vector `x` has been augmented to 5 elements. Now try

```
x(4)=log10(x(6))
```

What happened? Why? It is possible to construct big matrices from smaller ones. Enter the following two lines:

```
r=[1 2 3 4 5];
y=[x;r]
```

results in

```
y=
  -1.3000 1.7321 4.8000 0.0000 1.3000
   1.0000 2.0000 3.0000 4.0000 5.0000
```

Now try:

```
w=[x,r]
```

This puts `x` and `r` side-by-side in the new row vector `w`.

Little matrices can be extracted from big matrices using the colon delimiter. For example :

```
z=A(2:3,1:2)
```

produces

```
z=
 8  10
14  16
```

The statement defining z can be read as follows, "set z equal to the second through third rows and first and second columns of A." Now try a second example:

```
z=y(:,2:4)
```

results in

```
z=
  1.7321 4.8000 0.0000
  2.0000 3.0000 4.0000
```

The issued command can be interpreted as "z is equal to all rows of y and the second through fourth columns." Notice that the colon can be used to specify a range of rows and columns or it can be used to specify all of the rows and columns. This notation can lead to some very exotic submatrix references and can be used on either side of the equal sign, but it is always good practice to hand check the results to see if you're getting the expected submatrices.

More on entering MATLAB expressions

The most common format for issuing commands to MATLAB is in this form

variable = expression

However, the left side of that statement can be left out, and the results of the expression are automatically placed in a variable called ans. For example type:

```
1900/81
```

results in

```
ans=
  23.4568
```

The question often arises what if the expression to be entered is longer than one line on the screen? The correct way to do this is to use ellipses at the end of the current line.

For example

```
s=1 - 1/2 + 1/3 -  1/4 + 1/5 - 1/6 + 1/7 ...
    - 1/8 + 1/9 - 1/10 + 1/11 - 1/12
```

produces the expected result.

Getting workspace information

To get a list of the current variables defined in the workspace type :

```
who
```

produces a list of the currently defined variables. Try typing

```
whos
```

What is the difference? By typing what, a current list of m-files in the search path are shown. (Just for kicks, try why.)

You can also edit your workspace variables using a graphical user interface tool called Workspace Browser. You can access Workspace Browser from `File|Show Workspace` menu. In addition to your workspace variables, there are also several permanent variables that MATLAB uses. The variable `eps` is an example of such a variable. This variable represents the working precision of the computer system. Type `eps` and look at the result. Other built in variables exist e.g., `pi`, `i`, and `j`.

Numbers in MATLAB

Conventional decimal notation is used by MATLAB; i.e., 3, -99, 0.0001, 1.54e-10, and 3.13E3 are all valid ways of entering numbers. In all calculations, regardless of what is displayed, MATLAB uses double precision floating point numbers.

The traditional mathematical operators are available (e.g., +, -, *, /, and ^). In addition to these, the \ operator signifies left division (explained later) and nearly all mathematical functions found on scientific calculators can be used (e.g., cos, sin, abs, log, and log10).

MATLAB will return `Inf` when a number becomes infinite. Unlike on many other systems, MATLAB does not halt when a division by zero occurs. Instead, it carries a representation of infinity through the remaining calculations. Try typing `s=1/0`, and see the answer MATLAB gives. Another special "number" that is often seen is `NaN`. This means "Not a Number." This is produced during calculations like `0/0` or `Inf/Inf`.

Complex Numbers

Complex numbers are allowed in all mathematical expressions. The format for entering them is shown in the following examples: (This will not work as expected if `i` & `j` have been redefined by the user to something other than `sqrt(-1)`.)

```
z=3+4*i
z=3+4*j
r=1; theta=pi/4; w=r*exp(i*theta)
```

To enter a complex matrix :

```
A=[1 2;3 4] + i*[5 6;7 8]
A=[1+5*i 2+6*i;3+7*i 4+8*i]
```

Both of these result in

```
A=
  1.0000+5.0000i   2.0000+6.0000i
  3.0000+7.0000i   4.0000+8.0000i
```

Note: That in the second expression for `A`, there are no spaces around the plus signs. MATLAB would take these as separate entries.

Also Note: MATLAB will allow you to redefine the quantities `i` and `j`. If you do so you can re-create "`i`" by entering `i=sqrt(-1)`.

Formatting Output

When displaying matrices, MATLAB will display a matrix of exact integers as such. For example, type:

```
x=[-1 0 1]
```

this yields

```
x=
  -1   0   1
```

However, if there is at least one entry that is not an integer, there are several possible display formats. Try the following

```
x=[4/3 1.234567e-8]
```

this yields

```
x=[1.3333 0.0000]
```

Notice that the second element of `x` appears to be 0.0000. Since the first element is so much larger than the second, the second element IS zero in the displayed precision! Type each of the format statements in the table on next page and type `x` with a return to see how the display types differ. You can also set the output format from the `File|Preferences` menu.

`format short` (default)
`format short e`
`format short g`
`format long`
`format long e`
`format long g`
`format hex`
`format +`
`format bank`
`format rat`

With the short and long formats, if the largest element in the matrix is larger than 1000 or smaller than 0.001, a common scale factor is used. For example :

With `format long` or `short`, type

```
y=1.e20*x
```

Look carefully at the result. A common scale factor of 1.0E+020 is applied to the matrix. Now change the format to one of the "`e`" statements, and look at `x` again.

The + format is used to look at large matrices. It displays a +, -, or blank if the element is positive, negative, or zero.

Using Diary files

Type `help diary` for a full explanation.

Basically, you issue the statement

```
diary filename.dia
```

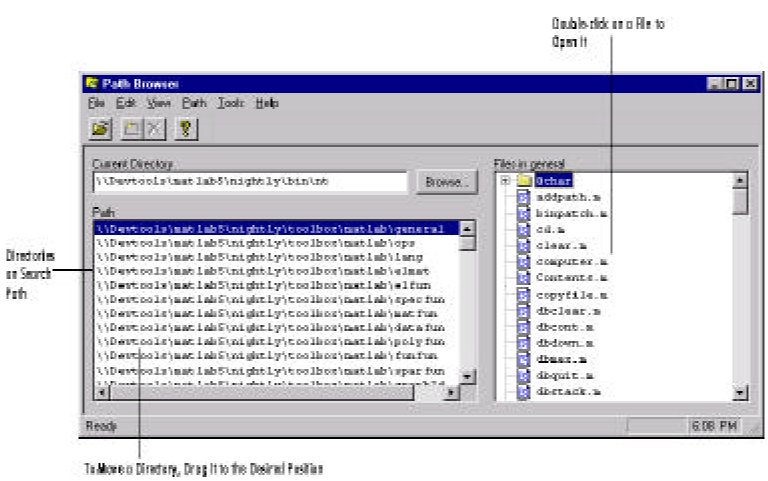
This will create a new file if it does not already exist, or start appending to the file if it does exist. MATLAB will start recording all command window text to the file until one of the following is typed.

- `diary newfile`,
- `diary,`
- `diary off,`
- or you leave MATLAB.

The diary file can be reinstated using `diary on`.

MATLAB Search Path

MATLAB uses a *search path* to find M-files. MATLAB's M-files are organized in directories or folders on your file system. Many of these directories of M-files are provided along with MATLAB, while others are available separately as toolboxes. You can edit the MATLAB Search Path using a graphical user interface tool called Path Browser. You can open the Path Browser from `File|Set Path` menu. The Path Browser lets you view and modify MATLAB's search path and see all of its files.



Getting HELP

To get help at any time simply type

```
help
```

If you have a specific topic in mind, e.g., how to use the plot command :

```
help plot
```

If you are really lost type

```
help help
```

In addition to `help` command there are other useful help commands in MATLAB.

- `helpwin` Provides Windows based help
- `helpdesk` Provides comprehensive HTML help tools (search, index options). You can also reach the online PDF manuals from within heldesk.
- `demo` Provides demo for MATLAB and related products by Math Works.
- `tour` Provides a tour of all MATLAB products like Toolboxes, Simulink, Stateflow, Blocksets, etc.
- `lookfor` Provides a search for keywords in M-files.

Clearing variables from the workspace

You may want to clear previously used unnecessary variables from the workspace. To clear variables and functions from memory, type `clear`.

To see its various options, type `help clear`.

It is a good practice to put this command in the beginning of m files before any other commands.

Quitting and saving the workspace

<u>WARNING: MATLAB does not automatically save your workspace when you quit or exit!</u>

To leave MATLAB, type `quit` or `exit`.

To save your workspace variables type

```
save
```

this saves all of your variables in a file called *MATLAB.mat*.

The full format for saving is

```
save filename variable_list
```

see `help save` for a full description.

To load an old set of workspace variables, type

```
load
```

By itself, `load` brings in the file *MATLAB.mat,* if it exists. By specifying a filename, that file will be loaded. When loading a file, the variable names used when saving are reinstated. If there is currently a variable with the same name, it will be overwritten without warning. Any (unique) variables already in the workspace are left untouched. More can be done with the `load` command, see `help load`.

2.3 Matrix Operations

Lets perform some basic matrix arithmetic.

Transpose.

The transpose operation is signified by a single apostrophe, '. If the matrix to be transposed is complex, the result is the complex conjugate transpose. This is sometimes, but not always, the desired result for complex matrices. To get around this use `conj(A')` or `A.'`. Try the following:

```
A=[1 2 3; 4 5 6; 7 8 0]
B=A'
```

results in

```
A=
  1  2  3
  4  5  6
  7  8  0
B=
  1  4  7
  2  5  8
  3  6  0
```

and

```
x=[-1 0 2]'
```

produces

```
x=
  -1
   0
   2
```

Addition and Subtraction and Multiplication.

These operations are straightforward. As is expected, the order of multiplication and subtraction matters while it does not for addition. The matrices must be of compatible size for any operation to be carried out. The only exception to this is if one of the quantities is a scalar (a 1x1 matrix), as will be demonstrated below.

Try the following:

```
C=A+B
D=A+x
b=A*x
```

The first expression was evaluated, and the second generated an error. Why? Now try these:

```
y=x-1
w=4;z=w*y
```

Note: that the scalar quantities are applied to all elements. Now try a mix of these operations, for example :

```
x'*y
y'*x
x*y'
x*y
c=A*x+y
```

Make sure you understand the results of these operations.

Matrix Division.

Matrix division is a little more complex. This is because there is no "division" operation per se. In actuality, it involves the "best" solution of a linear system. (Scalar division rules are as expected.) There is left and right division.

It is easier to explain by example :

`X=A\B` is equivalent to `X=inv(A)*B`. It solves the problem AX=B.

`X=B/A` is equivalent to `X=B*inv(A)`. It solves the problem XA=B.

If A is rectangular, A\B and B/A automatically find the least squares solution, whereas inv(A) is only valid for square matrices.

The use of matrix division is NOT recommended for simple operations because it can get confusing. Your work is easier to follow if you simply use the function `inv()`. For example, if you have the equation, b=A*x, and you wish to know x, type

```
x=inv(A)*b
```

Refer to the MATLAB manual tutorial for more information about matrix division.

Matrix Powers.

Raising a matrix to a matrix power, a matrix to a scalar power, a scalar to a matrix power, and a scalar to a scalar power are all possible. However, raising a matrix to a scalar power is most likely the only one you'll need. The only limitation is that the matrix must be square! Try these for practice :

```
A^4
x^3
```

Can you explain your results?

Miscellaneous Matrix Operations.

These operations are presented for your use, but their use is either too obvious or too rare to take up more space here. See the help facility or the Reference section of the MATLAB manual for more detail.

`det`	determinant of a matrix
`trace`	trace
`poly`	characteristic polynomial
`kron`	Kronecker tensor product
`expm`	exponential
`logm`	logarithm
`sqrtm`	square root
`funm`	arbitrary function

2.4 Array Operations

Array operations refer to element-by-element arithmetic operations, as opposed to matrix operations. By preceding the usual operators (e.g., / \ * ^) with a period . , the operation is carried out as an array operation. Many functions and all logical operators are considered array operations.

Addition and Subtraction.

SAME AS BEFORE.

Multiplication and Division.

For example:

```
x=[1 2 3]; y=[4 5 6];
z=x.*y
```

results in

```
z=
  4 10 18
```

Now try this :

```
z=x.\y
```

results in

```
z=
  4.0000  2.5000  2.0000
```

NOTE: The expression `x.\y` is equivalent to `y./x`

Array Powers.

Element by element powers are denoted by .^. Try these examples :

```
z=x.^y
```

produces

```
z=
  1 32 729
```

If the exponent is a scalar :

```
z=x.^2
```

produces

```
z=
  1   4   9
```

Or, the base can be a scalar :

```
z=2 .^[x y]
z=
  2   4   8   16   32   64
```

Relational and Logical Operations and Functions.

These operations (e.g., greater than, less than, ...) exist, but they are not often used outside of m-files. See the Reference section of the MATLAB manual, or `help <`.

Elementary Math Functions.

Many functions are inherently element by element in MATLAB. These functions are summarized below.

Trigonometric Functions		Elementary Math Functions	
`sin`	sine	`abs`	absolute value
`cos`	cosine	`angle`	phase angle
`tan`	tangent	`sqrt`	square root
`asin`	arcsine	`real`	real part
`acos`	arccosine	`imag`	imaginary part
`atan`	arctangent	`conj`	complex conjugate
`atan2`	four quadrant arctangent	`round`	round to nearest integer
`sinh`	hyperbolic sine	`fix`	round towards zero
`cosh`	hyperbolic cosine	`floor`	round towards -Inf
`tanh`	hyperbolic tangent	`ceil`	round towards +Inf
`asinh`	hyperbolic arcsine	`sign`	signum
`acosh`	hyperbolic arccosine	`rem`	remainder or modulus
`atanh`	hyperbolic arctangent	`exp`	exponential base e
		`log`	natural logarithm
		`log10`	log base 10

Generating Vectors and Matrices.

It is possible to generate vectors using operators and matrices using special functions. (Caution should be used here. If you are generating a large vector or matrix, be sure to end the statement with a semicolon. Otherwise, you'll get to see the entire vector (matrix) from beginning to end.)

To generate an evenly spaced vector try typing

```
t=1:5
```

this produces

```
t=
  1 2 3 4 5
```

Notice that this produces a row vector by default. Now try an increment other than unity:

```
y=(0:pi/4:pi)'
```

gives

```
y=
  0.0000
  0.7854
  1.5708
  2.3562
  3.1416
```

Negative increments are also possible.

Using the functions `linspace` and `logspace`, you can specify the number of points rather than the increment.

```
y=linspace(-pi,pi,4)
```

produces

```
y=
  -3.1416  -1.0472  1.0472  3.1416
```

The square identity matrix is defined by using the function `eye(n)`, where n is the number of columns or rows. The functions in the table below allow the user to generate special matrices. The arguments of these functions are either

- `(m,n)`, signifying the number of rows and columns
- `(m)`, signifying a square (m x m) matrix

or

- `(A)`, a matrix whose dimension you want to match

`rand`	random values
`zeros`	all zeros
`ones`	all ones

For example,

- ones(t), where t is a vector generates the unit step function.
- y(t)=3+t, for t=0, 1 2, ..., 10 is created by entering

  ```
  t=[0:1:10];y=3*ones(t)+t
  ```

One last useful matrix creation function is `diag`. Here is an example of its use:

```
x=[1 2 3 4];X=diag(x)
```

produces

```
x=
  1 0 0 0
  0 2 0 0
  0 0 3 0
  0 0 0 4
```

2.5 Polynomials

Polynomials are represented as vectors containing the polynomial coefficients in descending order. The `roots` command finds roots of polynomials. For example, the roots of the polynomial s^3+2s^2+3s+4 are found by

```
p=[1 2 3 4]; roots(p)
```

produces

```
ans =
  -1.6506
  -0.1747 + 1.5469i
  -0.1747 - 1.5469i
```

The `poly` command is used to form a polynomial from its roots. It returns the coefficients of the polynomial as a row vector:

```
p=poly([-1 2])
```

produces

```
p=
  1    -1    -2
```

A polynomial can be evaluated at a point using the `polyval` command. Its syntax is

```
ps=polyval(p,s)
```

where p is a polynomial and s is the point at which the polynomial is to be evaluated. The input s can be a vector or a matrix. In such cases, the evaluation is done element by element. For example, consider the polynomial p(s)=(s+1)(s+2):

```
p=[1 3 2]; s=[1 2;3 4]; polyval(p,s)
```

gives

```
ans =
      6    12
     20    30
```

Polynomials are multiplied and divided using the `conv` and `deconv` commands, respectively. The residue command performs partial fraction expansion.

2.6 Conclusion

You should now have a basic understanding of how to use MATLAB to perform simple matrix operations. At this time, it would probably be a good idea for you to try a complete problem on your own. It is recommended that you select a problem that you know the answer to, so you can verify that you have solved it correctly. Feel free to choose one of the problems below.

2.7 Sample Problems

1. If

$$A = \begin{pmatrix} 3 & 1 & 4 \\ -2 & 0 & 1 \\ 1 & 2 & 2 \end{pmatrix} \text{ and } B = \begin{pmatrix} 1 & 0 & 2 \\ -3 & 1 & 1 \\ 2 & -4 & 1 \end{pmatrix}$$

compute:

(a) 2A (b) A+B (c) 2A-3B

(d) $(2A)^T-(3B)^T$ (e) AB (f) BA

(g) A^TB^T (h) $(BA)^T$

2. Solve the following system of equations:

$$2x_1 + x_2 + x_3 = 4$$

$$x_1 - x_2 + 2x_3 = 2$$

$$3x_1 - 2x_2 - x_3 = 0$$

3. Calculate the eigenvalues and eigenvectors for A and B in #1. (Hint: `eig`)

Answers:

1. (a) $\begin{pmatrix} 6 & 2 & 8 \\ -4 & 0 & 2 \\ 2 & 4 & 4 \end{pmatrix}$ (b) $\begin{pmatrix} 4 & 1 & 6 \\ -5 & 1 & 2 \\ 3 & -2 & 3 \end{pmatrix}$ (c) $\begin{pmatrix} 3 & 2 & 2 \\ 5 & -3 & -1 \\ -4 & 16 & 1 \end{pmatrix}$ (d) $\begin{pmatrix} 3 & 5 & -4 \\ 2 & -3 & 16 \\ 2 & -1 & 1 \end{pmatrix}$

(e) $\begin{pmatrix} 8 & -15 & 11 \\ 0 & -4 & -3 \\ -1 & -6 & 6 \end{pmatrix}$ (f) $\begin{pmatrix} 5 & 5 & 8 \\ -10 & -1 & -9 \\ 15 & 4 & 6 \end{pmatrix}$ (g) and (h) $\begin{pmatrix} 5 & -10 & 15 \\ 5 & -1 & 4 \\ 8 & -9 & 6 \end{pmatrix}$

2. $x = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}^T$

3. If you used these commands

```
[Avec,Aval]=eig(A)

[Bvec,Bval]=eig(B)
```

you got the right answer. You should have used help eig to find out how to do this.

3. Plotting Tutorial

MATLAB has a very strong graphic capability that is well suited to scientific and engineering applications.

3.1 Creating Simple X-Y Graphs

Create the data.

First you must have some data to plot. Enter the following lines into the command window:

```
t=(0:pi/36:2*pi)';
y1=2*sin(6*t); y2=3*cos(2*t);
```

Plotting a single set of data.

To plot `y1`, enter the following command:

```
plot(y1)
```

Notice that if a vector is presented to the plot command, it plots the values against the index number of each value. The exception to this is if the vector contains complex numbers, then the vector is plotted as the imaginary part versus the real part. To illustrate this, type in the following command:

```
plot(t+2*t*i)
```

The result of this should be a straight line beginning at (0,0) and ending at (2*pi,4*pi).

To plot y1 vs. t, issue the command:

```
plot(t,y1)
```

Plotting multiple lines.

To plot multiple lines (e.g., y1 and y2), enter the following command:

```
plot([y1,y2])
```

Or to plot these same lines vs. t:

```
plot(t,[y1 y2])
```

However, for reasons that will become apparent shortly, the following command is often preferable:

```
plot(t,y1,t,y2)
```

3.2 Taking Control of the Plots

It is OK to be able to create these plots on the screen, but usually, you need a hardcopy of the graphs as well. In this section, you will find out how to print, add titles, gridlines, and axis labels to your plots. In addition, you'll see how to control the axis scaling, proportion, and line types and colors.

Changing line types and colors.

Let's say that you wanted to plot y1 as a green line, and y2 as a series of blue circles with a dashed red line through them. You would issue the following command:

```
plot(t,y1,'g',t,y2,'bo',t,y2,'r--')
```

Notice that the format is as below, and that you cannot plot lines and symbols in the same option string.

```
plot(xdata,ydata,'options',...)
```

The full list of options is shown in the table below:

Line Types		Point Types		Colors	
-	solid	.	point	y	yellow
--	dashed	o	circle	m	magenta
:	dotted	x	x-mark	c	cyan
-.	dashdot	+	plus	r	red
		*	star	g	green
		s	square	b	blue
		d	diamond	w	white
		v	triangle (down)	k	black
		^	triangle (down)		
		<	triangle (left)		
		>	triangle (right)		
		p	pentagram		
		h	hexagram		

Try, on your own, to plot y1 as a white dotted line and y2 as a red dash-dotted line.

Adding titles, grids, and axis labels.

To add grid lines, type

```
grid
```

To add a title, type

```
title('Your title information goes here.')
```

To add axis labels, type

```
xlabel('Time (sec)')
ylabel('V (volts)')
```

If you want to add a variable's value to a string, use something like this:

```
plot(t,y1+y2,'w--')
M=200;
title(['Response @ M=' num2str(M) ' lbs.'])
```

See the MATLAB Reference guide for more about this function and using strings or type:

```
help num2str
```

Controlling the plot axis scaling.

If you noticed, MATLAB automatically selected the plot axis scales for you. This may not always produce the desired result. The command used to change this is the `axis` command. The `axis` command, like many MATLAB commands, is a multi-purpose command. For the first sample of its use, type:

```
y3=cos(t); y4=sin(t); plot(y3,y4)
```

Notice that this should be a circle, but does not look like one. Now issue the following:

```
axis('square')
```

Now it looks more like a circle. The square option of `axis` tells MATLAB to place the graph in a square box. Try also `axis('equal').` To return to the default axis scaling with a rectangular box, type

```
axis('normal')
```

The second way to use `axis` is to manually specify the x-y ranges to show in the box. Type

```
axis([0 2*pi -5 5])
plot(t,y1,t,y2)
```

The format for the axis-vector is : `[xmin xmax ymin ymax]`

To return to the default mode with automatic ranging, type

```
Axis auto
```

Printing a graph.

The first step is to create the desired graph in the graphics window.

The print facility in MATLAB simply dumps the current graph window to the printer. The details on how to accomplish this vary from system to system.

To print the current contents of the Graphics window :

- From a workstation, type

```
print
```

- From Windows MATLAB

  use the menu bar in the graphics window

3.3 Advanced Graphing Stuff

Not everyone will need the following, but you'll be a better person for knowing that these options are available to you.

Clearing the graphics window, holding a plot, and adding text.

To clear the graphics window, type

```
clf
```

To hold a plot so that other lines can be added, type

```
hold on
```

And to release them so that the next plot command clears the old graph, type

```
hold off
```

To add text anywhere in the graphics window, use the `text` or `gtext` command. To find out how to use it, type `help text` or `help gtext`.

Creating logarithmic plots.

To create a log-log or semilog plot, first plot the data using the `plot` command, and then type one of the following:

```
loglog(…)
semilogx(…)
semilogy(…)
```

Making subplots.

This one is a little more complicated, but it is still not too bad. As always, refer to the manual or the help facility for more information. Let's create two separate plots, one above the other.

First, tell MATLAB that you want to divide the graphics window:

```
subplot(211)
```

This number's digits can be broken down, in order, as follows:

2 = two "rows" of plotting boxes

1 = one "column" of plotting boxes

1 = make the first plot box active, i.e., MATLAB will send the next plot command to the first box.

Windows are numbered from left to right and top to bottom.

Type

```
plot(t,y1), grid
ylabel('y1'), xlabel('t'), title('Graph 1')
subplot(212), plot(t,y2), grid
ylabel('y2'), xlabel('t'), title('Graph 2')
```

Watch the results carefully. The `subplot` command advances to the next graph box. Now type this

```
plot(t,y3)
```

Did you notice that the second graph was replaced when the third `plot` statement was issued? If you had wanted to make four separate boxes, you could have used

```
subplot(221)
```

3.4 Examples

(1) The equations for Mercury's orbit about the Earth are given by the following parametric equations:

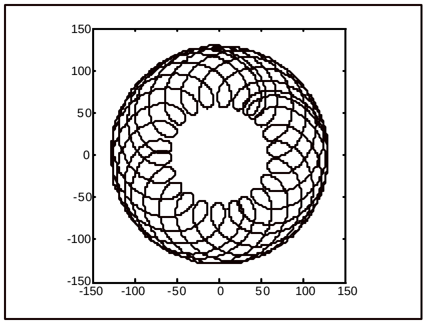
$$x(t) = 93\cos t + 36\cos 4.15t$$

$$y(t) = 93\sin t + 36\sin 4.15t$$

After 7-1/3 revolutions we get the following curve called an epitrochoid. Compute the vectors x and y and plot them against each other:

```
t=[0:pi/360:2*pi*22/3];
x=93*cos(t)+36*cos(t*4.15);
y=93*sin(t)+36*sin(t*4.15);
plot(x,y),axis('square')
```

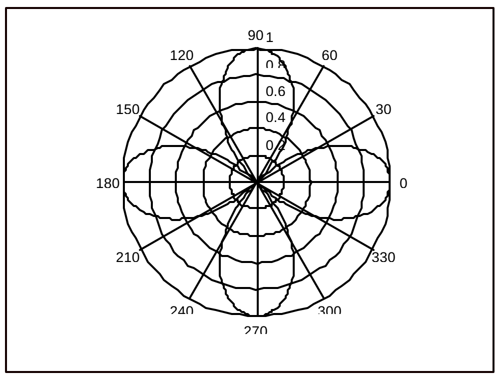
produces



(2) The equation of a four-leaf figure in polar coordinates is r=cos(2θ). The angle must be in radians for the `polar` command.

```
th=[pi/200:pi/200:2*pi]';
r=cos(2*th);
polar(th,r)
```

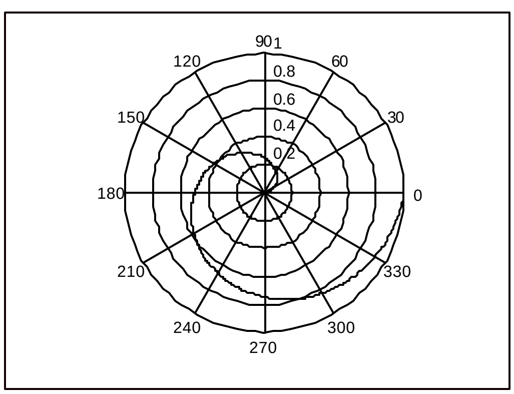
produces



(3) The equation for the Archimedes spiral is given by r=Kθ, K>0.

```
th=[pi/200:pi/200:2*pi]';
sa=th/(2*pi);
polar(th,sa)
```

produces



(4) The parametric equations of a circle with radius r and center at (a,b) are given by

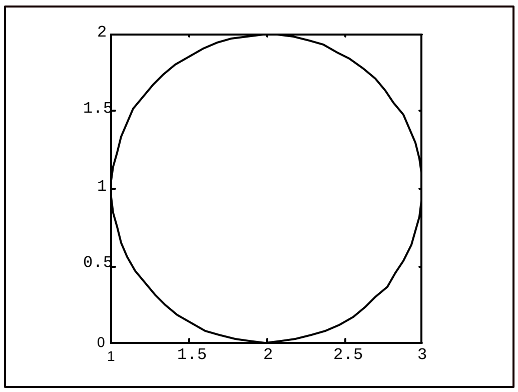
$$x(t) = r\cos(t) + a$$

$$y(t) = r\sin(t) + b$$

i.e.,
$$(x-a)^2 + (y-b)^2 = r^2$$

```
t=[0:0.1:2*pi];
a=2;b=1;
x=cos(t)+a;
y=sin(t)+b;
plot(x,y),axis('square')
```

produces



4. Transfer Function Tutorial

4.1 MATLAB and Transfer Functions

MATLAB, with the addition of the Control System Toolbox, can do quite a lot with respect to system analysis and simulation with transfer functions. This toolbox was written to handle the most general cases possible and can even accommodate MIMO (Multi-Input, Multi-Output) systems. This tutorial will only describe how to use MATLAB with SISO (Single-Input, Single-Output) systems. Keep this in mind when using the help facility for functions described here because some of the explanations include MIMO information.

MATLAB has very few limitations regarding transfer functions, the main one however is that the transfer function must be in polynomial form. There are, as will be discussed later, ways that MATLAB will help you get polynomial form from say pole-zero format; but nonlinear terms, like the exponential time-delay, are not valid.

In this tutorial, you will:

- learn how to manipulate TF's,
- learn how to generate Root Locus, Bode, Nyquist, and Nichols plots,
- learn how to simulate the responses of TF's,
- and be shown a complete analysis example.

This tutorial, however, makes no attempt to describe the development or "teach" these topics to the reader. It is assumed that the reader already has that knowledge.

4.2 LTI Models

You can specify linear time-invariant (LTI) systems as transfer function (TF) models, zero/pole/gain (ZPK) models, state-space (SS) models, or frequency response data (FRD) model. You can construct the corresponding models using the constructor functions.

```
sys = tf(num,den)                 % transfer function
sys = zpk(z,p,k)                  % zero/pole/gain
sys = ss(a,b,c,d)                 % state space
sys = frd(response, frequencies)  % frequency response data
```

To find out information about LTI models, type

```
ltimodels
```

The output `sys` is a model-specific data structure called a TF, ZPK, SS, or FRD object, respectively. These objects store the model data and enable you to manipulate the LTI model as a single entity. For example, type

```
h = tf(1,[1 1])
```

This results in :

```
Transfer function:
  1
-----
s + 1
```

4.3 Transfer Function Basics in MATLAB

Transfer function representation

In order to handle TF's, MATLAB has a specific format that must be followed. If G(s) is your transfer function in this form (a system with m zeros and n poles) :

$$G(s) = \frac{B(s)}{A(s)} = \frac{b_0 s^m + b_1 s^{m-1} + \ldots + b_{m-1} s + b_m}{a_0 s^n + a_1 s^{n-1} + \ldots + a_{n-1} s + a_n}$$

In MATLAB, the same system would be represented by two vectors containing the coefficients of the numerator and denominator in descending powers of s:

$$B = [b_0 \quad b_1 \quad \ldots \quad b_{m-1} \quad b_m]$$

$$A = [a_0 \quad a_1 \quad \ldots \quad a_{n-1} \quad a_n]$$

For example, lets say that G(s) is to be used in MATLAB where, $G(s) = \frac{s^2 + 1}{s^3 + 2s^2 + 3s + 4}$. The numerator and denominator polynomials would be as follows, type:

```
gnum=[1 0 1]
gden=[1 2 3 4]
```

Notice that the "missing" power of s is shown as a zero in the numerator. Similarly, if you want to represent H(s)=s, `hnum=[1 0]` and `hden=[1]`.

Finding poles and zeros.

Since the poles and zeros are simply the roots of the denominator and numerator, respectively, the `roots` command is used. For example, what are the poles and zeros of G(s) above? Type :

```
poles=roots(gden)
zeros=roots(gnum)
```

This results in :

```
poles=
  -1.6506
  -0.1747 + 1.5469i
  -0.1747 - 1.5469i
zeros=
   0 + 1.0000i
   0 - 1.0000i
```

(In case you were wondering, this command works with any polynomial, as long as the powers of the independent variable are in descending order.)

If you have defined a system as, for example, `sys = tf(gnum,gden)`, then the following commands give the same results as the above.

```
pole(sys)

zero(sys)
```

Multiplying TF's.

To multiply two polynomials together, you use the command `conv`. For example, if you have $H(s) = \frac{s}{s+1}$ and you want to multiply G(s) and H(s) simply issue the following commands:

```
hnum=[1 0]; hden=[1 1];
ghnum=conv(gnum,hnum)
ghden=conv(gden,hden)
```

This results in :

```
ghnum=
  1   0   1   0
ghden=
  1   3   5   7   4
```

There is a better way to "connect" two blocks in series, but this command is often useful. If, for example, you had the poles and zeros of a system, this command could be used to help create the equivalent transfer function:

```
z1=-1-i; z2=-1+i;
p1=-1; p2=-3-.5*i;  p3=-3+.5*i;
num=conv([1 -z1],[1 -z2]);
den=conv([1 -p1],conv([1 -p2],[1 -p3]));
```

results in

```
num=
  1  2  2
den=
  1.0000  7.0000  15.2500  9.2500
```

Partial fraction expansion.

To get the partial fraction expansion of a TF, you use the command `residue`. To demonstrate this, there are three examples below, each representing one of the three major cases in partial fraction expansion.

Example #1: Distinct real poles. Let $E_1(s) = \frac{s^3 + 5s^2 + 9s + 7}{s^2 + 3s + 2}$. To enter $E_1(s)$ into MATLAB :

```
e1num=[1 5 9 7]; e1den=[1 3 2];
```

To perform PFE:

```
[re1,pe1,ke1]=residue(e1num,e1den)
```

This results in:

```
re1=
   2
  -1
pe1=
  -1
  -2
ke1=
   1   2
```

The variable `re1` represents the numerators of the expanded fraction, the variable `pe1` represents the poles of the new denominators, and the variable `ke1` represents the constant terms. In other words, using `re1`, `pe1`, and `ke1`, the TF can be represented as follows:

$$E_1(s) = (s+2) + \frac{2}{s+1} - \frac{1}{s+2}$$

See `help residue` for a description of the general format.

Example #2: Distinct imaginary poles. Let $E_2(s) = \frac{s+1}{s^3 + s^2 + s}$. Enter this TF into MATLAB:

```
e2num=[1 1]; e2den=[1 1 1 0];
```

perform PFE:

```
[re2,pe2,ke2]=residue(e2num,e2den)
```

results in:

```
re2=
  -0.5000 - 0.2887i
  -0.5000 + 0.2887i
   1.0000
pe2=
  -0.5000 + 0.8660i
  -0.5000 - 0.8660i
        0
ke2=
    []
```

This corresponds to:

$$E_2(s) = \frac{-0.5-0.2887i}{s+0.5-0.866i} + \frac{-0.5+0.2887i}{s+0.5+0.866i} + \frac{1}{s}$$

Example #3: Repeated poles. Let $E_3(s) = \dfrac{s^2+4s+6}{s^3+3s^2+3s+1}$. Enter this TF :

```
e3num=[1 4 6]; e3den=[1 3 3 1];
```

Have MATLAB perform PFE:

```
[re3,pe3,ke3]=residue(e3num,e3den)
```

This results in:

```
re3=
   1.0000
   2.0000
   3.0000
pe3=
  -1.0000
  -1.0000
  -1.0000
ke3=
     []
```

In expanded form:

$$E_3(s) = \frac{1}{s+1} + \frac{2}{(s+1)^2} + \frac{3}{(s+1)^3}$$

This example was rigged so that you would be able to see how to create the expanded form from re3, pe3, and ke3. It is up to the user to recognize what to do with repeated poles in an answer. Be aware that the first "numerator" in a series of repeated poles corresponds to the first power denominator, and so on.

4.4 Connecting block diagrams of TF's

There are several commands that are designed explicitly for deriving system transfer functions from the connection of its component blocks. These commands are

`series`	to connect two TF's in series
`feedback`	to connect two TF's in feedback
`parallel`	to connect two TF's in parallel

To maintain brevity, this tutorial will only describe how to use these commands. The help facility for these functions is very good. Trying out examples is left to the reader. Also note, that these blocks are for SISO systems only when using transfer function notation.

For all of the following descriptions, assume that you have already defined G(s) and H(s), and that the connected system transfer function is F(s).

Series connections.

To connect G(s) and H(s) in series:

`[Fnum,Fden]=series(gnum,gden,hnum,hden)` or

`Fsys=series(gsys,hsys)`

where `gsys = tf(gnum,gden)` and `hsys = tf(hnum,hden)`

Unity feedback.

To connect G(s) in unity feedback:

`[Fnum,Fden]=feedback(gnum,gden,1,1,sign)` or

`Fsys=feedback(gsys,1,1,sign)`

where, `sign=+1` for positive feedback, and `sign=-1` for negative feedback.

General feedback.

To connect G(s) and H(s) in feedback, where G(s) is in the feedforward path, and H(s) is in the feedback path:

```
[Fnum,Fden]=feedback(gnum,gden,hnum, ...
 hden,sign)
```
or

`Fsys=feedback(gsys,hsys,sign)`

Parallel connection.

To connect G(s) and H(s) in parallel:

`[Fnum,Fden]=parallel(gnum,gden,hnum,hden)` or

`Fsys=parallel(gsys,hsys)`

Notice that there is no `sign` option here. It is assumed that the outputs are added together.

4.5 Transfer Function Analysis Plots

Several analysis plots are available in MATLAB. The most commonly used commands are listed below. Some of these commands have useful subcommands associated with them, these are discussed in their respective sections.

`rlocus`	Evan's root locus plot
`bode`	Bode diagram
`nyquist`	Nyquist plot
`nichols`	Nichols plot

These analysis tools will be demonstrated by example.

Evan's root locus.

This function calculates (and optionally plots) the closed-loop poles of an open-loop transfer function at different gains. The gain is assumed to be applied in the feedforward path, and it is also assumed that there is negative feedback.

Assume that $G(s)=\frac{s+1}{s^2+s+1}$ in the feedforward path, and $H(s)=\frac{1}{s+2}$ is in the feedback path.

To make a root locus plot of the closed-loop system:

```
gsys=tf([1 1],[1 1 1]);hsys=tf(1,[1 2]);
ghsys=series(gsys,hsys);
rlocus(ghsys);
```

To find out numerical values for specific closed-loop poles, you can specify a gain vector:

```
K=0:1:100; clpols=rlocus(ghsys,K);
```

Associated commands are `rlocfind` and `sgrid`. Use help to find out more about `rlocus`, `locfind`, and `sgrid`.

Bode plots.

This command is used in a similar manner to that of `rlocus`.

Assume that we want to make an open-loop bode plot of the same G(s) and H(s) as before, type:

```
bode(ghsys)
```

Or, if we want to make a closed-loop bode plot:

```
sys=feedback(gsys,hsys,-1);
bode(sys)
```

Notice that the units of the plot are db and degrees. If numerical values for phase and gain amplitude are desired, you can make a frequency vector:

```
w=logspace(-2,2,100);
[mag,phase]=bode(sys,w);
```

Unlike the plots, the units of mag and phase are magnitude ratio and degrees. An associated command that you may want to use is `margin`. This command calculates the gain and phase margins of a system. See help on `logspace`, `margin`, and `bode`.

Nichols and Nyquist plots.

The commands for these plots are `nichols` and `nyquist`. It is no mistake that these too work in a similar manner. Try the following

```
nichols(sys)
nyquist(ghsys)
```

See also `ngrid` and `margin`.

4.6 Simulating TF System Responses

Several simulation algorithms are available to help characterize the response of a transfer function to an input. There are three main commands that you can use:

`step`	calculate the system's response to a step
`impulse`	calculate the system's response to an impulse
`lsim`	calculate the system's response to a user supplied input

The help facility on these commands is also very good.

Step response.

To calculate a system's unit step response and plot it, simply type :

```
step(sys)
```

To calculate a system's step response and get numerical values out, type:

```
[y,t]=step(sys);
```

or

```
t=0:.05:10; [y,x]=step(sys,t);
```

Where `y` is the system output, `x` is the state history (don't worry about it if you don't need this), and `t` is the time vector used during the simulation. Notice that you can specify the time vector, or let MATLAB automatically select one.

Impulse response.

To calculate a system's response to an unit impulse input and plot it, type:

```
impulse(sys)
```

To get numerical values, use the same format as for `step`.

General input response.

The only two main differences between this and the above is that 1) you must create input and time vectors, and 2) MATLAB will not automatically plot the response. Let's try a unit ramp input:

```
t=0:.05:10;
u=t;
y=lsim(sys,u,t);
```

or

```
[y,ts]=lsim(sys,u,t);
plot(ts,y,t,u);
```

This command is very powerful for simulating systems. The most difficult part becomes creating the input.

4.7 Conclusion

At this point, you should be able to use transfer functions in MATLAB reasonably well. There is a lot more that can be done with MATLAB in this manner, and you are encouraged to use MATLAB's help facility to find out more.

4.8 A Complete Example

Lets assume that we have a transfer function $G(s) = \frac{s+5.0}{s^3 + 4.01s^2 + 13.04s + 0.13}$. We want to create root locus plots, OL bode plots, CL bode plots, and simulate this system's CL and OL response to a step input and an impulse.

Define the transfer function:

```
num=[1 5];
den=[1 4.01 13.04 0.13];
sys=tf(num,den);
```

First find out the poles and zeros:

```
olpoles=pole(sys)
olzeros=zero(sys)
```

Second, get the root locus plot:

```
rlocus(sys)
```

Lets pick a gain to use for the closed-loop system:

```
[k,poles]=rlocfind(sys)
```

Use the mouse to pick some point along one of the branches. Notice that this function also returns the CL poles in `poles`.

Create OL bode plot:

```
bode(sys)
```

Now let's build the closed-loop system using the gain you selected with rlocfind.

```
syscl=feedback(k*sys,1,1,-1)
```

Let's take a look at the CL bode plot:

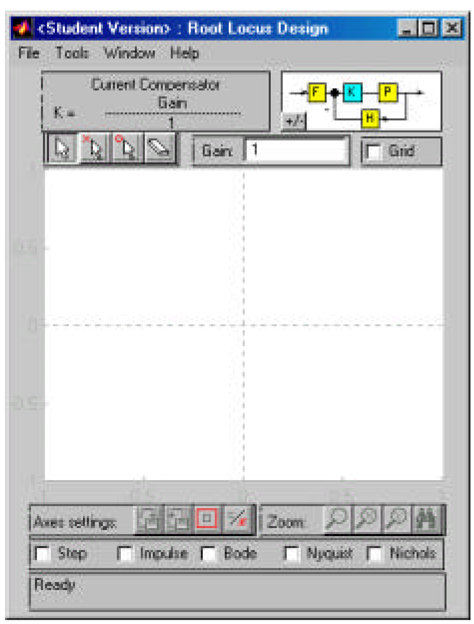
```
bode(syscl)
```

Finally lets simulate the system's CL and OL response to a step input and an impulse.

```
ysol=step(sys);
yscl=step(syscl);
yiol=impulse(sys);
yicl=impulse(syscl);
```

4.9 Rltool Example

Another useful tool for transfer function analysis and control design for single-input/single-output is the command **rltool**. Start it by typing **rltool** in the command window. The following window should appear. If you want help about a button move the mouse over the button or go to the help menu.



Root Locus Design Window

To use this tool import a model into the program. For this example enter the following transfer function into the command window.

$$G = \frac{0.0222}{s^2 + 0.1111s + 0.872}$$

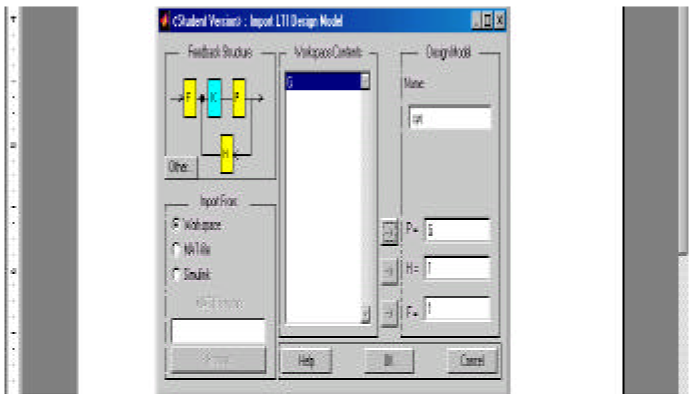
Matlab Command

```
G = tf( 0.02222, [1 0.1111 0.872])
```

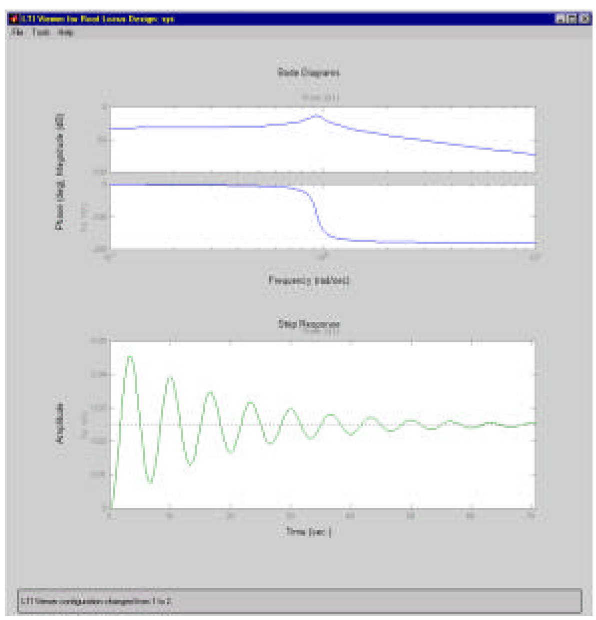
Import the model into the program,

- Select File|Import Model

The following window should appear in which you can select the model from the workspace. Import it by clicking on the arrow next to the letter P. Once you import the model you can look at the bode plot and step response by selecting the appropriate boxes.



Window to Import Model



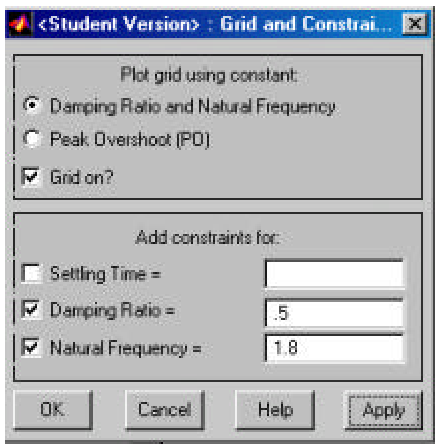
Step Response and Bode Plot

The next step is to specify the desired performance. For this example, there are two performance criteria.

- ù$_n$ greater than 1.8 for a rise time specification
- Damping ratio greater than 0.5 for an overshoot specification

Show these design specifications on the root locus

- Select Tools|Add Grid Boundary (the following figure should appear)
- Add a grid by selecting **Grid on?**
- Input the values for the natural frequency and damping ratio
- Select the appropriate boxes and click apply or OK



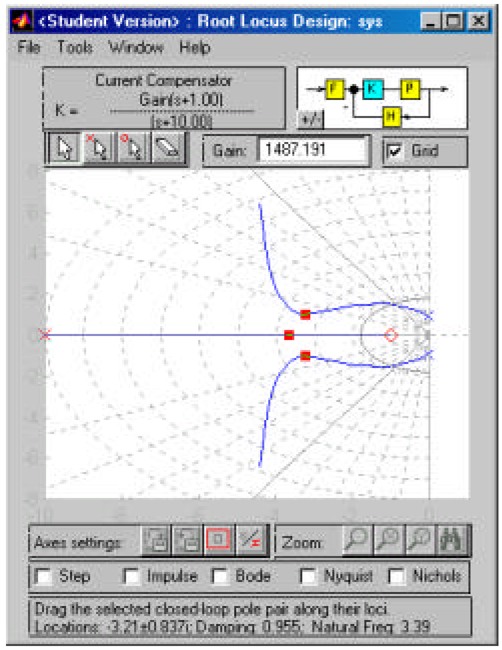
Add Design Specifications

The root locus plot should now show the acceptable region for the close loop poles. The next step is to design the compensator.

- Select Tools|Edit Compensator (The following figure should appear)
- Add a zero at –1 and a pole at –10
- Click on OK

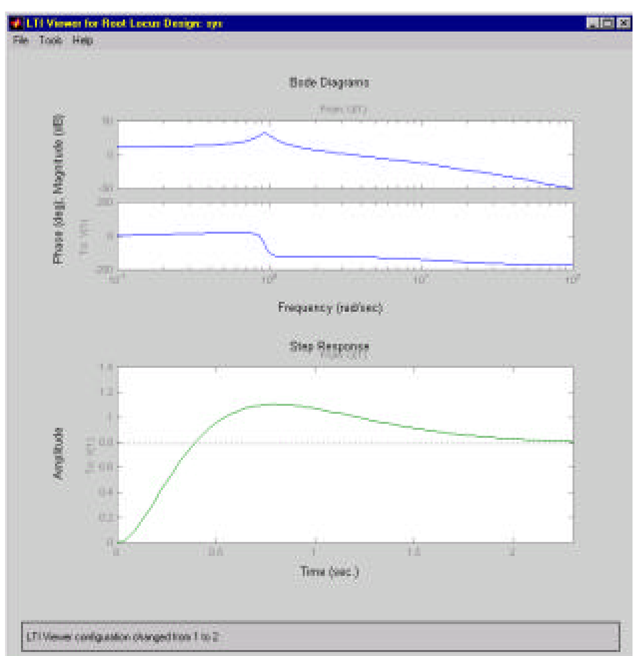
Edit Compensator Window

As you can see, the root locus changed shape, and the next step is to select an appropriate gain to move the closed loop poles. Do this by selecting the poles with the mouse and moving them to the desired location as in the following figure. You can also type the value for the gain into the corresponding box.



Root Locus Plot

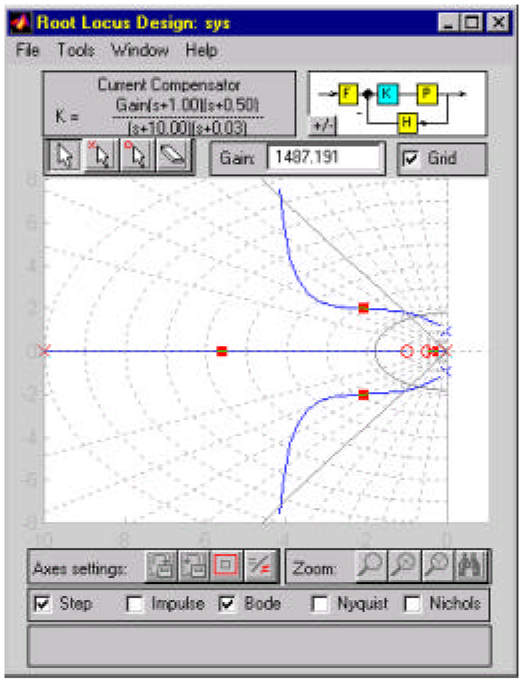
To analyze your design, various plots are available, which are seen in the bottom of the window. Plot the step response and the bode plots for the closed loop system by selecting the appropriate boxes. They should appear like the following figure. As you can see the steady state error is too large. To improve it we will add a lag compensator.



Step Response and Bode Plot

Create a lag controller by adding a pole and zero at the listed locations. The root locus should change to look like the following figure.

- zero at –0.5
- pole at –0.025



Root Locus Plot

Plot the step response and bode plot again for this system. By right clicking on the plot window a menu pops up with pull down lists. If you select **Characteristics**, more options will appear depending on the plot you selected. The following table lists the options available for the step and bode plots.

Step Response	Bode Plot
Peak Response	Peak Response
Settling Time	Stability Margins
Rise Time	
Steady State	

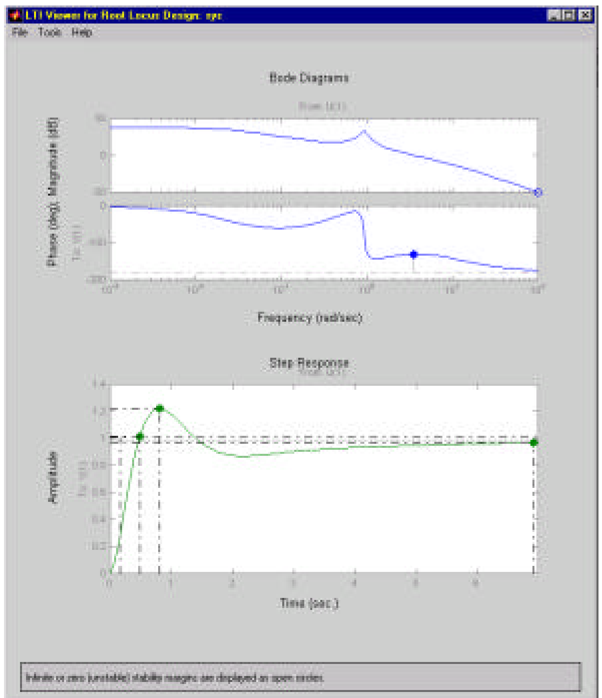
Display information about the rise time, overshoot, and steady state error on the step response.

- Select Characteristics|Rise Time
- Select Characteristics|Peak Response
- Select Characteristics|Settling Time

Display information about the phase margin on the bode plot.

- Select Characteristics|Stability Margins

A point appears representing the rise time, and if you click on the dot it will display the value for the rise time. You can do the same with the other points. The graphs should appear like the following figure.



Step Response and Bode Plot

Another useful feature is the ability to quickly see changes in the root locus as poles and zeros are moved. For example, select the pole at the far left and move it towards the origin. How does it change the root locus? Try this with the other poles and zeros.

5. Simulink Tutorial

5.1 What is Simulink?

Simulink is an extension of MATLAB that assists in the building of simulations. Simulink models are created much like a block diagram is created, i.e., by connecting up blocks and subsystems. Simulink is a VERY powerful simulation tool, and this tutorial is only a brief introduction to it. If you have a need for more information, refer to the Simulink manual. This tutorial is intended to only give an introduction, and get you started as a Simulink user. With the knowledge you gain here, you should become able to use Simulink at a reasonable level of proficiency.

5.2 Simulink's Structure

To be able to use Simulink effectively, you must understand something about how the program is structured.

Blocks

A Simulink block is a subsytem with inputs and/or outputs. Within the block, some rule exists that relates the input and output. Blocks can be almost anything. Laplace transfer functions, nonlinear relations, signal sources, and signal display devices are all examples of the kinds of blocks available.

Models

A Simulink model is a collection of individual blocks, connected in such a way that they form the model for a specific system or perform a specific task.

Libraries

A Simulink library is actually a special type of model. These libraries are collections of blocks that are used by you to create other models. You can, if you like, create your own libraries of commonly used blocks simply by copying the blocks you want into a model window and saving the new "model."

Workspace

The regular MATLAB workspace is important in Simulink. It allows you to start the Simulink interface and collect simulation data for further manipulation. In the workspace, you can also control important simulation variables.

Simulation

In order to simulate a block diagram, an integration method must be chosen. The choice of method depends on the type of system to be simulated.

5.3 Simple Simulink Tasks

In this section, you will be taught the basics of Simulink. To do this, you will create a simple model, and run a simulation.

Starting Simulink.

Simulink can only be run from within MATLAB. To start Simulink, type

```
simulink
```

from the command window. You can also start Simulink by entering the name of a model that you already created.

Using the Simulink Master Library.

Once Simulink is started, a small window is opened that contains some blocks with titles under them. This is the Simulink Master Library. All of the blocks that were supplied with Simulink are accessible from this window.

Simulink's blocks are organized into categories, subcategories, subsubcategories, etc. Currently, you are looking at the top level of this library. To look at the next level down, click on the plus sign next to the desired category, or double click the category name. All of the basic blocks are found in the *Simulink* category. For example, click on the *Simulink* plus sign, then again on the plus sign in front of *Sources*. This opens up the *source* block library. Alternatively, you can right click on the name of a category to open up the library in its own window. At some point, now or after this tutorial, you should go on a "tour" of the Simulink libraries. Look through all of the libraries to see the extent of blocks that are available. This will assist you later when you are building you own models.

To close this library, click on the minus sign next to the category name or you can select File|Close if a window is open. The same method is used to close all Simulink windows including the Master Library.

It is strongly recommended that you close the libraries once you get what you want from them. This will keep your workspace from getting cluttered, and will improve the stability of Simulink. Keeping them open gobbles up your available memory.

Now, let's create a new model for this demonstration.

- Click on the icon with a blank piece of paper

This will open a new window similar to the first, but without the blocks. Your new file is now ready for you to build your simulation.

To save a model file,

- select File|Save or File|Save As

It is recommended that you do this often. Simulink files are saved with extension, *mdl*.

Copying blocks between/within windows.

To place a block into your model, first that block must be displayed in the library. Let's copy a Signal Generator block in your file.

- Open the Sources library from the Master Library (if it is not already).
- Then, using the left mouse button, click and drag the Signal Generator block from the Sources library to your model window.
- Release it.

You have now copied this block to your file.

To copy a block within your model, you use the right mouse button.

- Click and drag the Signal Generator block using the right button somewhere else within your model.
- Release it.

This operation is a convenience when you need several copies of the same block or type of block within a model.

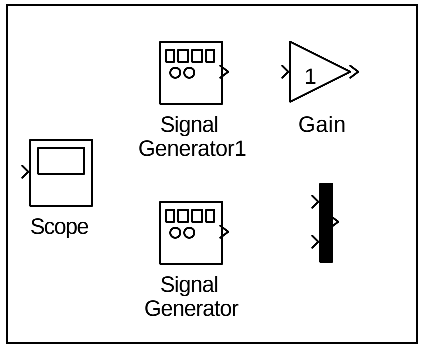
To continue the demonstration, go get copies of the following blocks:

Format: block_name (library)

- Scope (Sinks)
- Gain (Math)
- Mux (Signals and Systems)

If you have trouble finding the blocks, you can enter its name in the search box next to the binocular icon on the library browser.

Your model window would look something like this.



Deleting, moving, block parameters, connecting, branching and cleaning up model files

We do not need the second signal generator, so let's delete it.

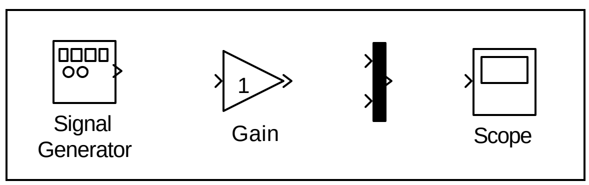
- With a single click of the left mouse button, select the second signal generator.
- Press the delete key.

It is gone. This same method can be used to delete anything within a model, including connections and text.

Lets move the individual blocks into a more logical order (preparation for connecting them). To move a block,

- simply click and drag it to its new location with the left mouse button.

Order the blocks into approximately this configuration:



Most, but not all, blocks have block parameters associated with them. Block parameters define the behavior of the blocks, and their use varies depending on the block type. For example, let's set the gain block to have a gain of 2.5.

- Double click on the gain block to open its dialogue box.
- Edit the gain value so that it reads 2.5, and click OK.

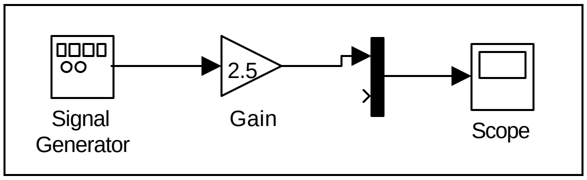
Block parameters can also be entered as names of variables in the workspace. You can also change the number of inputs on the Mux block. Go ahead and make the change. In these dialogue boxes, you can also get help on this type of block by clicking on the help button.

To connect the blocks, you use the small arrows on the sides of the block.

- Click on the "out" arrow on the Signal Gen. block and drag the connection around. Notice that the line behaves differently in free space and near an input arrow-tail.
- Release the line in a random location. Notice that it leaves the line's end in place with an arrow head on it.
- Now click on this arrowhead and drag a new connection to the input of the gain block.

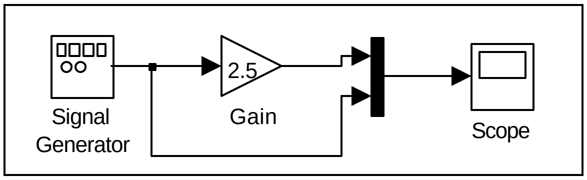
You have now connected the two blocks. (These operations also work dragging from the input side of one block to the output side of another.)

To straighten this model up, delete the connection you just made. (Select the connecting line and press delete.) Now connect up all of the blocks as shown below:



To branch (solder) a new connection line onto an existing connection, you use the right mouse button.

- Click on the **connection** from the Signal Gen. block with the right mouse button,
- drag it straight downward.
- Then using the same techniques described above, connect this branch to the second input on the Mux block as shown below.



To clean up model files, it is often useful to move several blocks around and alter connecting lines. To group some blocks together for moving, copying, or deleting,

- simply click with the left mouse button and drag a rectangle around the blocks that you want to alter.
- Now the grouped features are highlighted (small handles appear in the corners of the blocks),
- and as far as editing goes, they will behave as one block. (This is different from Edit|Create Subsystem.)

To modify a connection line,

- first you select it.
- Then you can grab any of the line's vertices and move them around.

You can also change the names of the blocks to more descriptive names.

- Click on the text below the gain block.
- Now that it is highlighted, it can be edited. If you press return, it adds a new line to the name.
- To finish editing the block name, click elsewhere in the model window.

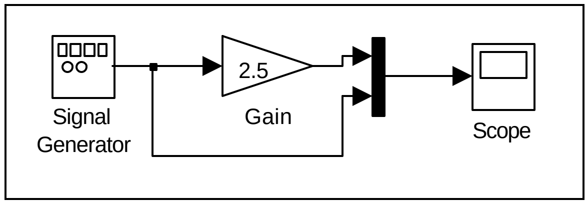
Every block must have a name, and it must be unique.

Finally, you can change the size of blocks.

- Select the gain block.
- Drag one of the handles that appear at the corners outward from the center of the block to make it larger.

If a block is made large enough, more information about the contents of that block can often be seen.

Play with these features until you get the following.



The reader is encouraged to explore Format|Flip Block and Format|Rotate Block on his/her own.

5.4 Running Simulations

The basics of running a simulation are very simple, so in addition to explaining this, you will be introduced to several other useful aspects of Simulink that are related to simulations.

Controlling a simulation

To start a simulation, select Simulation|Start from the menu bar. If you do this with the current model, nothing appears to be happening. This is because there are no 'open' display devices on the screen.

To stop the simulation, select Simulation|Stop from the menu bar.

There are several other parameters that the user can control in a simulation. These options are summarized below:

Start Time	When to start simulation (sec)
Stop Time	When to stop simulation (sec)
Solver Options	This controls the algorithm that Simulink uses to simulate your model. (see descriptions below) Variable-step solvers can modify their step sizes during the simulation. Fixed-step solvers take the same step size during the simulation.
Max Step Size	The largest time step the solver can take
Initial Step Size	A suggested first step size
Relative Tolerance	A percentage of the state's value
Absolute Tolerance	The acceptable error as the value of the measured state approaches zero

The individual solvers are described briefly below, including what they're good at solving and what they are not.

- ode45 --- The best solver to apply as a "first try" for most problems.
- ode23 --- More efficient than ode45 at crude tolerances and in the presence of mild stiffness.
- ode113 --- More efficient than ode45 at stringent tolerances
- ode15s --- Use this solver if you suspect that a problem is stiff or if ode45 failed or was very inefficient.
- ode23s --- This method can solve some kinds of stiff problems for which ode15s is not effective.
- ode23t --- Use this solver if the problem is only moderately stiff and you need a solution without numerical damping.
- ode23tb --- More efficient than ode15s at crude tolerances
- discrete (variable-step) --- The solver Simulink chooses when it detects that your model has no continuous states
- ode5 --- The fixed-step version of ode45

- ode4 --- The fourth-order Runge-Kutta formula
- ode3 --- The fixed-step version of ode23
- ode2 --- Heun's method, also known as the improved Euler formula
- ode1 --- Euler's method. Suffer from accuracy and stability problems. Not recommended for use as anything except to verify results.
- discrete (fixed-step) --- Suitable for models having no states and for which zero crossing detection and error control are not important

Watching a simulation

The scope block allows you to view a signal as it runs. To do this,

- Double click on the scope icon. This opens up a small window with a grid.
- Before we start the simulation, change the horizontal range to 10 by clicking the `Properties` icon and changing `Time range`, and the vertical range to 3 by right clicking on an axes and choosing `Properties…`. Don't close the scope unless you no longer wish to view it. Leave this one open for now.
- Change the Simulation|Parameters so that 0.01 is the maximum step size.
- Now select Simulation|Start.

You should see a sine wave on the scope's grid. Let the simulation continue while you move on to the next section.

Changing parameters "on-the-fly"

One of the nice features of Simulink is that you can alter some system parameters *while the simulation is running*. Try this.

- Double click the Signal Gen. icon.
- Select the sawtooth wave, and see how the scope graph changes.
- Try changing the amplitude of the input.

Some blocks even allow you to change their parameters while running a simulation.

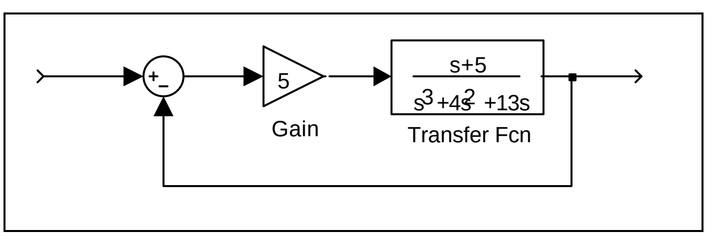
- Open the gain block's dialogue box, and change the gain.

As soon as you select OK, the value has been changed. If you try to change a fixed parameter during a simulation (e.g., the order of a transfer function block), Simulink will tell you that you cannot do that or it will stop the simulation.

Play around a little, but stop the simulation before continuing on to the next section.

5.5 Using the Workspace

This section will demonstrate how to pass information back and forth between Simulink and the MATLAB workspace. To demonstrate this, we will use an entirely new example. Create a new file as shown below.



Be sure to redefine the block parameters as follows:

- Sum (Math) ... +-
- Gain (Math) ... 5
- Transfer Fcn (Continuous) ... numerator [1 5] and denominator [1 4 13 0]

With this done, you can proceed.

Sending/Receiving data between the workspace and Simulink

There are two blocks that allow you to do this:

1. to workspace (Sinks library)
2. from workspace (Sources library)

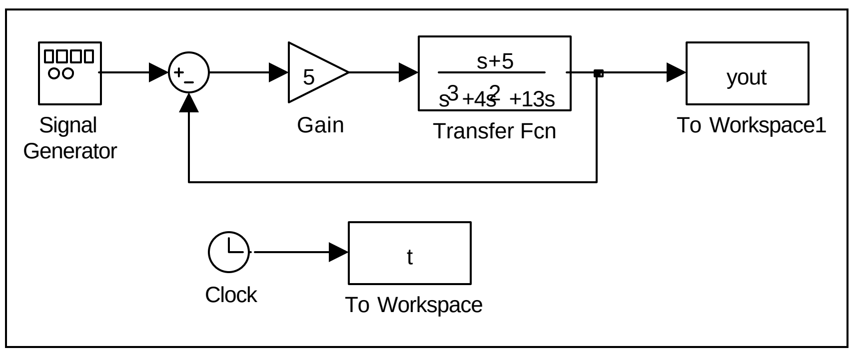
Lets begin by using the "to workspace" block (it's a little simpler).

- Modify your model so that it appears as shown below. The clock block is found in the Sources library.
- The default variable name for To Workspace is `simout`. This is easily changed by double clicking on the block.

Be sure to change the data type from *Structure* to *Matrix*. While a *Structure* type holds more information, *Matrix* is simpler to manipulate in the workspace.

The simulation will automatically send to the workspace the time steps in the variable *tout*, but it is always a good idea to add the clock setup to any system where you use To Workspace. This is because most of the simulation algorithms use adaptive step sizing, and this is the simplest way to record the time vector associated with a simulation.

For this simulation, a maximum number of rows equal to 1000 is adequate.



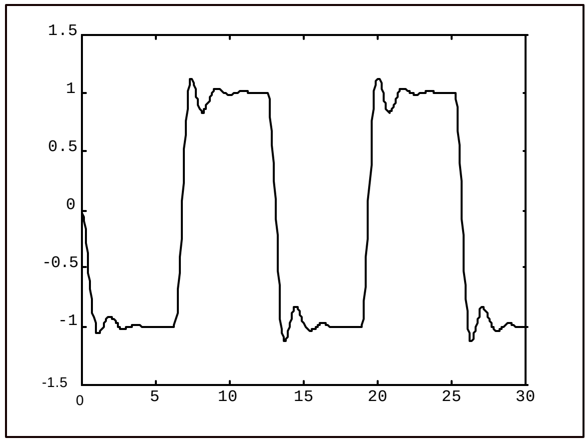
Before you start the simulation:

- Change the Signal Generator to create square waves with a frequency of 0.5 rad/sec.
- Change the Simulation|Parameters|Solver options to ode45, the stop time to 30 sec, and the max step size to 0.1 sec.
- Start the simulation.
- You'll hear a beep when the simulation is done. Now that the system output is saved in a workspace variable, it can be manipulated and plotted like any other simulation variable.

To see the results of the simulation, open the MATLAB command window and type

```
plot(t,yout)
```

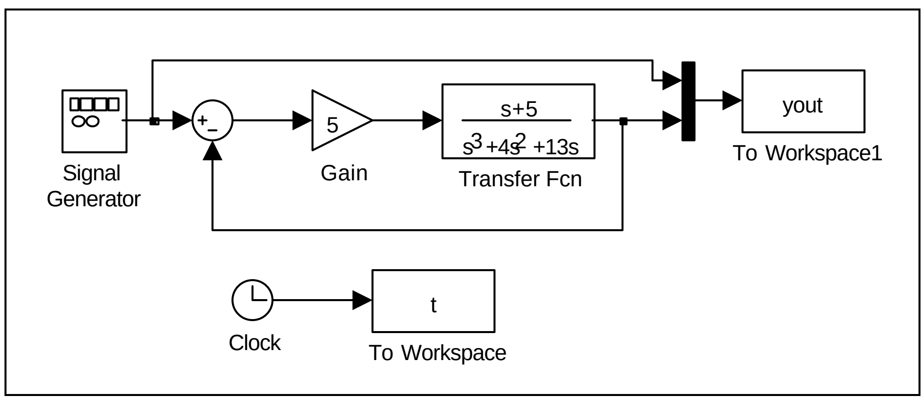
The result looks like this:



A To Workspace block can be connected to each branch that you desire the data for (with different variable names of course). However there is a more efficient method. By using the Mux block from the Connections library, a multiple-column matrix can be stored in one variable.

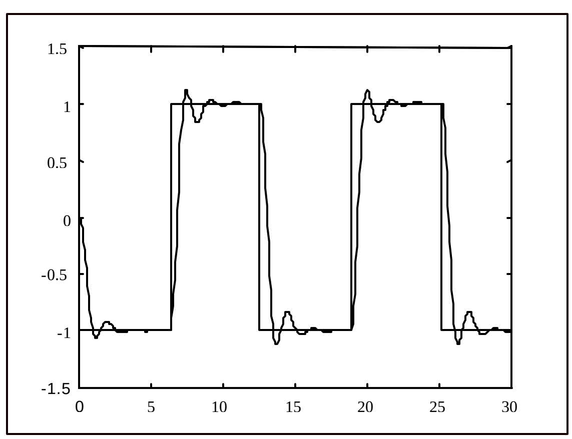
For example,

- modify your model to be as shown below:



- Now run the simulation and issue the same `plot` statement in the command window.

There is now a second line on the graph. The first and second columns of `yout` correspond to the first and second inputs to the Mux block, the signal generator output and the system output.



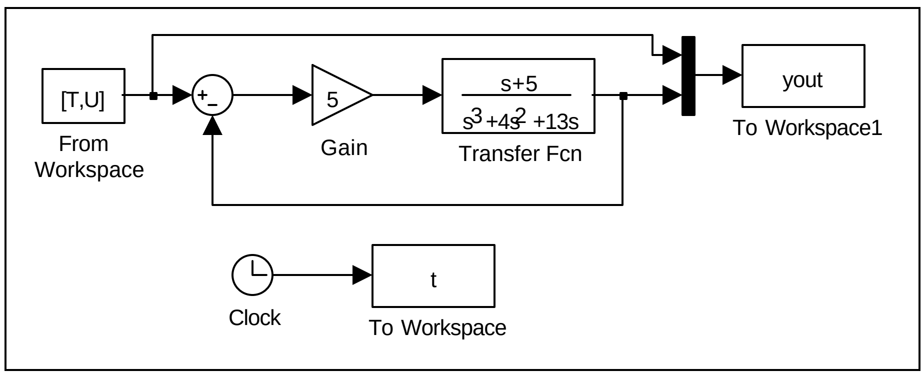
The block From Workspace works in a similar but different manner. Substitute a From Workspace block for the Signal Generator and label it as shown in the figure below. With the From Workspace block, the workspace variables, [T,U], must be defined in the workspace prior to starting the simulation.

To define [T,U], use the following statements: (don't forget the transpose operators and the ;)

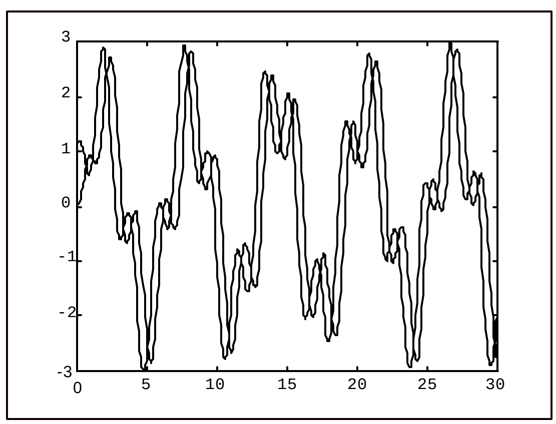
```
T=[0:.05:30]';
U=(2*sin(T)+cos(3.3*T));
```

Simulink will interpolate between the data points as it simulates.

- Start the simulation.
- From MATLAB, plot the results using the same command as before. Simulink used *your* input to the system.



The result should be

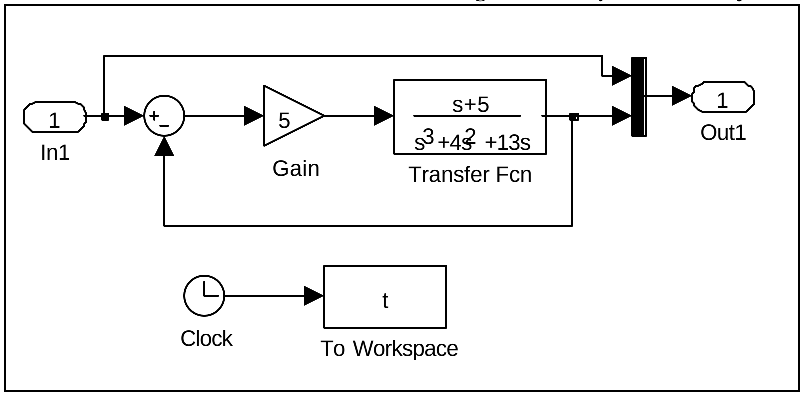


That is essentially all that you'll need to pass data back and forth between the workspace and Simulink. There are also blocks within Simulink to load values from files. It is left up to the reader to figure these out.

Getting a workspace linear model for a system.

MATLAB has a command `linmod` that allows the user to create a state-space model of a Simulink model file. This tutorial will only explain how to use this command with linear systems. (`linmod` can also be used with nonlinear systems to produce a linearized system.)

- Modify the model that you just created to the following and save it as *simtutor.mdl*.
- The In1 and Out1 blocks are contained in the *Signals and Systems* library.



For MIMO systems, the numbers that you assign to the Inports and Outports become the order of the inputs/outputs for the `linmod` system. One limitation to keep in mind here is that Inports and Outports can only handle scalar quantities.

Now lets create the state space model for this system:

```
[A,B,C,D]=linmod('simtutor')
```

For those who wish to use transfer function notation, you can then issue the command

```
[num,den]=ss2tf(A,B,C,D)
```

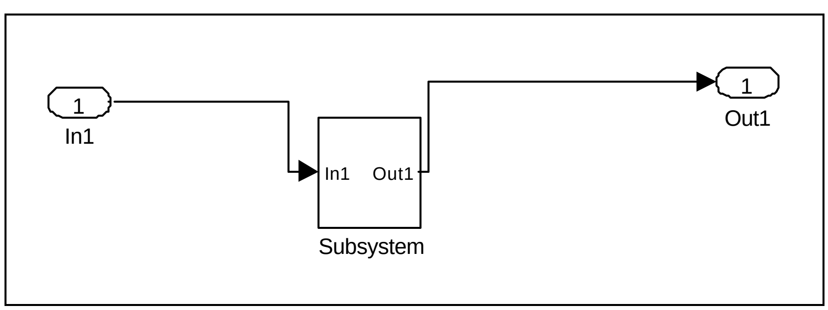
See `help linmod` and `help ss2tf` for more information.

5.6 Creating Subsytems

For simulations of large systems, it is recommended that the system be broken down into smaller parts. This is done using Edit|Create subsystem from the menu bar.

- Click the left mouse button in a clear area of the window.
- Drag a rectangle from this point so that you have surrounded the entire model except the in and out ports.
- Release the button. All of the blocks except the ports will be highlighted.
- Select Edit|Create subsystem, and now the model should look like this:

You can now treat this block like any other block that you might use. You can give the block a name, copy it to other models, copy it within the same model, etc.



To see the new block's component parts and edit them,

- double click on the "grouped" block. This will open another window.

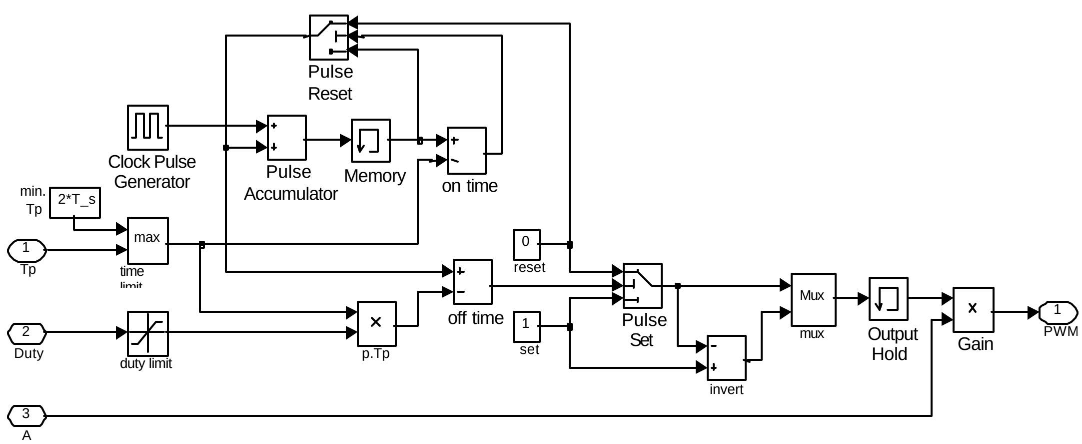
5.7 Masking Subsystem

By masking a subsystem you can build your own Simulink block that is composed of the built-in Simulink blocks and behaves the same way as the built-in blocks do. Masking enables you to customize the dialog box and icon for a subsystem. With masking, you can:

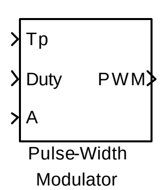
- Simplify the use of your model by replacing many dialog boxes in a subsystem with a single one.
- Provide a more descriptive and helpful user interface by defining a dialog box with your own block description, parameter field labels, and help text.
- Define commands that compute variables whose values depend on block parameters.
- Create a block icon that depicts the subsystem's purpose.
- Prevent unintended modification of subsystems by hiding their contents behind a customized interface.
- Create dynamic dialogs.

To create the mask for a subsystem, select the subsystem block and choose `Mask Subsystem` from the Edit menu. The *Mask Editor* pops up with which you can define dialog prompts and their characteristics, the masked block description and help text, and the commands that creates the masked block icon. For detailed instructions on the *Mask Editor*, refer to `Help` in the *Mask Editor*.

The following is an example of a masked subsystem.



A complicated diagram made by the built-in Simulink blocks is masked into a block icon that has a dialog box of its own.



5.8 Linear Analysis Tool (LTI Viewer)

LTI Viewer is an interactive environment for comparing time and frequency responses of LTI systems. The Viewer can contain up to 6 response areas, where each response area can show a different response type and be independently manipulated. The LTI Viewer controls are found in two main locations:

1. The Figure menus (*File*, *Tools*, and *Help*)
2. Right click menus (from any axes displaying a response plot)

The Figure menus provide high level tools for manipulating data in the LTI Viewer, and configuring the appearance of the Viewer.

File allows you to Import/Export/Delete LTI Objects from the Viewer's workspace, or Open/Close/Print the Viewer and related windows.

Tools opens additional windows for configuring the number of response areas to show on the Viewer, as well as setting up response and line style preferences.

Help provides tips on using the Viewer and related windows.

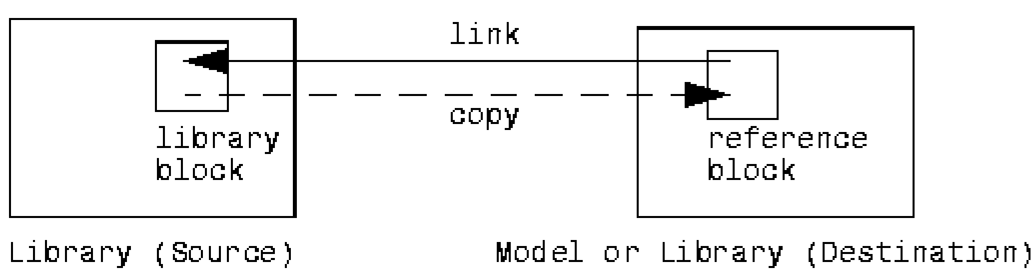
`Simulink` provides access to Simulink model in use to get the linear model.
The right click menus provide tools for manipulating the actual responses.
The LTI Viewer can be called from Tools|Linear Analysis menu in the Simulink model file. When it is called the LTI Viewer brings *Input Point* and *Output Point* block window for use in linear analysis.

5.9 Creating Libraries

To create a library, select `Library` from the `New` submenu of the `File` menu. Simulink displays a new window, labeled `Library: untitled`. If an untitled window already appears, a sequence number is appended.
You can create a library from the command line using this command.

```
new_system('newlib', 'Library')
```

This command creates a new library named `'newlib'`. To display the library, use the `open_system` command. The library must be named (saved) before you can copy blocks from it.



It is important to understand the terminology used with this feature.

Library - A collection of library blocks. A library must be explicitly created using `New Library` from the `File` menu.
Library block - A block in a library.
Reference block - A copy of a library block.
Link - The connection between the reference block and its library block that allows Simulink to update the reference block when the library block changes.
Copy - The operation that creates a reference block from either a library block or another reference block.

5.10 Conclusion

You should be capable of building and simulating systems using Simulink at this point. Remember that, if you have not done so, you should take some time to explore the libraries and see what is there. Practice and experience will, of course, improve your knowledge of Simulink. It is recommended that you try to model a simple system from beginning to end for practice.

There are extensive examples and demos available under the Extra library. Take a look at these to see how more complicated systems can be built.

6. M-File Tutorial

MATLAB supports some basic programming structures that allow looping and conditioning commands along with relational and logical operations. These new commands combined with ones already discussed can create powerful programs or new functions in MATLAB.

6.1 Introduction to m-files

An m-file is simply a text file that is created to perform a specific task. M-files can be broken down into two categories:

- Script files
- User functions

Script files are analogous to batch files. They are executed as if the user had entered the commands directly, while user functions behave like the "built in" MATLAB functions. As a matter of fact, many of the MATLAB functions are simply m-files that were written by the Mathworks.

The differences between these types of m-files are fairly simple to understand. Because script files are executed as if they were typed at the command prompt, variable references within the script file affect the workspace. Functions, on the other hand, have an input parameter list and output parameter list. Any variables used within the function are local and do not affect the workspace.

In this tutorial, you will be shown examples of both types of files and given a brief set of commands that are special to (or most commonly used in) m-files. To create your own m-files, all that you will need is an editor that creates ASCII text files. The only restriction on this editor is that it cannot insert anything other than ASCII characters into the file.

MATLAB now has its own text editor and debugger for creating and editing m-files. The editor is simple to use and has features specifically useful to MATLAB applications. You can create a new m-file by going to File|New|M-File

6.2 Script files

example1.m : A simple flow control demonstration.

This program will demonstrate the use of for-loops and if-then-elseif-else statements.

```
% This is a comment line.  The %-symbol tells MATLAB to ignore
% the rest of the line.
%
% In addition, the comments at the beginning of ANY m-file are
% considered to be the help screen for that function.  The first
% line that is not a comment terminates the help section.

% this line will not be printed if
%     help example1
% is issued from the command line.

% define some constants
t=linspace(0,1,100);
% a, b, and c are assumed to be defined in the workspace

% calculations in a for loop
for n=1:100       % for loop starts
      f(n)=t(n)^5 + a*t(n)^3 + b*t(n) + c;
end         % for loop ends

% if statement demonstration
if (any(f==0))    % are there any exact roots
      disp('An exact root for f(t) was found in [0,1].')
elseif (any(f<0) & any(f>0))  % are there any roots
      disp('f(t) has a root in [0,1]')
else  % if none of the above conditions...
      disp('f(t) has no root in [0,1]')
end

% plot results
plot(t,f), grid
```

In the workspace, enter the coefficients :

```
a=-40; b=-20; c=5;
```

Run the program by typing

```
example1
```

Change the coefficients as follows:

```
c=100;
```

and run it again.

Finally, try the program with

```
c=0;
```

This script file has created several variables while it was running. Type

```
whos
```

If there had been other variables with the same name in the workspace, they would have been overwritten by the script file. Keep this in mind when you use script files.

The `disp` command displays strings and variables without displaying the variable name.

The for-loop is basically of this format:

```
for variable=start:increment:stop
    your commands
end
```

It is possible to nest for loops, but each one must have its own `end` statement. It is also possible to use a matrix as the loop variable. See `help for` for information on this. See `help while` for information on while-loops.

The conditional used in this program is the most general one possible. The statement must be of the following format :

```
if (conditional)
    your commands
elseif (conditional) (optional)
    your commands
else (optional)
    your commands
end
```

The `elseif` and `else` parts are optional, and you can have nested conditionals.

example2.m : A demonstration of user input/output.

```
% This file demonstrates the use of user input in
% a script file.  In addition, more complicated
% output is also demonstrated.


% get the user's desired value for search range and coeff's
tlow=input('Enter the lower bound for t : ');
tup=input('Enter the upper bound for t : ');
```

```
a=input('Enter the first coeff : ');
b=input('Enter the second      : ');
c=input('Enter the third       : ');

% define some constants
t=linspace(tlow,tup,100);

% calculations in a for loop
for n=1:100       % for loop starts
      f(n)=t(n)^5 + a*t(n)^3 + b*t(n) + c;
end         % for loop ends

% define string to contain bracket e.g. '[0,5]'
T=['[' num2str(tlow) ',' num2str(tup) ']'];



% if statement demonstration
if (any(f==0))    % are there any exact roots
      disp(['An exact root for f(t) was found in ' T])
elseif (any(f<0) & any(f>0))  % are there any roots
      disp(['f(t) has a root in ' T])
else  % if none of the above conditions...
      disp(['f(t) has no root in ' T])
end

% plot results
plot(t,f), grid
```

This script file demonstrates the use of the `input` command, and shows you how to incorporate variable strings into a `disp` command. Now run the program by typing

```
example2
```

and play around with the ranges and coefficients.

6.3 Functions

example3.m : A function with inputs only.

```
function example3(tlow,tup,a,b,c)
% This file is a simple function with no outputs

% define some constants
t=linspace(tlow,tup,100);

% calculations in a for loop
for n=1:100       % for loop starts
      f(n)=t(n)^5 + a*t(n)^3 + b*t(n) + c;
end         % for loop ends
```

```
% define string to contain bracket e.g. '[0,5]'
T=['[' num2str(tlow) ',' num2str(tup) ']'];

% if statement demonstration
if (any(f==0))    % are there any exact roots
      disp(['An exact root for f(t) was found in ' T])
elseif (any(f<0) & any(f>0))  % are there any roots
      disp(['f(t) has a root in ' T])
else  % if none of the above conditions...
      disp(['f(t) has no root in ' T])
end

% plot results
plot(t,f), grid

return
```

This function is the same as the previous script file except that the user variables are now in a parameter list. The only way that MATLAB knows that a function is a function and not a script file is by the `function` statement at the first line in the file. This statement defines the format for the function.

To run this function type

```
      example3(0,1,-40,-20,5)
```

Notice that when you run this function, the workspace variables are unaffected.

example4.m : A function with a single output.

```
function status=example4(tlow,tup,a,b,c)
% This file is a simple function with one output.

% define some constants
t=linspace(tlow,tup,100);

% calculations in a for loop
for n=1:100       % for loop starts
      f(n)=t(n)^5 + a*t(n)^3 + b*t(n) + c;
end         % for loop ends

% if statement demonstration
if (any(f==0))    % are there any exact roots
      status=0;
elseif (any(f<0) & any(f>0))  % are there any roots
      status=1;
else  % if none of the above conditions...
      status=2;
end

return
```

Notice how the `function` statement has changed. Now the function will return a number between 0 and 2 depending on whether roots were found.

Type :

```
example4(0,1,-40,-20,5)
```

You should see the following :

```
ans=
    1
```

Now type :

```
result=example4(0,1,-40,-20,0)
```

this results in

```
result=
    0
```

example5.m : A multi-output function.

```
function [root,status]=example5(tlow,tup,a,b,c)
% This file is a simple function with two outputs.

% define some constants
t=linspace(tlow,tup,100);

% calculations in a for loop
for n=1:100       % for loop starts
      f(n)=t(n)^5 + a*t(n)^3 + b*t(n) + c;
end         % for loop ends

% if statement demonstration
if (any(f==0))    % are there any exact roots
      status=0;
      root=t(find(f==0));     % return the exact roots
elseif (any(f<0) & any(f>0))  % are there any roots
      status=1;
      for n=1:99        % find brackets of roots
            if ((f(n)*f(n+1))<0)
                  root=[t(n) t(n+1)];
            end
      end
else  % if none of the above conditions...
      status=2;
      root=[];
end

return
```

Notice how the function statement has changed. In this function, the roots (or bracketed ranges of the roots) and the status will be returned.

Try the following examples :

```
[r1,s1]=example5(0,1,-40,-20,0)
[r2,s2]=example5(0,1,-40,-20,5)
[r3,s3]=example5(0,1,-40,-20,100)
```

```
example5(0,1,-40,-20,5)
r4=example5(0,1,-40,-20,10)
```

Notice that if both output variables are specified, they are returned. Otherwise, only the first output variable is returned.

6.4 More M-file commands

<table>
<tr><th colspan="4">Relational and Logical Operators</th></tr>
<tr><td><</td><td>less than</td><td>&</td><td>AND</td></tr>
<tr><td><=</td><td>less than or equal</td><td>|</td><td>OR</td></tr>
<tr><td>></td><td>greater than</td><td>~</td><td>NOT</td></tr>
<tr><td>>=</td><td>greater than or equal</td><td></td><td></td></tr>
<tr><td>==</td><td>equal</td><td></td><td></td></tr>
<tr><td>~=</td><td>not equal</td><td></td><td></td></tr>
</table>

Flow Control Commands	
`if`	conditionally execute statements
`elseif`	used with if
`else`	used with if
`end`	terminates if, for, while statements
`for`	repeat statements for a number of times
`while`	do while
`break`	breaks out of for and while loops
`return`	return from functions
`pause`	pause

Programming	
`input`	get numbers from keyboard
`keyboard`	call keyboard as m-file
`error`	display error message
`function`	define function
`eval`	interpret text as command
`feval`	evaluate function given be string
`echo`	enable command echoing
`exist`	check if variable exists
`nargin`	get the number of input arguments
`nargout`	get the number of output arguments
`menu`	select item from menu
`etime`	get the elapsed time

This is only a partial list taken from the reference section of the MATLAB manual.

6.5 M-files : Some final notes

M-files can call other m-files. In fact, it is recommended that you break down a program into several routines.

Take advantage of the help section at the top of the file. It is the curse of programming that you will forget what your own functions do after a while. A well written help section will prevent many headaches for you.

If you need more instruction, the MATLAB manual has a good description of how to use m-files. In addition, you can look at the Mathworks' m-files for many of the commands. These are often helpful in finding programming tricks...

6.6 Useful M-files for Mechatronics

(1) Stepmesh.m

```
% This m-file computes the step response of a second-order system
% for values of zeta ranging from 0.1 to 1 and will create a mesh
% plot of the step responses. The natural frequency is set at 1.

n=1;
y=zeros(200,1);
i=1;
```

```
for del=0.1:0.1:1
      d=[1 2*del 1];
      t=[0:0.1:19.9]';
      y(:,i)=step(n,d,t);
      i=i+1;
end
mesh(fliplr(y), [-120 30])
```

(2) Stepchar.m

```
function[pos,tr,ts,tp]=stepchar(t,y)
% This m-file computes the % overshoot, peak time, rise time,
% and 1% settling time for a unit step response.
[mp,ind]=max(y); dimt=length(t); yss=y(dimt);
pos=100*(mp-yss)/yss; tp=t(ind);
for i=1:dimt,
      if y(i) > 1.01*yss,
            ts=t(i);
      elseif y(i) < 0.99*yss,
            ts=t(i);
      end
      end
for i=1:dimt
      if y(i) < 0.1*yss
            t1=t(i);
      elseif y(i)==mp,
            break;
      end
end
for i=1:dimt;
      if y(i) < 0.9*yss,
            t2=t(i);
      elseif y(i)==mp,
            break
      end;
end
tr=t2-t1;
```

(3) Stepmat.m

```
function [pos,tr,ts,tp]=stepmat(t,y)
% This program finds the step response characteristics
% for any number of step responses.
% The columns of y are the step responses, i.e., y=[y1 y2 ...]
[dimt,col_y]=size(y);
```

```
% This is the main loop to search over the columns of y.
      for ii=1:col_y
            yss=y(dimt,:);
% Finding the rise time, tr
      ind1=find(y(:,ii)<=0.1*yss(ii)); max(ind1); t1=t(ans);
      ind2=find(y(:,ii)<=0.9*yss(ii)); max(ind2); t2=t(ans);
      tr(ii)=t2-t1;
% Finding the 1 percent settling time, ts
      jj=dimt;
            for jj=1:dimt;
            if y(jj,ii)>=1.01*yss(ii); ts(ii)=t(jj);
            elseif y(jj,ii)<=0.99*yss(ii); ts(ii)=t(jj);
            end
      end
end
% Finding the percent overshoot, pos, and the peak time, tp.
[mp,ind]=max(y); tp=t(ind)'; pos=100*(max(y)-yss)./yss;
```

(4) Freqmat.m

```
function [mr,bw]=freqmat(w,m)
% This program calculates Mr and BW from the magnitude Bode plot.
% Use [m,w]=bode(sys); to gather the data.
% Columns of m are the absolute magnitudes, i.e., m=[m1 m2 ...]
[dimw,col_m]=size(m); mdb=ones(dimw,col_m); bw=ones(1,col_m);
for ii=1:col_m;
      jj=1;
      while m(jj,ii)>m(1,ii)/sqrt(2);

            jj=jj+1;
      end
bw(ii)=w(jj);
end
mr=20*log10(max(m));
```

Brushless Direct-Current Motors

- Features Common to Rotating Magnetic Field Electromechanical Devices
  - Introduction
  - Windings
  - Air Gap mmf – Sinusoidally-Distributed Windings
  - Rotating Air Gap mmf – Two-Pole Devices
- Introduction to Several Electromechanical Motion Devices
  - Reluctance Devices
  - Induction Machines

- Synchronous Machines
- Permanent-Magnet Devices

- Brushless DC Motors
  - Introduction
  - Two-Phase Permanent-Magnet Synchronous Machine
  - Voltage Equations and Winding Inductances
  - Torque
  - Machine Equations in the Rotor Reference Frame
  - Time-Domain Block Diagrams and State Equations

Features Common to Rotating Magnetic Field Electromechanical Devices

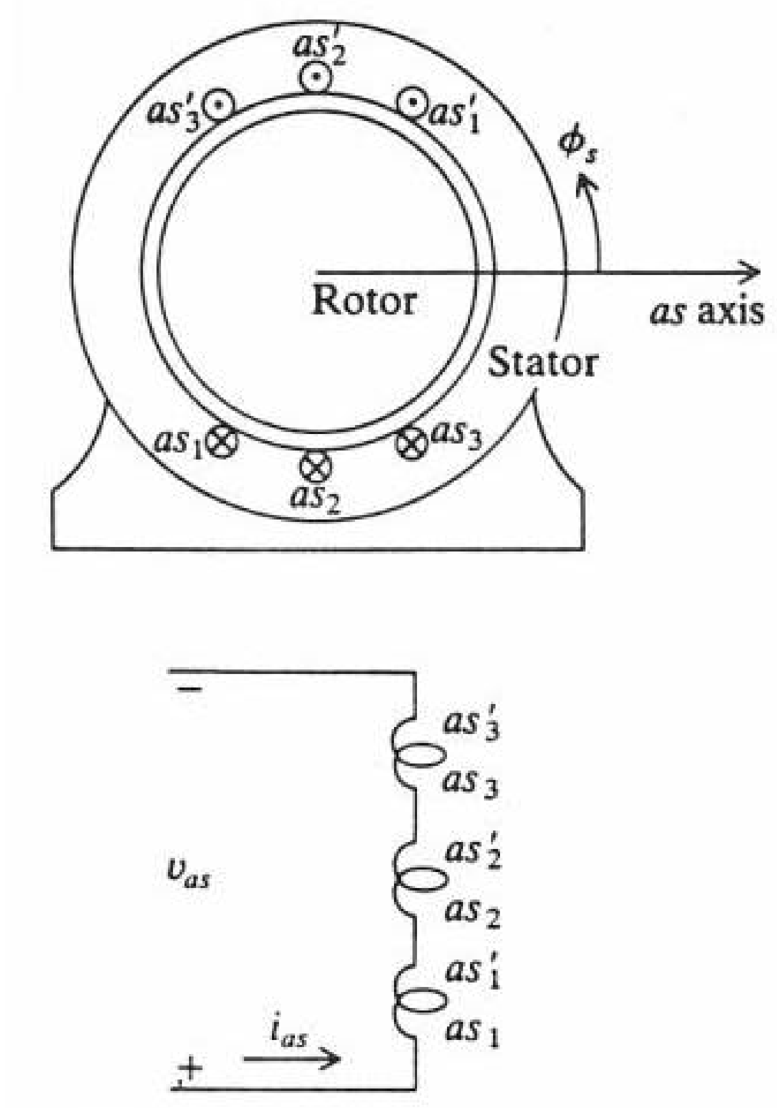
- Introduction
  - A *dc* machine has windings on both the stationary and rotating members, and these circuits are in relative motion whenever the armature (rotor) rotates. However, due to the action of the commutator, the resultant mmf produced by currents flowing in the rotor windings is stationary.
  - The rotor windings appear to be stationary, magnetically.
  - With constant current in the field (stator) winding, torque is produced and rotation results owing to the force established to align two stationary, orthogonal magnetic fields.

- In rotational electromechanical devices other than *dc* machines, torque is produced as a result of one or more magnetic fields which rotate about the air gap of the device.
- Reluctance machines, induction machines, synchronous machines, stepper motors, and brushless dc motors (permanent-magnet synchronous machines), all develop torque in this manner.
- There are features of these devices which are common to all, in particular:
  - Winding arrangement of the stator
  - Method of producing a rotating magnetic field due to stator currents
- Hence, we cover these common features now.

- Windings
  - Consider the diagram of the elementary two-pole, single-phase stator winding.
  - Winding *as* is assumed distributed in slots over the inner circumference of the stator, which is more characteristic of the stator winding than is a concentrated winding.
  - The winding is depicted as a series of individual coils. Each coil is placed in a slot in the stator steel.
  - Follow the path of positive current i_{as} flowing in the *as* winding.
  - Note that as_1 and as_1' are placed in stator slots which span π radians; this is characteristic of a two-pole machine.
  - as_1 around to as_1' is referred to as a coil; as_1 or as_1' is a coil side. In practice a coil will contain more than one conductor.

- The number of conductors in a coil side tells us the number of turns in this coil. This number is denoted as nc_s.
- Repeat this winding process to form the $as_2 - as_2'$ coil and the $as_3 - as_3'$ coil, assuming that the same number of turns, nc_s, make up each coil.
- With the same number of turns in each of these coils, the winding is said to be distributed over a span from as_1 to as_3 or 60°.
- The right-hand rule is used to give meaning to the *as* axis; it is the principal direction of magnetic flux established by positive current flowing in the *as* winding. It is said to denote the positive direction of the magnetic axis of the *as* winding.

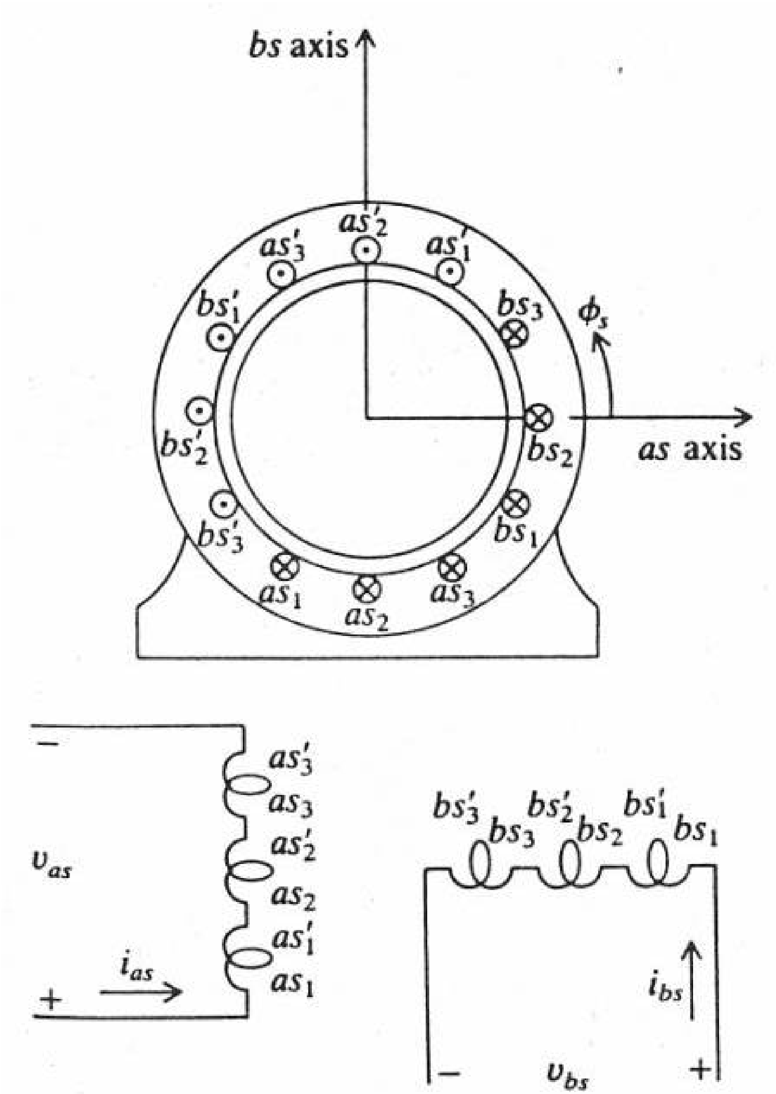
Elementary Two-Pole, Single-Phase Stator Winding



- Now, consider the diagram of the elementary two-pole, two-phase stator windings. Here we have added a second winding – the *bs* winding.
- The magnetic axis of the *bs* winding is displaced ½ π from that of the *as* winding.
- Assume that the positive direction of i_{bs} is such that the positive magnetic axis of the *bs* winding is at $\phi_s = \frac{1}{2}\pi$ where ϕ_s is the angular displacement about the stator referenced to the *as* axis.
- This is the stator configuration for a two-pole, two-phase electromechanical device.
- The stator windings are said to be *symmetrical* (as it is used in electromechanical devices) if the number of turns per coil and resistance of the *as* and *bs* windings are identical.

– For a two-pole, three-phase, symmetrical electromechanical device, there are three identical stator windings displaced 120° from each other. Essentially all multiphase electromechanical devices are equipped with symmetrical stators.

Elementary Two-Pole, Two-Phase Stator Windings



- Air Gap mmf – Sinusoidally-Distributed Windings
  - It is generally assumed that the stator windings (and in many cases the rotor windings) may be approximated as sinusoidally-distributed windings.
  - The distribution of a stator phase winding may be approximated as a sinusoidal function of ϕ_s, and the waveform of the resulting mmf dropped across the air gap (air gap mmf) of the device may also be approximated as a sinusoidal function of ϕ_s.
  - To establish a truly sinusoidal air gap mmf, the winding must also be distributed sinusoidally, and it is typically assumed that all windings may be approximated as sinusoidally-distributed windings.

– In the figure, we have added a few coils to the *as* winding, which now span 120°.

– For the purpose of establishing an expression for the air gap mmf, we employ the developed diagram of the cross-sectional view obtained by "flattening out" the rotor and stator.

– Note that displacement ϕ_s is defined to the left of the *as* axis since this allows us to position the stator above the rotor.

– The winding distributions may be approximated as:

$$N_{as} = N_p \sin\phi_s \qquad \text{for } 0 < \phi_s < \pi$$

$$N_{as} = -N_p \sin\phi_s \qquad \text{for } \pi < \phi_s < 2\pi$$

– N_p is the peak turns density in turns/radian.

– If N_s represents the number of turns of the equivalent sinusoidally distributed winding (not the total turns of the winding) that corresponds to the fundamental component of the actual winding distribution, then:

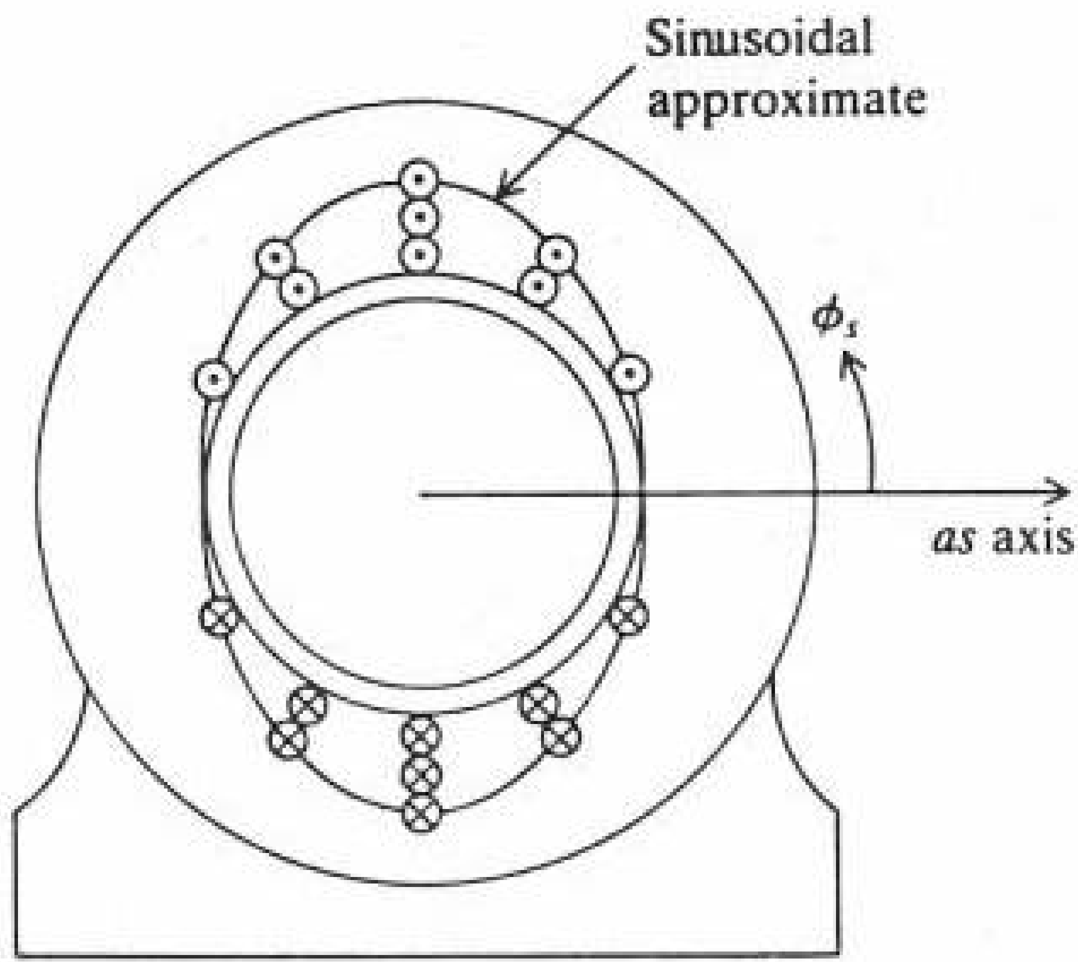
$$N_s = \int_0^{\pi} N_p \sin(\phi_s)\, d\phi_s = 2N_p$$

– The sinusoidally-distributed winding will produce a mmf that is positive in the direction of the *as* axis (to the right in the figure for positive i_{as}).

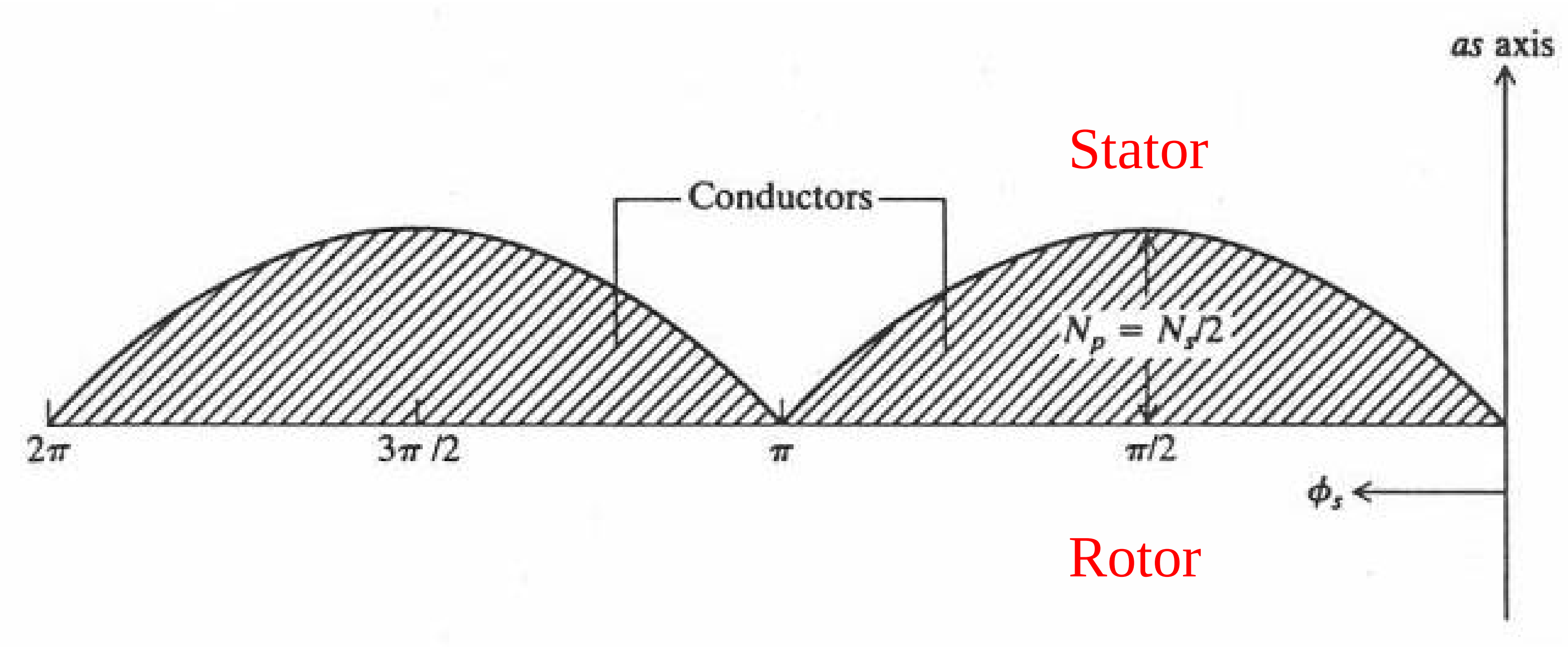
– We assume that all of the mmf is dropped across the air gap, as the reluctance of the steel is much smaller (neglecting saturation) than the reluctance of the air gap.

– So if the windings are sinusoidally-distributed in space, then the mmf dropped across the air gap will also be sinusoidal in space.

Approximate Sinusoidal Distribution of the *as* Winding



Developed Diagram with Sinusoidally-Distributed Stator Winding



- We need to develop an expression for the air gap mmf, mmf_{as}, associated with the *as* winding. We will apply Ampere's Law to two closed paths, shown in the diagram.
- For closed path (a), the total current enclosed is $N_s i_{as}$ and, by Ampere's Law, this is equal to the mmf drop around the given path ($\oint \vec{H} \cdot \vec{dL}$).
- If the reluctance of the rotor and stator steel is small compared with the air-gap reluctance, we can assume that ½ of the mmf is dropped across the air gap at $\phi_s = 0$ and ½ at $\phi_s = \pi$.
- By definition mmf_{as} is positive for a mmf drop across the air gap from the rotor to the stator. Thus mmf_{as} is positive at $\phi_s = 0$ and negative at $\phi_s = \pi$, assuming positive i_{as}.

– This suggests that for arbitrary ϕ_s, mmf_{as} might be expressed as:

$$mmf_{as}(0) = \frac{N_s}{2} i_{as}$$

$$mmf_{as}(\pi) = -\frac{N_s}{2} i_{as}$$

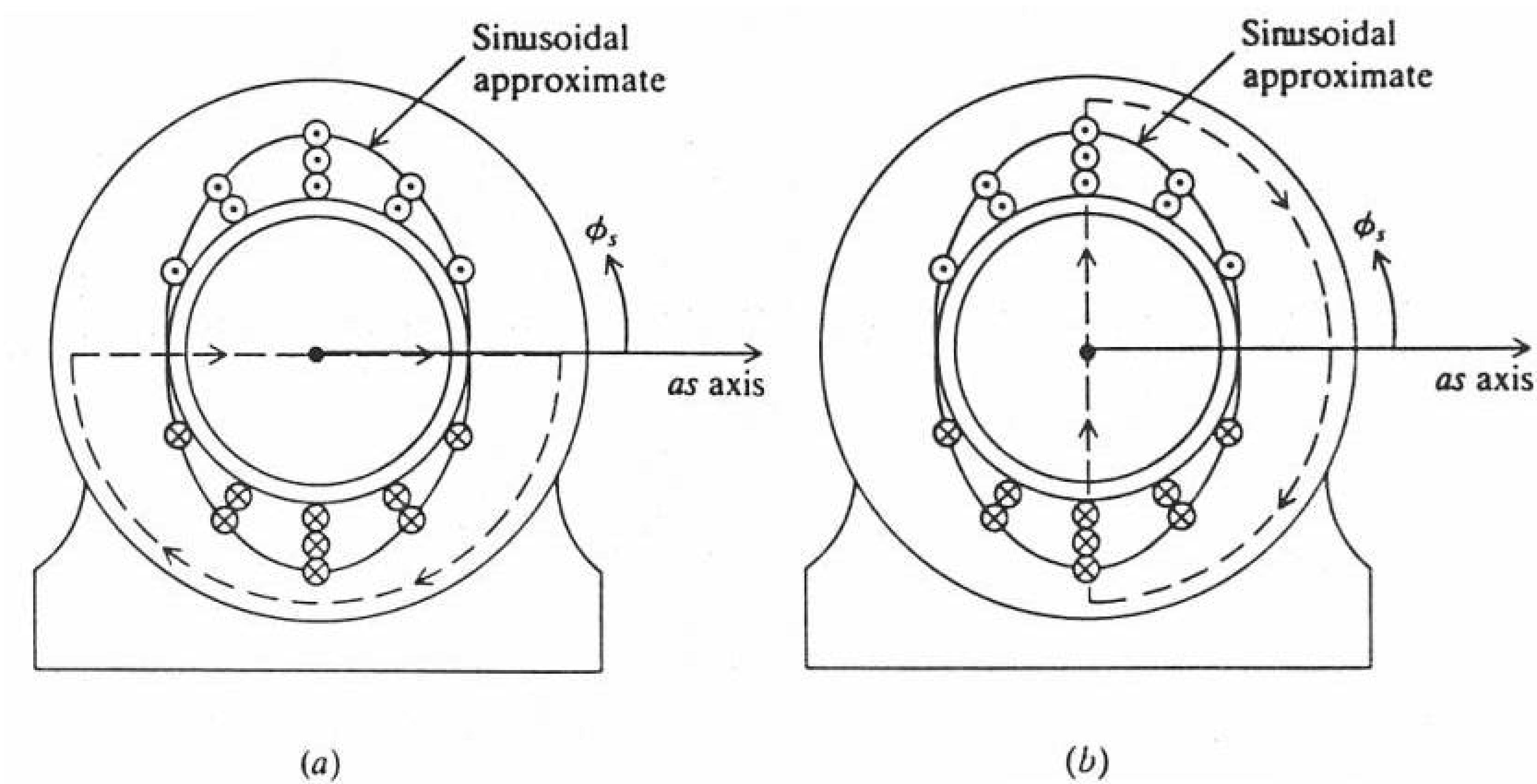
$$\Rightarrow \quad mmf_{as} = \frac{N_s}{2} i_{as} \cos\phi_s$$

– This tells that the air gap mmf is zero at $\phi_s = \pm \frac{1}{2}\pi$. Check this by applying Ampere's Law to the second closed path in the figure, path (b). The net current enclosed is zero, and so the mmf drop is zero along the given path, implying that $mmf_{as} = 0$ at $\phi_s = \pm \frac{1}{2}\pi$.

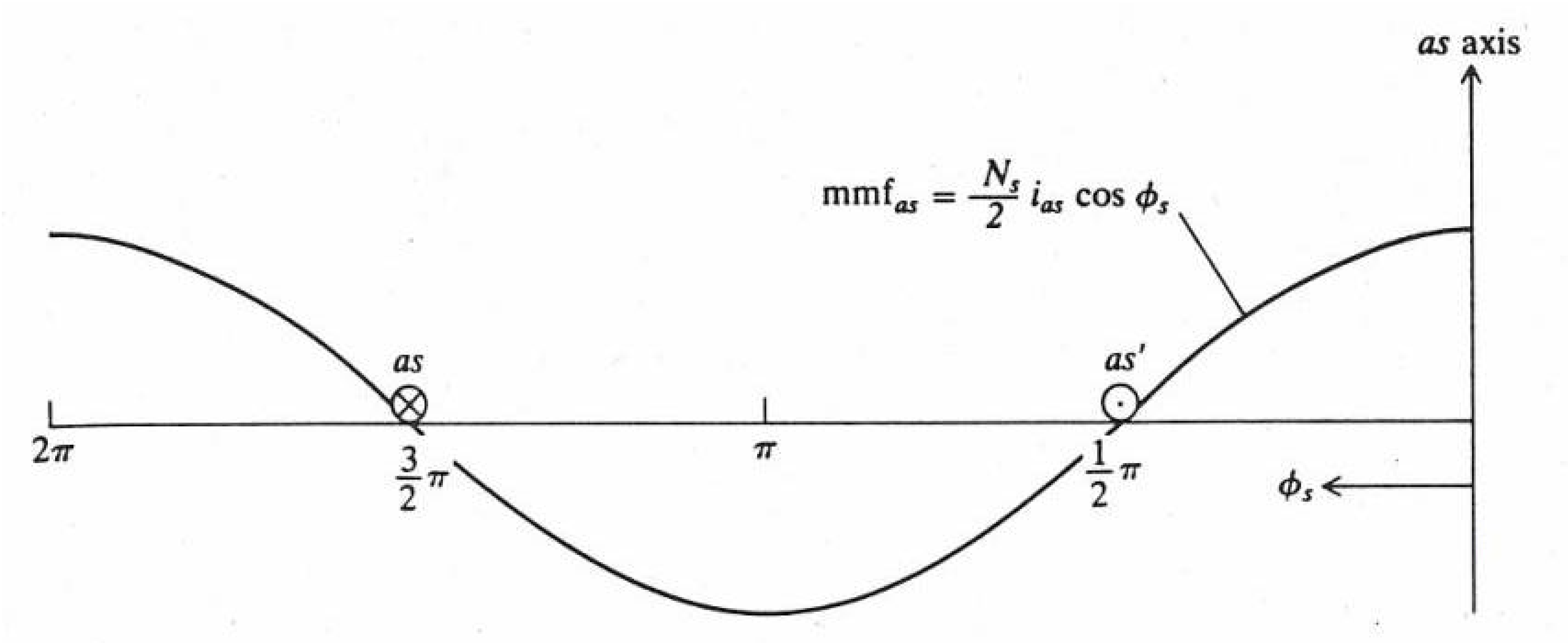
– Let's consider the *bs* winding of a two-phase device. The air gap mmf due to a sinusoidally-distributed *bs* winding may be expressed as:

$$\text{mmf}_{bs} = \frac{N_s}{2} i_{bs} \sin \phi_s$$

Closed Paths used to Establish mmf_{as}



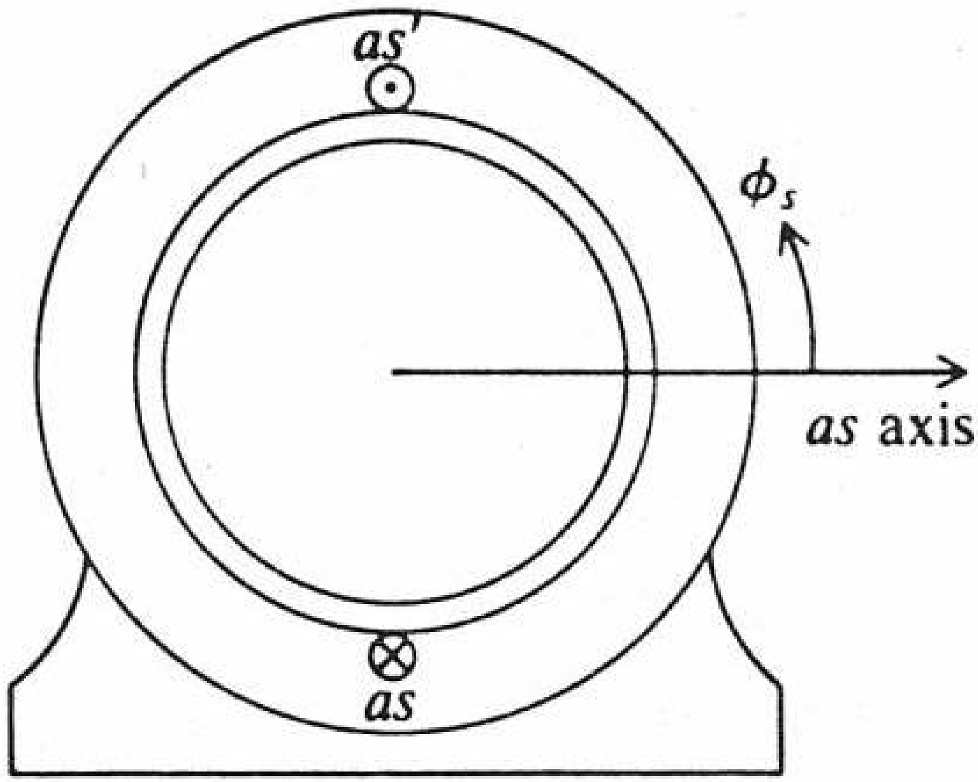
A mmf_{as} Due to Sinusoidally Distributed *as* Winding



- <u>Rotating Air Gap mmf – Two-Pole Devices</u>
  – Considerable insight into the operation of electromechanical motion devices can be gained from an analysis of the air gap mmf produced by current flowing in the stator winding(s).
  – Let's consider the rotating air gap mmf's produced by currents flowing in the stator windings of single-, two-, and three-phase devices.

- *Single-Phase Devices*
  - Consider the device shown which illustrates a single-phase stator winding. Assume the *as* winding is sinusoidally distributed, with *as* and *as*´ placed at the point of maximum turns density.
  - Assume that the current flowing in the *as* winding is a constant. Then the *as* winding would establish a stationary magnetic system with a N pole from $0.5\pi < \phi_s < 1.5\pi$ and a S pole from $-0.5\pi < \phi_s < 0.5\pi$.
  - The air gap mmf is directly related to these poles; indeed, the flux flowing from the N pole and into the S pole is caused by the air gap mmf.

Elementary Two-Pole, Single-Phase Sinusoidally-Distributed Stator Winding



– What happens when the current flowing in the *as* winding is a sinusoidal function of time? Let's assume steady-state operation:

$$I_{as} = \sqrt{2} I_s \cos\left[\omega_e t + \theta_{esi}(0)\right]$$

– Capital letters denote steady-state instantaneous variables; I_s is the rms value of the current; ω_e is the electrical angular velocity; $\theta_{esi}(0)$ is the angular position corresponding to the time zero value of the instantaneous current.

– The air gap mmf expressed for the *as* winding is:

$$mmf_{as} = \frac{N_s}{2} i_{as} \cos\phi_s = \frac{N_s}{2} \sqrt{2} I_s \cos\left[\omega_e t + \theta_{esi}(0)\right] \cos\phi_s$$

– Consider this expression for a moment. It appears that all we have here is a stationary, pulsating magnetic field. Let us rewrite this expression using a trig identity:

$$\text{mmf}_{as} = \frac{N_s}{2}\sqrt{2}I_s\left\{\frac{1}{2}\cos\left[\omega_e t + \theta_{esi}(0) - \phi_s\right] + \frac{1}{2}\cos\left[\omega_e t + \theta_{esi}(0) + \phi_s\right]\right\}$$

– The arguments of the cosine terms are functions of time and displacement ϕ_s. If we can make an argument constant, then the cosine of this argument would be constant.

$$\left.\begin{aligned} \omega_e t + \theta_{esi}(0) - \phi_s &= C_1 \\ \omega_e t + \theta_{esi}(0) + \phi_s &= C_2 \end{aligned}\right\} \qquad \begin{aligned} \frac{d\phi_s}{dt} &= \omega_e \\ \frac{d\phi_s}{dt} &= -\omega_e \end{aligned}$$

– What does this mean?

– If you run around the air gap in CCW direction at an angular velocity ω_e, the first term in the expression for mmf_{as} will appear as a constant mmf and hence a constant set of N and S poles. On the other hand, if we run CW at ω_e, the second term in the expression for mmf_{as} will appear as a constant mmf.

– In other words, the pulsating air gap mmf we noted standing at $\phi_s = 0$ (or any fixed value of ϕ_s) can be thought of as two, one-half amplitude, oppositely-rotating air gap mmf's (magnetic fields), each rotating at the angular speed of ω_e, which is the electrical angular velocity of the current.

– Since we have two oppositely rotating sets of N and S poles (magnetic fields), it would seem that the single-phase machine could develop an average torque as a result of interacting with either.

– A single-phase electromechanical device with the stator winding as shown can develop an average torque in either direction of rotation.
– Note that this device is a two-pole device, even though there are two two-pole sets, as only one set interacts with the rotor to produce a torque with a nonzero average.

- *Two-Phase Devices*
  - Consider the two-pole, two-phase sinusoidally distributed stator windings shown.
  - For balanced (i.e., variables are equal-amplitude sinusoidal quantities and 90° out of phase) steady-state conditions, the stator currents may be expressed as:

$$I_{as} = \sqrt{2} I_s \cos\left[\omega_e t + \theta_{esi}(0)\right]$$

$$I_{bs} = \sqrt{2} I_s \sin\left[\omega_e t + \theta_{esi}(0)\right]$$

  - The reason for selecting this set of stator currents will become apparent.
  - The total air gap mmf due to both stator windings (assumed to be sinusoidally distributed) may be expressed by adding mmf_{as} and mmf_{bs} to give mmf_s.

– The total air gap mmf due to the stator windings is:

$$\left.\begin{aligned} mmf_{as} &= \frac{N_s}{2} i_{as} \cos\phi_s \\ mmf_{bs} &= \frac{N_s}{2} i_{bs} \sin\phi_s \end{aligned}\right\} mmf_s = \frac{N_s}{2}\left(i_{as}\cos\phi_s + i_{bs}\sin\phi_s\right)$$

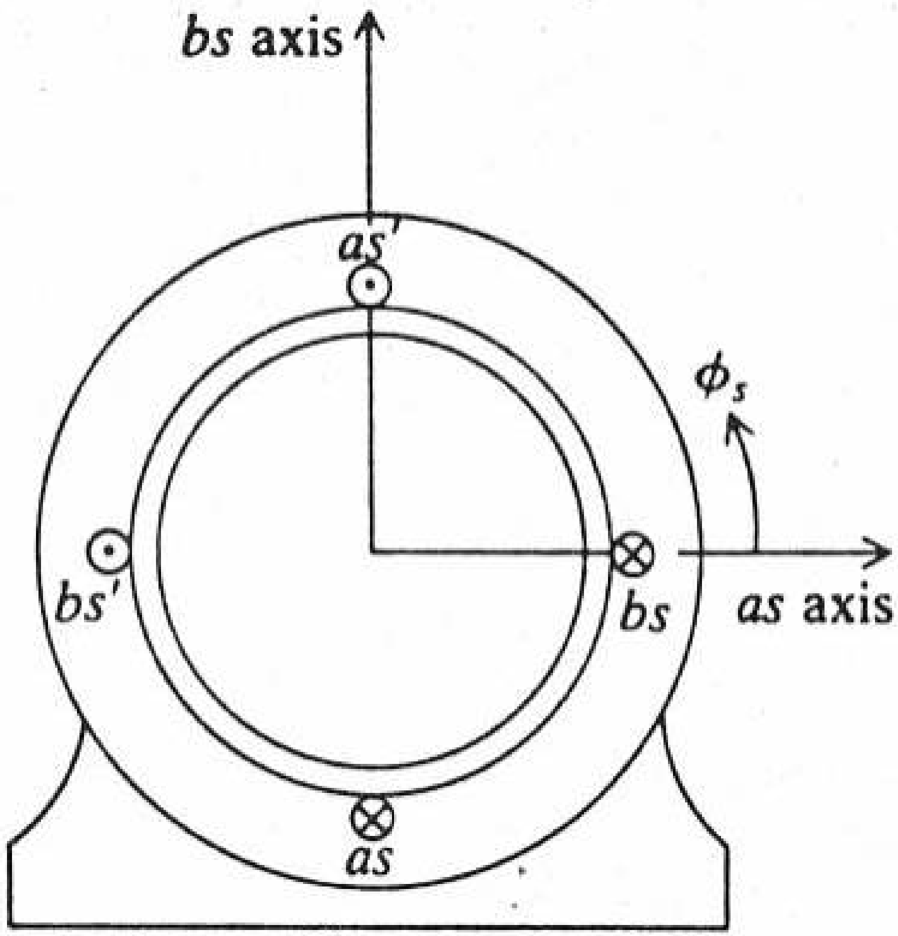
– Substitution:

$$\left.\begin{aligned} I_{as} &= \sqrt{2} I_s \cos\left[\omega_e t + \theta_{esi}(0)\right] \\ I_{bs} &= \sqrt{2} I_s \sin\left[\omega_e t + \theta_{esi}(0)\right] \end{aligned}\right\} mmf_s = \frac{N_s}{2}\left(i_{as}\cos\phi_s + i_{bs}\sin\phi_s\right)$$

– Result:

$$mmf_s = \frac{N_s}{2}\sqrt{2} I_s \cos\left[\omega_e t + \theta_{esi}(0) - \phi_s\right]$$

Elementary Two-Pole, Two-Phase Sinusoidally-Distributed Stator Winding



– It is interesting to note that we have only one rotating air gap mmf or rotating magnetic field.

– Set the argument equal to a constant, take the derivative with respect to time, and we find that the argument is constant if

$$\frac{d\phi_s}{dt} = \omega_e$$

– If we travel around the air gap in the CCW direction at ω_e, we will always see a constant mmf_s for the balanced set of currents

$$I_{as} = \sqrt{2} I_s \cos\left[\omega_e t + \theta_{esi}(0)\right]$$

$$I_{bs} = \sqrt{2} I_s \sin\left[\omega_e t + \theta_{esi}(0)\right]$$

– Hence a single rotating air gap mmf is produced. The actual value that we would see as we travel around the air gap at ω_e would depend upon the selection of time zero and our position on the stator at time zero.

- With the assigned positive direction of current in the given arrangement of the *as* and *bs* windings shown, the balanced set of stator currents produces a mmf_s that rotates CCW, which is desired for conventional purposes.
- In the case of the single-phase stator winding with a sinusoidal current, the air gap mmf can be thought of as two oppositely-rotating, constant-amplitude mmf's. However, the instantaneous air gap mmf is pulsating even when we are traveling with one of the rotating air gap mmf's. Unfortunately, this pulsating air gap mmf or set of poles gives rise to steady-state pulsating components of electromagnetic torque.

– In the case of the two-phase stator with balanced currents, only one rotating air gap mmf exists. Hence, the steady-state electromagnetic torque will not contain a pulsating or time-varying component; it will be a constant with the value determined by the operating conditions.

- *Three-Phase Devices*
  - The stator windings of a two-pole, three-phase device are shown in the figure.
  - The windings are identical, sinusoidally distributed with N_s equivalent turns and with their magnetic axes displaced 120°; the stator is symmetrical.
  - The positive direction of the magnetic axes is selected so as to achieve counterclockwise (CCW) rotation of the rotating air gap mmf with balanced stator currents of the *abc* sequence.

– The air gap mmf's established by the stator windings may be expressed by inspection as:

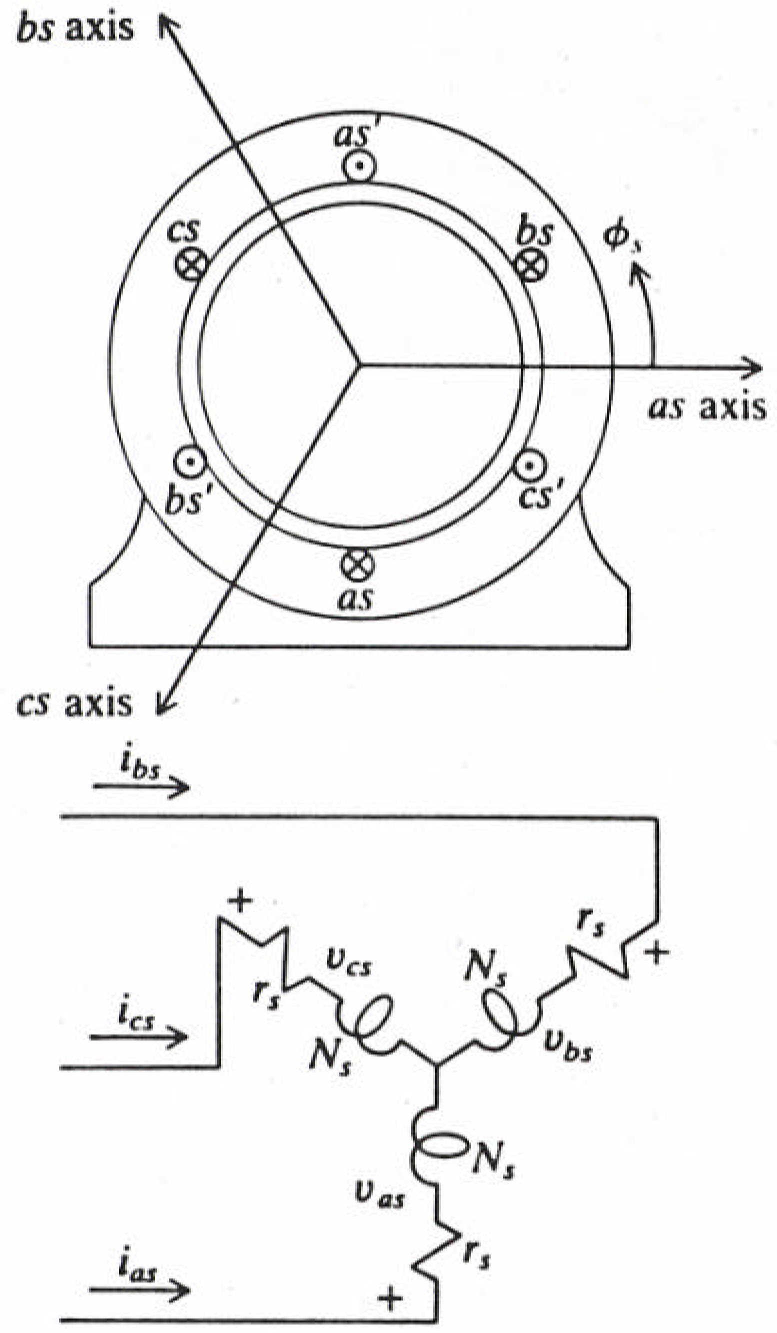
$$\mathrm{mmf}_{as} = \frac{N_s}{2} i_{as} \cos\phi_s$$

$$\mathrm{mmf}_{bs} = \frac{N_s}{2} i_{bs} \cos\left(\phi_s - \frac{2}{3}\pi\right)$$

$$\mathrm{mmf}_{cs} = \frac{N_s}{2} i_{cs} \cos\left(\phi_s + \frac{2}{3}\pi\right)$$

– As before, N_s is the number of turns of the equivalent sinusoidally distributed stator windings and ϕ_s is the angular displacement about the stator.

Elementary Two-Pole, Three-Phase Sinusoidally-Distributed Stator Windings



– For balanced steady-state conditions, the stator currents for an *abc* sequence may be expressed as:

$$I_{as} = \sqrt{2} I_s \cos\left[\omega_e t + \theta_{esi}(0)\right]$$

$$I_{bs} = \sqrt{2} I_s \cos\left[\omega_e t - \frac{2}{3}\pi + \theta_{esi}(0)\right]$$

$$I_{cs} = \sqrt{2} I_s \cos\left[\omega_e t + \frac{2}{3}\pi + \theta_{esi}(0)\right]$$

– Substitution:

$$I_{as} = \sqrt{2} I_s \cos\left[\omega_e t + \theta_{esi}(0)\right]$$

$$I_{bs} = \sqrt{2} I_s \cos\left[\omega_e t - \frac{2}{3}\pi + \theta_{esi}(0)\right]$$

$$I_{cs} = \sqrt{2} I_s \cos\left[\omega_e t + \frac{2}{3}\pi + \theta_{esi}(0)\right]$$

$$\mathrm{mmf}_{as} = \frac{N_s}{2} i_{as} \cos\phi_s$$

$$\mathrm{mmf}_{bs} = \frac{N_s}{2} i_{bs} \cos\left(\phi_s - \frac{2}{3}\pi\right)$$

$$\mathrm{mmf}_{cs} = \frac{N_s}{2} i_{as} \cos\left(\phi_s + \frac{2}{3}\pi\right)$$

– Add the resulting expressions to yield an expression for the rotating air gap mmf established by balanced steady-state currents flowing in the stator windings:

$$\text{mmf}_s = \frac{N_s}{2}\sqrt{2}I_s\frac{3}{2}\cos\left[\omega_e t + \theta_{esi}(0) - \phi_s\right]$$

– Compare this with the mmf_s for a two-phase device:

$$\text{mmf}_s = \frac{N_s}{2}\sqrt{2}I_s\frac{3}{2}\cos\left[\omega_e t + \theta_{esi}(0) - \phi_s\right]$$ 3-Phase

$$\text{mmf}_s = \frac{N_s}{2}\sqrt{2}I_s\cos\left[\omega_e t + \theta_{esi}(0) - \phi_s\right]$$ 2-Phase

– They are identical except that the amplitude of the mmf for the 3-phase device is 3/2 times that of a 2-phase device.

– It can be shown that this amplitude for multiphase devices changes from that of a two-phase device in proportion to the number of phases divided by 2.
– It is important to note that with the selected positive directions of the magnetic axes a counterclockwise rotating air gap mmf is obtained with a three-phase set of balanced stator currents of the *abc* sequence.

Introduction to Several Electromechanical Motion Devices

- Rotational electromechanical devices fall into three general classes:
  - Direct-current
  - Synchronous
  - Induction
- We have already covered *dc* machines.
- Synchronous Machines
  - They are so called because they develop an average torque only when the rotor is rotating in synchronism (synchronous speed) with the rotating air gap mmf established by currents flowing in the stator windings.

- Examples are: reluctance machines, stepper motors, permanent-magnet machines, brushless dc machines, and the machine which has become known as simply the synchronous machine.

- Induction Machines
  - Induction is the principle means of converting energy from electrical to mechanical.
  - The induction machine cannot develop torque at synchronous speed in its normal mode of application.
  - The windings on the rotor are short-circuited and, in order to cause current to flow in these windings which produce torque by interacting with the air gap mmf established by the stator windings, the rotor must rotate at a speed other than synchronous speed.

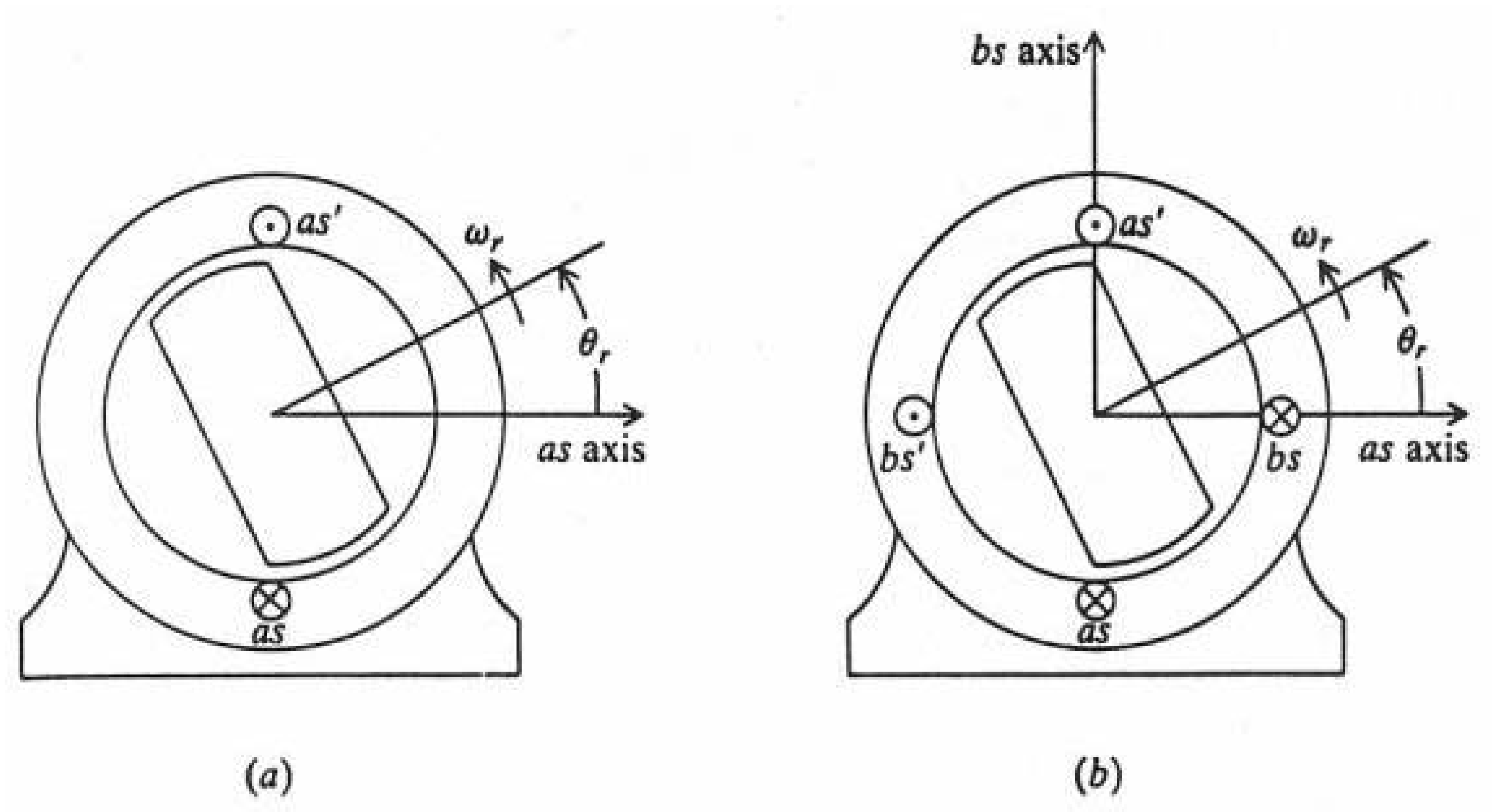
- Here we will show the winding arrangement for elementary versions of these electromechanical devices and describe briefly the principle of operation of each.

- Reluctance Drives
  - Elementary single-, and two-phase two-pole reluctance machines are shown in the figure.
  - Stator windings are assumed to be sinusoidally distributed.
  - The principal of operation is quite straightforward.
    - In an electromagnetic system a force (torque) is produced in an attempt to minimize the reluctance of the magnetic system.
    - We have established that, with an alternating current flowing in the winding of the single-phase stator, two oppositely-rotating mmf's are produced.
    - Therefore, once the rotor is rotating in synchronism with either of the two oppositely-rotating air gap mmf's, there is a force (torque) created by the magnetic system in an attempt to align the minimum-reluctance path of the rotor with the rotating air gap mmf.

- When there is no load torque on the rotor, the minimum-reluctance path of the rotor is in alignment with the rotating air gap mmf.
- When a load torque is applied, the rotor slows ever so slightly, thereby creating a misalignment of the minimum-reluctance path and the rotating air gap mmf.
- When the electromagnetic torque produced in an attempt to maintain alignment is equal and opposite to the load torque on the rotor, the rotor resumes synchronous speed.
- If the load torque is larger than the torque which can be produced to align, the rotor will fall out of synchronism and, since the machine cannot develop an average torque at a speed other than synchronous, it will slow to stall.
- The operation of a two-phase device differs from that of the single-phase device in that only one constant-amplitude rotating air gap mmf is produced during balanced steady-state conditions.
- Hence, a constant torque will be developed at synchronous speed rather than a torque which pulsates about an average value as is the case with the single-phase machine.

- Although the reluctance motor can be started from a source which can be switched at a frequency corresponding to the rotor speed as in the case of stepper or brushless dc motors, the devices cannot develop an average starting torque when plugged into a household power outlet.
- Many stepper motors are of the reluctance type. Some stepper motors are called variable-reluctance motors. Operation is easily explained. Assume that a constant current is flowing the *bs* winding of the figure with the *as* winding open-circuited. The minimum reluctance path of the rotor will be aligned with the *bs* axis, i.e., assume θ_r is zero. Now let's reduce the *bs* winding current to zero while increasing the current in the *as* winding to a constant value. There will be forces to align the minimum-reluctance path of the rotor with the *as* axis; however, this can be satisfied with $\theta_r = \pm \frac{1}{2} \pi$. There is a 50-50 chance as to which way it will rotate. We see that we need a device different from a single- or two-phase reluctance machine to accomplish controlled stepping. Two common techniques are single-stack and multistack variable-reluctance steppers.

Elementary Two-Pole Reluctance Machines: Single-Phase and Two-Phase



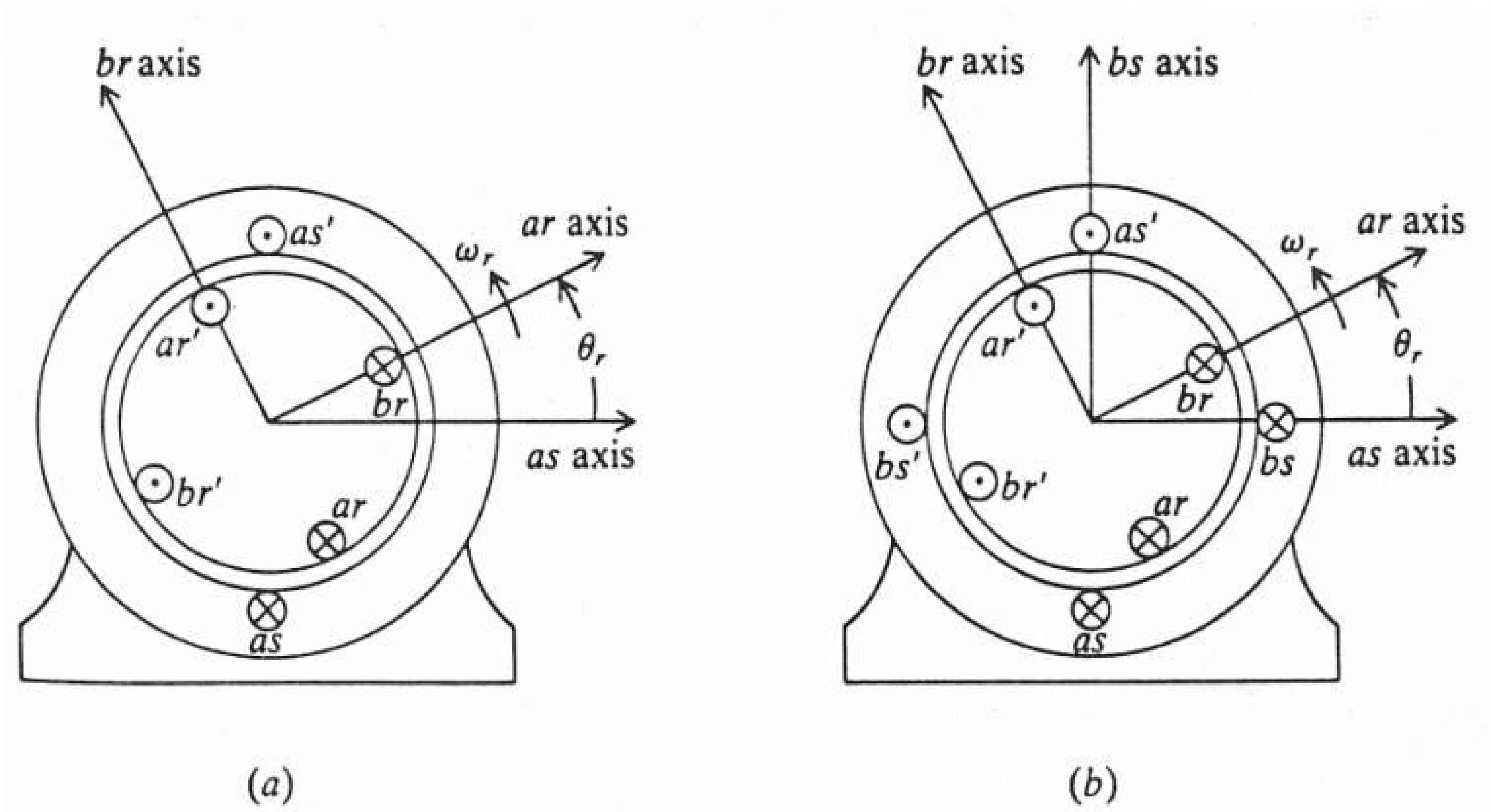
- Induction Machines
  - Elementary single- and two-phase induction machines are shown in the figure.
  - The rotors of both devices are identical in configuration; each has the equivalent of two orthogonal windings which are assumed to be sinusoidally distributed. The *ar* and *br* windings are equivalent to a symmetrical two-phase set of windings and, in the vast majority of applications, these rotor windings are short-circuited.
  - Let's look at the operation of the two-phase device first.
    - For balanced steady-state operation, the currents flowing in the stator windings produce an air gap mmf which rotates about the air gap at an angular velocity of ω_e.

- With the rotor windings short-circuited, which is the only mode of operation we will consider, a voltage is induced in each of the rotor windings only if the rotor speed ω_r is different from ω_e.
- The currents flowing in the rotor circuits due to induction will be a balanced set with a frequency equal to $\omega_e - \omega_r$, which will produce an air gap mmf that rotates at $\omega_e - \omega_r$ relative to the rotor or ω_e relative to a stationary observer.
- Hence, the rotating air gap mmf caused by the currents flowing in the stator windings induces currents in the short-circuited rotor windings which, in turn, establish an air gap mmf that rotates in unison with the stator rotating air gap mmf (mmf_s).
- Interaction of these magnetic systems (poles) rotating in unison provides the means of producing torque on the rotor.
- An induction machine can operate as a motor or a generator. However, it is normally operated as a motor.

- As a motor it can develop torque from $0 < \omega_r < \omega_e$. At $\omega_r = \omega_e$, the rotor currents are not present since the rotor is rotating at the speed of the stator rotating air gap mmf and, therefore, the rotor windings do not experience a change of flux linkages, which is, of course, necessary to induce a voltage in the rotor windings.
- The single-phase induction motor is perhaps the most widely used electromechanical device. The figure shown is not quite the whole picture of a single-phase induction motor. Recall that the single-phase stator winding produces oppositely rotating air gap mmf's of equal amplitude.
- If the single-phase induction motor is stalled, $\omega_r = 0$, and if a sinusoidal current is applied to the stator winding, the rotor will not move. This device does not develop a starting torque. Why?
- The rotor cannot follow either of the rotating mmf's since it develops as much torque to go with one as it does to go with the other. If, however, you manually turn the rotor in either direction, it will accelerate in that direction and operate normally.

- Although single-phase induction motors normally operate with only one stator winding, it is necessary to use a second stator winding to start the device. Actually single-phase induction motors we use are two-phase induction motors with provisions to switch out one of the windings once the rotor accelerates to between 60 and 80 percent of synchronous speed.
- How do we get two-phase voltages from a single-phase household supply? Well, we do not actually develop a two-phase supply, but we approximate one, as far as the two-phase motor is concerned, by placing a capacitor (start capacitor) in series with one of the stator windings. This shifts the phase of one current relative to the other, thereby producing a larger rotating air gap mmf in one direction than the other. Provisions to switch the capacitor out of the circuit is generally inside the housing of the motor.

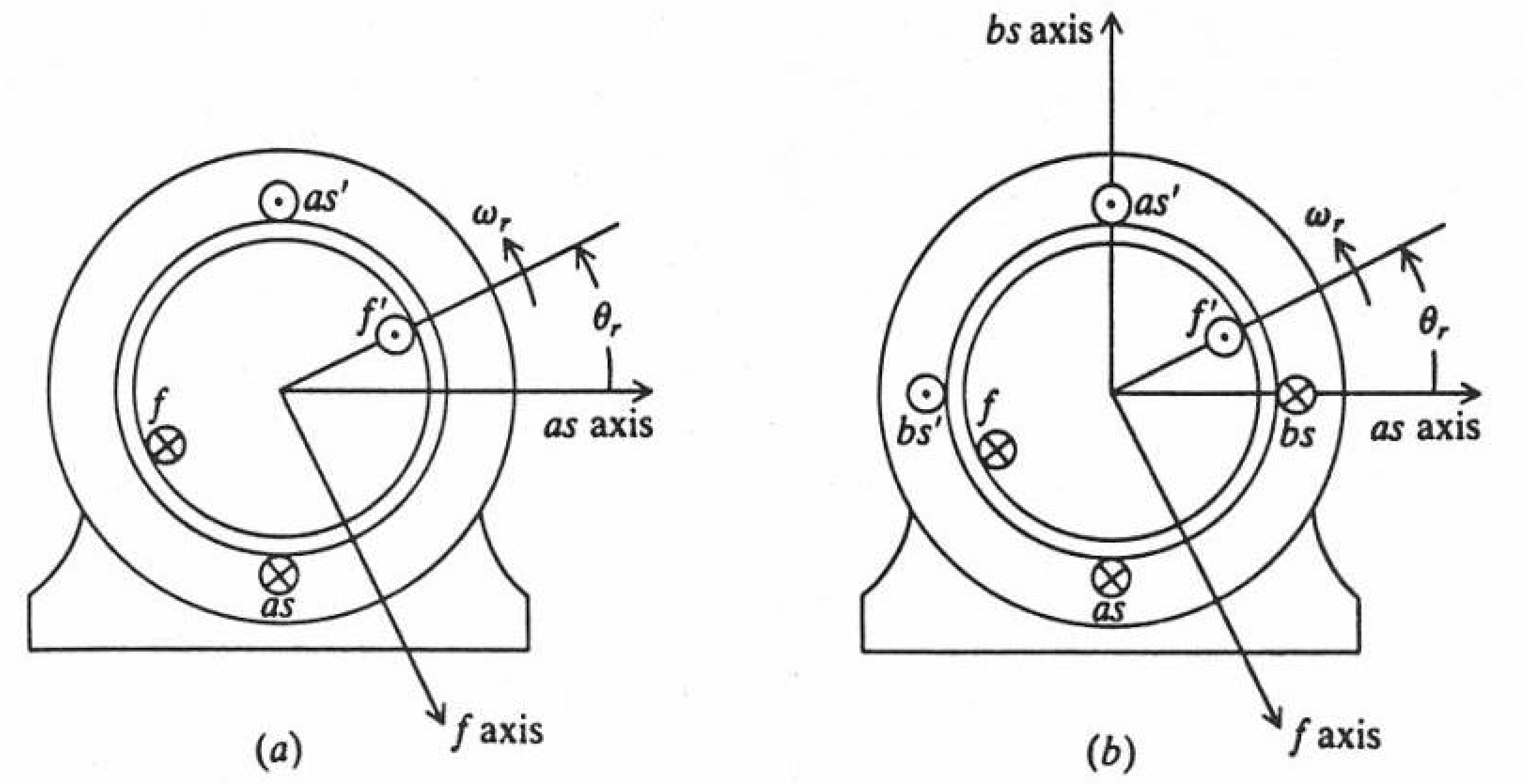
Elementary Two-Pole Induction Machines: Single-Phase and Two-Phase



- Synchronous Machines
  - Shown are elementary single- and two-phase two-pole synchronous machines. However, they are but one of several devices which fall into the synchronous machine category. We honor convention here and refer to these devices as synchronous machines.
  - The single-phase synchronous machine has limited application. The same can be said about the two-phase synchronous machine. It is the three-phase synchronous machine which is used to generate electric power in power systems such as in some automobiles, aircraft, utility systems, and ships. Nevertheless, the theory of operation of synchronous machines is adequately introduced by considering the two-phase version.

- The elementary devices shown have only one rotor winding – the field winding (*f* winding). In practical synchronous machines, the rotor is equipped with short-circuited windings in addition to the *f* winding which help to damp oscillations about synchronous speed and, in some cases, these windings are used to start the unloaded machine from stall as an induction motor.
- The principle of operation is apparent once we realize that the current flowing in the field winding is direct current. Although it may be changed in value by varying the applied field voltage, it is constant for steady-state operation of a balanced two-phase synchronous machine.
- If the stator windings are connected to a balanced system, the stator currents produce a constant-amplitude rotating air gap mmf. A rotor air gap mmf is produced by the direct current flowing in the field winding.
- To produce a torque or transmit power, the air gap mmf produced by the stator and that produced by the rotor must rotate in unison about the air gap of the machine. Hence, $\omega_r = \omega_e$.

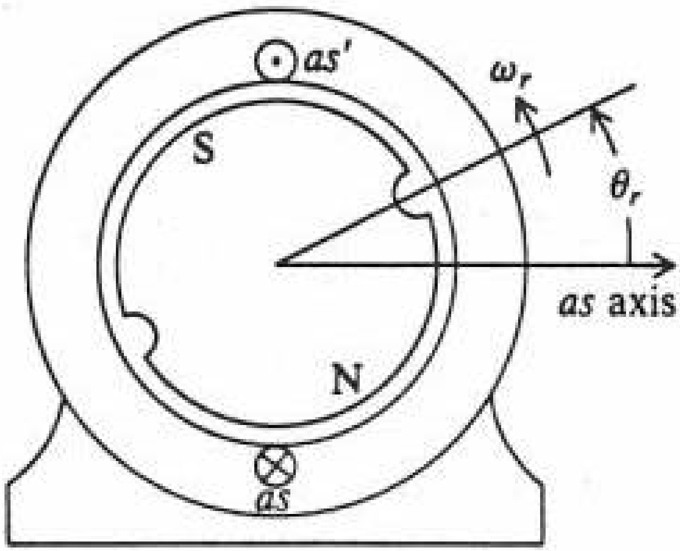
Elementary Two-Pole Synchronous Machines: Single-Phase and Two-Phase



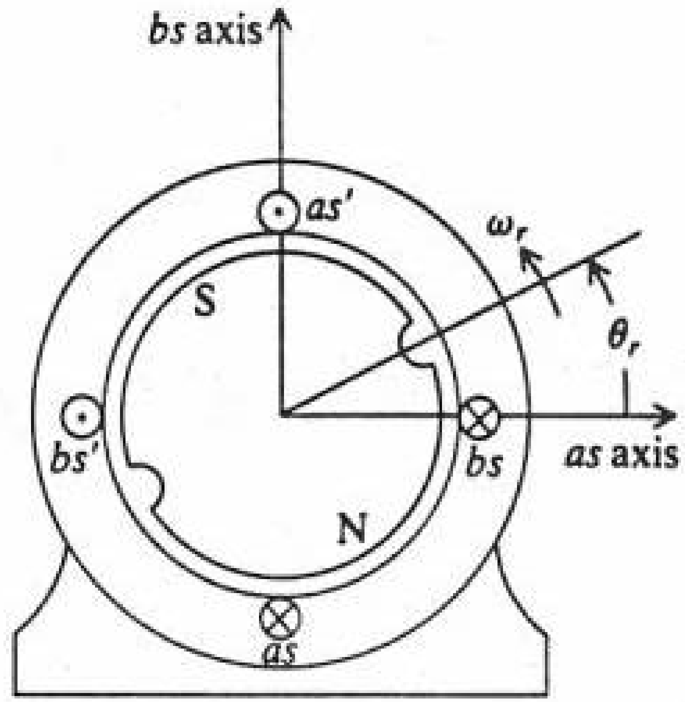
- Permanent-Magnet Devices
  - If we replace the rotor of the synchronous machines just considered with a permanent-magnet rotor, we have the so-called permanent-magnet devices shown in the figure.
  - The operation of these devices is identical to that of the synchronous machine.
  - Since the strength of the rotor field due to the permanent magnet cannot be controlled as in the case of the synchronous machine which has a field winding, it is not widely used as a means of generating power.
  - It is, however, used widely as a drive motor.

– In particular, permanent-magnet motors are used as stepper motors and, extensively, as brushless dc motors, wherein the voltages applied to the stator windings are switched electronically at a frequency corresponding to the speed of the rotor.

Elementary Two-Pole Permanent-Magnet Devices: Single-Phase and Two-Phase



(a)



(b)

Brushless DC Motors

- Introduction
  - Permanent magnet DC motors all have brushes to transmit power to the armature windings. Brush arcing causes electronic noise and maintenance problems from excessive wear.
  - A Brushless DC Motor has been developed to overcome these problems. It substitutes electronic commutation for the conventional mechanical brush commutation.
  - Because the electronic commutation exactly duplicates the brush commutation in conventional DC motors, the brushless DC motor exhibits the same linear torque-speed curve, has the same motor constants, and obeys the same performance equations.

Advantages of Brushless DC Motors

- High Reliability
  - The life of brushless DC motors is almost indefinite. Bearing failure is the most likely weak point.
- Quiet
  - A lack of mechanical noise from brushes makes it ideal for a people environment. An added advantage is that there is no mechanical friction.
- High Speed
  - Brush bounce limits DC motors to 10,000 RPM. Brushless DC motors have been developed for speeds up to 100,000 RPM, limited by the mechanical strength of the permanent magnet rotors.

- High Peak Torque
    - Brushless DC motors have windings on the stator housing. This gives efficient cooling and allows for high currents (torque) during low-duty-cycle, stop-start operation. Peak torques are more than 20 times their steady ratings compared to 10 times or less for conventional DC motors. Maximum power per unit volume can be 5 times conventional DC motors.

- Disadvantages of Brushless DC Motors
  - Cost
    - The relatively high cost of brushless DC motors is usually acceptable when considering complex machinery where normal downtime and maintenance are not only costly in itself, but often unacceptable.
  - Choice
    - Choice is restricted because there are few manufacturers.

- Types of Brushless DC Motors
  - Windings on the stator with the rotor on the inside; inside rotors have less inertia and are better suited for start-stop operation.
  - Windings on the stator with the rotor on the outside; outside rotors are better for constant load, high-speed applications.

- The brushless dc motor is becoming widely used as a small-horsepower control motor. It is a permanent-magnet synchronous machine. When it is supplied from a source, the frequency of which is always equal to the speed of its rotor, it becomes a brushless dc motor, not because it looks like a dc motor but because its operating characteristics can be made to resemble those of a dc shunt motor with a constant field current.
- How do we supply the permanent-magnet synchronous machine from a source the frequency of which always corresponds to the rotor speed?

– First we must be able to measure the rotor position. The rotor position is most often sensed by Hall-effect sensors which magnetically sense the position of the rotor poles.
– Next we must make the frequency of the source correspond to the rotor speed. This is generally accomplished with a dc-to-ac inverter, wherein the transistors are switched on and off at a frequency corresponding to the rotor speed. We are able to become quite familiar with the operating features of the brushless dc motor without getting involved with the actual inverter.
– If we assume that the stator variables (voltages and currents) are sinusoidal and balanced with the same angular velocity as the rotor speed, we are able to predict the predominant operating features of the brushless dc motor without becoming involved with the details of the inverter.

- Two-Phase Permanent-Magnet Synchronous Machine
  - A two-pole two-phase permanent-magnet synchronous machine is shown.
  - The stator windings are identical, sinusoidally distributed windings each with N_s equivalent turns and resistance r_s.
  - The magnetic axes of the stator windings are *as* and *bs* axes.
  - The angular displacement about the stator is denoted by ϕ_s, referenced to the *as* axis.
  - The angular displacement about the rotor is ϕ_r, referenced to the *q* axis.
  - The angular velocity of the rotor is ω_r and θ_r is the angular displacement of the rotor measured from the *as* axis to the *q* axis.
  - Thus $\phi_s = \phi_r + \theta_r$

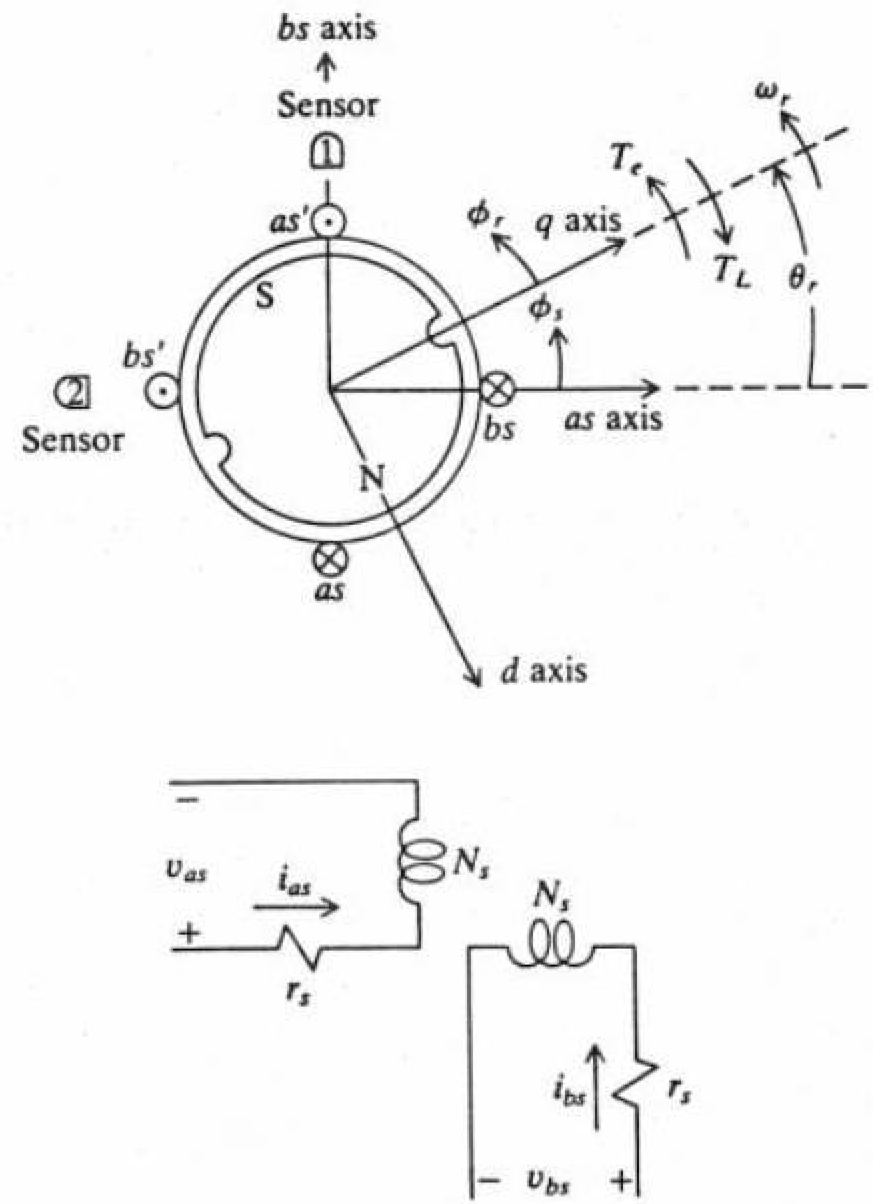
- The *d* axis (direct axis) is fixed at the center of the N pole of the permanent-magnet rotor and the *q* axis (quadrature axis) is displaced 90° CCW from the *d* axis.
- The electromechanical torque T_e is assumed positive in the direction of increasing θ_r and the load torque T_L is positive in the opposite direction.
- In the following analysis, it is assumed that:
  - The magnetic system is linear.
  - The open-circuit stator voltages induced by rotating the permanent-magnet rotor at a constant speed are sinusoidal.
  - Large stator currents can be tolerated without significant demagnetization of the permanent magnet.
  - Damper windings (short-circuited rotor windings) are not considered. Neglecting damper windings, in effect, neglects currents circulating in the surface of the rotor (eddy currents).

– The two sensors shown are Hall-effect sensors. When the N pole is under a sensor, its output is nonzero; with a S pole under the sensor, its output is zero.
– In the brushless dc motor applications, the stator is supplied from a dc-to-ac inverter the frequency of which corresponds to the rotor speed.
– The states of the sensors are used to determine the switching logic for the inverter which, in turn, determines the output frequency of the inverter.
– In the actual machine, the sensors are not positioned over the rotor. Instead, they are placed over a ring which is mounted on the shaft external to stator windings and which is magnetized by the rotor.

- The electromagnetic torque is produced by the interaction of the poles of the permanent-magnet rotor and the poles resulting from the rotating air gap mmf established by currents flowing in the stator windings.
- The rotating mmf (mmf_s) established by symmetrical two-phase stator windings carrying balanced two-phase currents is given by:

$$mmf_s = \frac{N_s}{2}\sqrt{2}I_s \cos\left[\omega_e t + \theta_{esi}(0) - \phi_s\right]$$

Two-Pole, Two-Phase Permanent-Magnet Synchronous Machine



- <u>Voltage Equations and Winding Inductances</u>
  - The voltage equations for the two-pole, two-phase permanent-magnet synchronous machine may be expressed as:

$$v_{as} = r_s i_{as} + \frac{d\lambda_{as}}{dt}$$

$$v_{bs} = r_s i_{bs} + \frac{d\lambda_{bs}}{dt}$$

  - The flux linkage equations may be expressed as:

$$\lambda_{as} = L_{asas} i_{as} + L_{asbs} i_{bs} + \lambda_{asm} \qquad \lambda_{asm} = \lambda'_m \sin\theta_r$$

$$\lambda_{bs} = L_{bsas} i_{as} + L_{bsbs} i_{bs} + \lambda_{bsm} \qquad \lambda_{bsm} = -\lambda'_m \cos\theta_r$$

  - λ'_m is the amplitude of the flux linkages established by the permanent magnet as viewed from the stator phase windings.

– In other words, the magnitude of $\lambda_m^{'}$ is proportional to the magnitude of the open-circuit sinusoidal voltage induced in each stator phase winding. Visualize the permanent-magnet rotor as a rotor with a winding carrying a constant current and in such a position to cause the N and S poles to appear as shown in the diagram.
– Assume that the air gap of the permanent-magnet synchronous machine is uniform. This may be an oversimplification.
– With this assumption of uniform air gap, the mutual inductance between the *as* and bs windings is zero.
– Since the windings are identical, the self-inductances L_{asas} and L_{bsbs} are equal and denoted as L_{ss}.

– The self-inductance is made up of a leakage and a magnetizing inductance:

$$L_{ss} = L_{\ell s} + L_{ms}$$

– The machine is designed to minimize the leakage inductance; it generally makes up approximately 10% of L_{ss}. The magnetizing inductances may be expressed in terms of turns and reluctance:

$$L_{ms} = \frac{N_s^2}{\Re_m}$$

– The magnetizing reluctance $\Re_m$ is an equivalent reluctance due to the stator steel, the permanent magnet, and the air gap. Assume that it is independent of rotor position θ_r.

- **Torque**
  - An expression for the electromagnetic torque may be obtained from:

$$T_e(\vec{i},\theta) = \frac{\partial W_c(\vec{i},\theta)}{\partial \theta}$$

  - The co-energy W_c may be expressed as:

$$W_c = \frac{1}{2} L_{ss}\left(i_{as}^2 + i_{bs}^2\right) + \lambda'_m i_{as} \sin\theta_r - \lambda'_m i_{bs} \cos\theta_r + W_{pm}$$

  - Where W_{pm} relates to the energy associated with the permanent magnet, which is constant for the device under consideration.

– Taking the partial derivative with respect to θ_r yields:

$$T_e\left(\vec{i},\theta\right) = \frac{\partial W_c\left(\vec{i},\theta\right)}{\partial \theta} = \frac{P}{2}\lambda'_m\left(i_{as}\cos\theta_r + i_{bs}\sin\theta_r\right)$$

– This expression is positive for motor action. The torque and speed may be related as:

$$T_e = J\left(\frac{2}{P}\right)\frac{d\omega_r}{dt} + B_m\left(\frac{2}{P}\right)\omega_r + T_L$$

- ## Machine Equations in the Rotor Reference Frame
  - A change of variables is helpful in the analysis of the permanent-magnet synchronous machine.
  - The objective of a change of variables is to transform all machine variables to a common reference frame, thereby eliminating θ_r from the inductance equations.
  - But the inductance L_{ss} is not a function of θ_r. However the flux linkages L_{asm} and L_{bsm} are functions of θ_r. In other words, the magnetic system of the permanent magnet is viewed as a time-varying flux linkage by the stator windings.
  - For a two-phase system the transformation is:

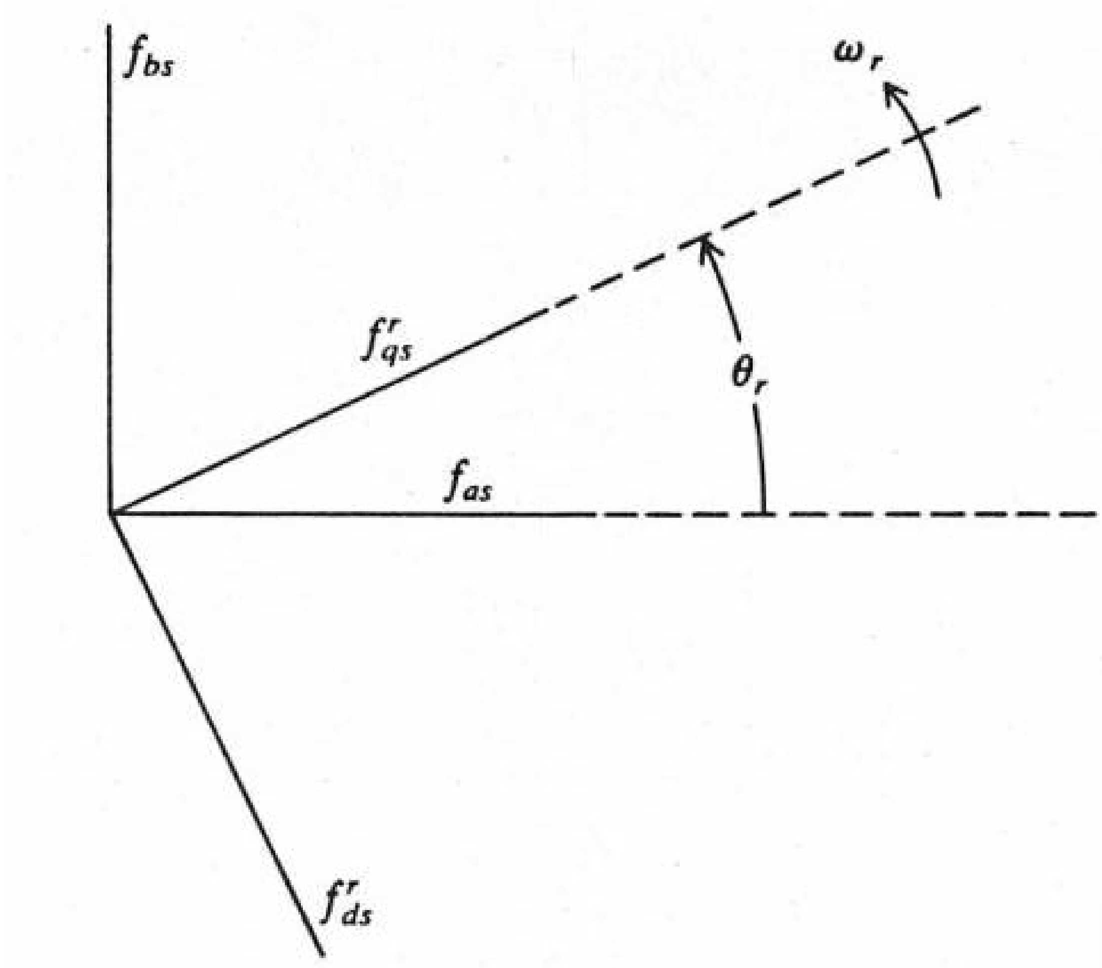
$$\begin{bmatrix} f_{qs}^r \\ f_{ds}^r \end{bmatrix} = \begin{bmatrix} \cos\theta_r & \sin\theta_r \\ \sin\theta_r & -\cos\theta_r \end{bmatrix} \begin{bmatrix} f_{as} \\ f_{bs} \end{bmatrix}$$

$$f_{qds}^r = K_s^r f_{abs}$$

$$f_{abs} = \left[K_s^r\right]^{-1} f_{qds}^r = K_s^r f_{qds}^r$$

- *f* can represent either voltage, current, or flux linkage and θ_r is the rotor displacement.
- The *s* subscript denotes stator variables and the *r* superscript indicates that the transformation is to a reference frame fixed in the rotor.
- Shown in the figure is a trigonometric interpretation of the change of stator variables.
  - The direction of f_{as} and f_{bs} variables is the positive direction of the magnetic axes of the *as* and *bs* windings, respectively.
  - The f^r_{qs} and f^r_{ds} variables are associated with fictitious windings the positive magnetic axes of which are in the same direction as the direction of f^r_{qs} and f^r_{ds}.
  - The *s* subscript denotes association with the stator variables.
  - The superscript *r* indicates that the transformation is to the rotor reference frame, which is the only reference frame used in the analysis of synchronous machines.

Trigonometric Interpretation of the Change of Stator Variables



– Transformation

$$v_{as} = r_s i_{as} + \frac{d\lambda_{as}}{dt}$$

$$v_{bs} = r_s i_{bs} + \frac{d\lambda_{bs}}{dt}$$

Matrix Form

$$\begin{bmatrix} v_{as} \\ v_{bs} \end{bmatrix} = \begin{bmatrix} r_s & 0 \\ 0 & r_s \end{bmatrix} \begin{bmatrix} i_{as} \\ i_{bs} \end{bmatrix} + \frac{d}{dt} \begin{bmatrix} \lambda_{as} \\ \lambda_{bs} \end{bmatrix}$$

$$v_{abs} = r_s i_{abs} + p\lambda_{abs}$$

$$f_{abs} = \begin{bmatrix} f_{as} \\ f_{bs} \end{bmatrix} \qquad r_s = \begin{bmatrix} r_s & 0 \\ 0 & r_s \end{bmatrix}$$

– Transformation

$$\lambda_{as} = L_{asas} i_{as} + L_{asbs} i_{bs} + \lambda_{asm}$$

$$\lambda_{bs} = L_{bsas} i_{as} + L_{bsbs} i_{bs} + \lambda_{bsm}$$

$$\lambda_{abs} = L_s i_{abs} + \lambda'_m$$

$$\begin{bmatrix} \lambda_{as} \\ \lambda_{bs} \end{bmatrix} = \begin{bmatrix} L_{asas} & L_{asbs} \\ L_{bsas} & L_{bsbs} \end{bmatrix} \begin{bmatrix} i_{as} \\ i_{bs} \end{bmatrix} + \lambda'_m \begin{bmatrix} \sin\theta_r \\ -\cos\theta_r \end{bmatrix}$$

$$\lambda'_m = \begin{bmatrix} \lambda_{asm} \\ \lambda_{bsm} \end{bmatrix} = \lambda'_m \begin{bmatrix} \sin\theta_r \\ \cos\theta_r \end{bmatrix}$$

$$\begin{bmatrix} L_{asas} & L_{asbs} \\ L_{bsas} & L_{bsbs} \end{bmatrix} = \begin{bmatrix} L_{ss} & 0 \\ 0 & L_{ss} \end{bmatrix} = L_s$$

– <u>Transformation</u>

$$f_{qds}^{r} = K_s^r f_{abs}$$

$$\begin{bmatrix} f_{qs}^{r} \\ f_{ds}^{r} \end{bmatrix} = \begin{bmatrix} \cos\theta_r & \sin\theta_r \\ \sin\theta_r & -\cos\theta_r \end{bmatrix} \begin{bmatrix} f_{as} \\ f_{bs} \end{bmatrix} \qquad f_{abs} = \left[K_s^r\right]^{-1} f_{qds}^{r} = K_s^r f_{qds}^{r}$$

$$f_{abs} = \left[K_s^r\right]^{-1} f_{qds}^{r}$$

substitute ⇒

$$v_{abs} = r_s i_{abs} + p\lambda_{abs}$$

$$\left(K_s^r\right)^{-1} v_{qds}^{r} = r_s \left(K_s^r\right)^{-1} i_{qds}^{r} + p\left[\left(K_s^r\right)^{-1} \lambda_{qds}^{r}\right]$$

<u>Multiply by</u> K_s^r

$$\left(K_s^r\right)\left(K_s^r\right)^{-1} v_{qds}^{r} = \left(K_s^r\right) r_s \left(K_s^r\right)^{-1} i_{qds}^{r} + \left(K_s^r\right) p\left[\left(K_s^r\right)^{-1} \lambda_{qds}^{r}\right]$$

$$v_{qds}^{r} = r_s i_{qds}^{r} + \left(K_s^r\right) p\left[\left(K_s^r\right)^{-1} \lambda_{qds}^{r}\right]$$

– <u>Transformation</u>

$$v_{qds}^{r} = r_s i_{qds}^{r} + \left(K_s^r\right) p\left[\left(K_s^r\right)^{-1} \lambda_{qds}^{r}\right]$$

$$v_{qds}^{r} = r_s i_{qds}^{r} + \left(K_s^r\right)\left[p\left(K_s^r\right)^{-1}\right]\lambda_{qds}^{r} + \left(K_s^r\right)\left(K_s^r\right)^{-1} p\lambda_{qds}^{r}$$

$$= r_s i_{qds}^{r} + \omega_r \lambda_{dqs}^{r} + p\lambda_{qds}^{r}$$

$$\lambda_{dqs}^{r} = \begin{bmatrix} \lambda_{ds}^{r} \\ -\lambda_{qs}^{r} \end{bmatrix}$$

<u>For a magnetically linear system</u>: $\lambda_{abs} = L_s i_{abs} + \lambda_m^{'}$

$$\left(K_s^r\right)^{-1} \lambda_{qds}^{r} = L_s \left(K_s^r\right)^{-1} i_{qds}^{r} + \lambda_m^{'}$$

– Transformation

$$K_s^r \left(K_s^r\right)^{-1} \lambda_{qds}^r = K_s^r L_s \left(K_s^r\right)^{-1} i_{qds}^r + K_s^r \lambda_m'$$

$$\lambda_{qds}^r = L_s i_{qds}^r + K_s^r \lambda_m'$$

$$\lambda_{qds}^r = \begin{bmatrix} L_{\ell s} + L_{ms} & 0 \\ 0 & L_{\ell s} + L_{ms} \end{bmatrix} \begin{bmatrix} i_{qs}^r \\ i_{ds}^r \end{bmatrix} + \lambda_m^{'r} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

– In our new system of variables, the flux linkage created by the permanent magnet appears constant. Hence, our fictitious circuits are fixed relative to the permanent magnet and, therefore, fixed in the rotor. We have accomplished the goal of eliminating flux linkages which vary with θ_r.

– In expanded form, the voltage equations are:

$$v_{qs}^r = r_s i_{qs}^r + \omega_r \lambda_{ds}^r + p\lambda_{qs}^r$$

$$v_{ds}^r = r_s i_{ds}^r - \omega_r \lambda_{qs}^r + p\lambda_{ds}^r$$

– where

$$\lambda_{qs}^r = L_{ss} i_{qs}^r = (L_{\ell s} + L_{ms}) i_{qs}^r$$

$$\lambda_{ds}^r = L_{ss} i_{ds}^r + \lambda_m^{'r} = (L_{\ell s} + L_{ms}) i_{ds}^r + \lambda_m^{'r}$$

– Substitution

$$v_{qs}^r = (r_s + pL_{ss})i_{qs}^r + \omega_r L_{ss} i_{ds}^r + \omega_r \lambda_m^{'r}$$

$$v_{ds}^r = (r_s + pL_{ss})i_{ds}^r - \omega_r L_{ss} i_{qs}^r$$

$$\begin{bmatrix} v_{qs}^r \\ v_{ds}^r \end{bmatrix} = \begin{bmatrix} r_s + pL_{ss} & \omega_r L_{ss} \\ -\omega_r L_{ss} & r_s + pL_{ss} \end{bmatrix} \begin{bmatrix} i_{qs}^r \\ i_{ds}^r \end{bmatrix} + \begin{bmatrix} \omega_r \lambda_m^{'r} \\ 0 \end{bmatrix}$$

– The electromagnetic torque is obtained by expressing i_{as} and i_{bs} in terms of i^r_{qs} and i^r_{ds}. In particular:

$$T_e = \frac{P}{2} \lambda_m^{'r} i_{qs}^r$$

(positive for motor action)

- Time-Domain Block Diagrams and State Equations
  - Nonlinear System Equations
  - Voltage Equations

$$\begin{bmatrix} v_{qs}^r \\ v_{ds}^r \end{bmatrix} = \begin{bmatrix} r_s + pL_{ss} & \omega_r L_{ss} \\ -\omega_r L_{ss} & r_s + pL_{ss} \end{bmatrix} \begin{bmatrix} i_{qs}^r \\ i_{ds}^r \end{bmatrix} + \begin{bmatrix} \omega_r \lambda_m'^r \\ 0 \end{bmatrix}$$

  - Relationship between torque and rotor speed

$$T_e = J\left(\frac{2}{P}\right)\frac{d\omega_r}{dt} + B_m\left(\frac{2}{P}\right)\omega_r + T_L$$

– Rewriting the equations:

$$v_{qs}^{r} = r_s\left(1 + \tau_s p\right) i_{qs}^{r} + r_s \tau_s \omega_r i_{ds}^{r} + \lambda_m^{'r} \omega_r$$

$$v_{ds}^{r} = r_s\left(1 + \tau_s p\right) i_{ds}^{r} - r_s \tau_s \omega_r i_{qs}^{r}$$

$$\tau_s = \frac{L_{ss}}{r_s}$$

$$T_e - T_L = \frac{2}{P}\left(B_m + Jp\right)\omega_r$$

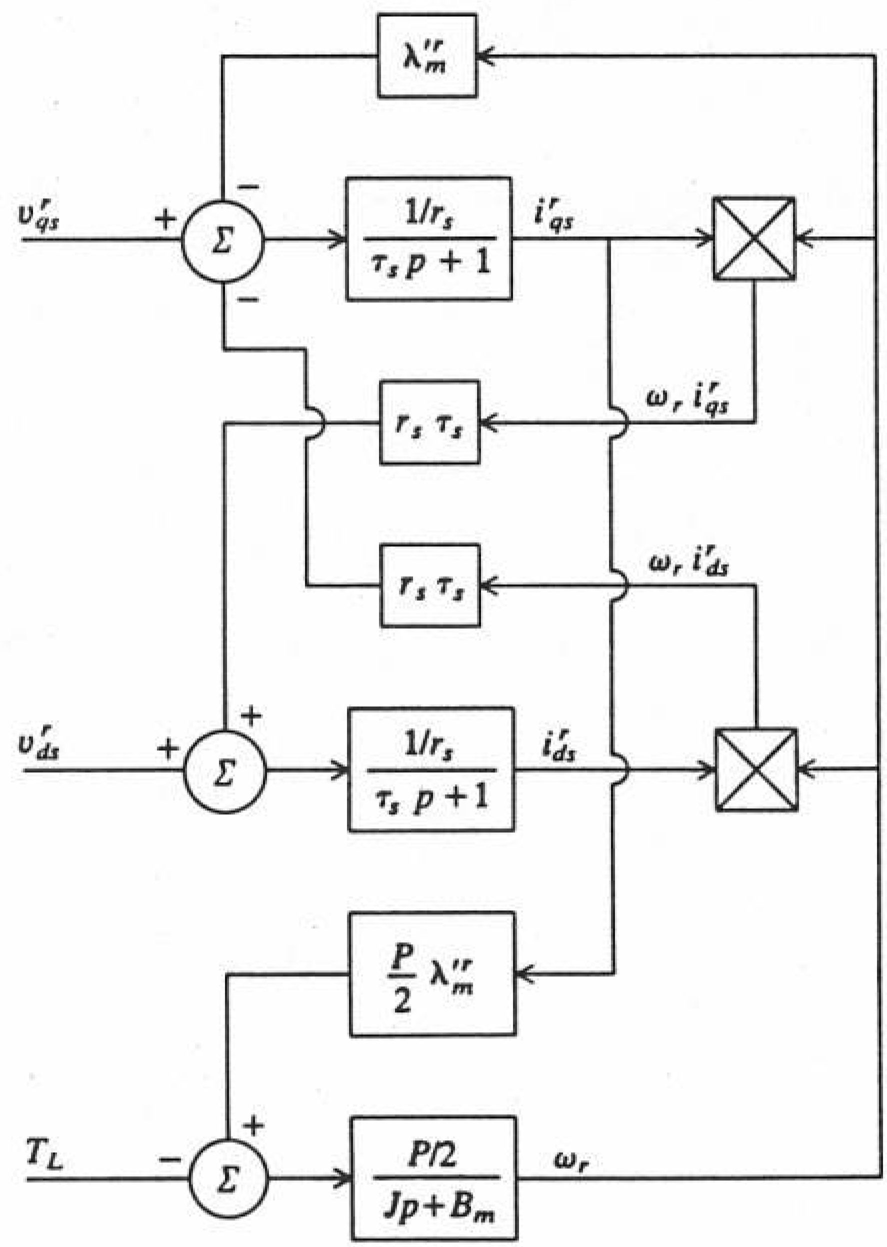
$$T_e = \frac{P}{2}\lambda_m^{'r} i_{qs}^{r}$$

$$i_{qs}^{r} = \frac{1/r_s}{\tau_s p + 1}\left(v_{qs}^{r} - r_s \tau_s \omega_r i_{ds}^{r} - \lambda_m^{'r}\omega_r\right)$$

$$i_{ds}^{r} = \frac{1/r_s}{\tau_s p + 1}\left(v_{ds}^{r} + r_s \tau_s \omega_r i_{qs}^{r}\right)$$

$$\omega_r = \frac{P/2}{Jp + B_m}\left(T_e - T_L\right)$$

Time-Domain Block Diagram of a Brushless DC Machine

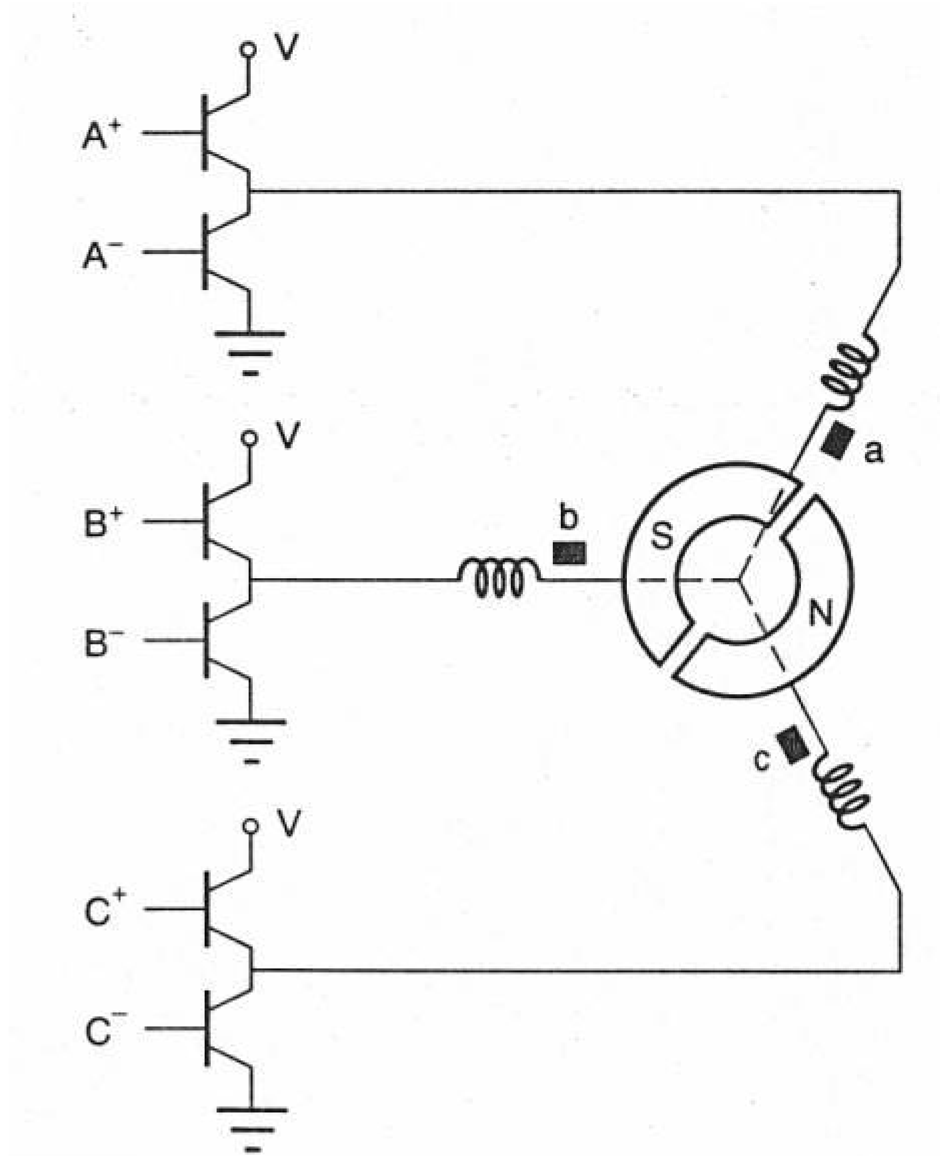


– State Variables

– The state variables are the stator currents i^r_{qs} and i^r_{ds}, the rotor speed ω_r, and the rotor position θ_r. Here we will omit θ_r since it is considered a state variable only when shaft position is a controlled variable. Also it can be established from ω_r.

– In Matrix Form the state-variable equations are:

$$\frac{d}{dt}\begin{bmatrix} i^r_{qs} \\ i^r_{ds} \\ \omega_r \end{bmatrix} = \begin{bmatrix} -\frac{r_s}{L_{ss}} & 0 & -\frac{\lambda'^r_m}{L_{ss}} \\ 0 & -\frac{r_s}{L_{ss}} & 0 \\ \left(\frac{P}{2}\right)^2 \frac{\lambda'^r_m}{J} & 0 & -\frac{B_m}{J} \end{bmatrix} \begin{bmatrix} i^r_{qs} \\ i^r_{ds} \\ \omega_r \end{bmatrix} + \begin{bmatrix} -\omega_r i^r_{ds} \\ \omega_r i^r_{qs} \\ 0 \end{bmatrix} + \begin{bmatrix} \frac{1}{L_{ss}} & 0 & 0 \\ 0 & \frac{1}{L_{ss}} & 0 \\ 0 & 0 & -\frac{P}{2}\frac{1}{J} \end{bmatrix} \begin{bmatrix} v^r_{qs} \\ v^r_{ds} \\ T_L \end{bmatrix}$$


V
A+
A−
V
B+
B−
V
C+
C−
a
b
c
S
N

Stepping Motors

- Stepping motors are electromechanical motion devices used primarily to convert information in digital form to mechanical motion.
- Introduction and Motivation
  - Stepping Motor Applications
  - Inkjet Printer Application and Testbed
- Stepper Motor Basics

Main References

- Stepping Motors: A Guide to Modern Theory and Practice, 3rd Edition
  - P. P. Acarnley, IEE, 1992
- Stepping Motors and Their Microprocessor Controls, 2nd Edition
  - T. Kenjo and A. Sugawara, Oxford, 1995
- Step Motor System Design Handbook, 2nd Edition
  - A. Leenhouts, Litchfield Engineering Co., 1997

Stepping Motor Applications

- Computer Peripherals
  - Main area for stepping motor applications
    - Printers
      - Carriage transport
      - Paper feed
      - Rotation of the photosensitive drum
      - Toner stirring unit
    - Graph (X-Y) plotters
      - Paper sheet is driven in the forward or backward direction with a stepping motor
      - Pen is driven by another stepping motor in the horizontal direction

- Stepping motor is used to drive the mechanism for replacing ink holders
- Disk drives
  - Head positioning in disk drives
- Numerical Control
  - XY-tables and index tables
  - Milling machines
  - Sewing machines
- Office Machines
  - Copiers
  - Facsimile machines

- Applications in Semiconductor Technology
  - Stepping motors used in high vacuums
  - Goniometer – instrument used to determine crystalline structure
  - Electron-beam microfabricator
- Space Vehicles and Satellites
- Other Applications
  - Time pieces
  - Cameras
  - Heavy industry applications
  - Medical equipment

- Whenever stepping from one position to another is required, whether the application is industrial, military, or medical, the stepper motor is generally used.

Inkjet Printer Application & Testbed

- Printer Description
  - Problem Statement
    - Potential Causes
    - Potential Countermeasures
    - Goal
    - Typical Nominal Motion Trajectory

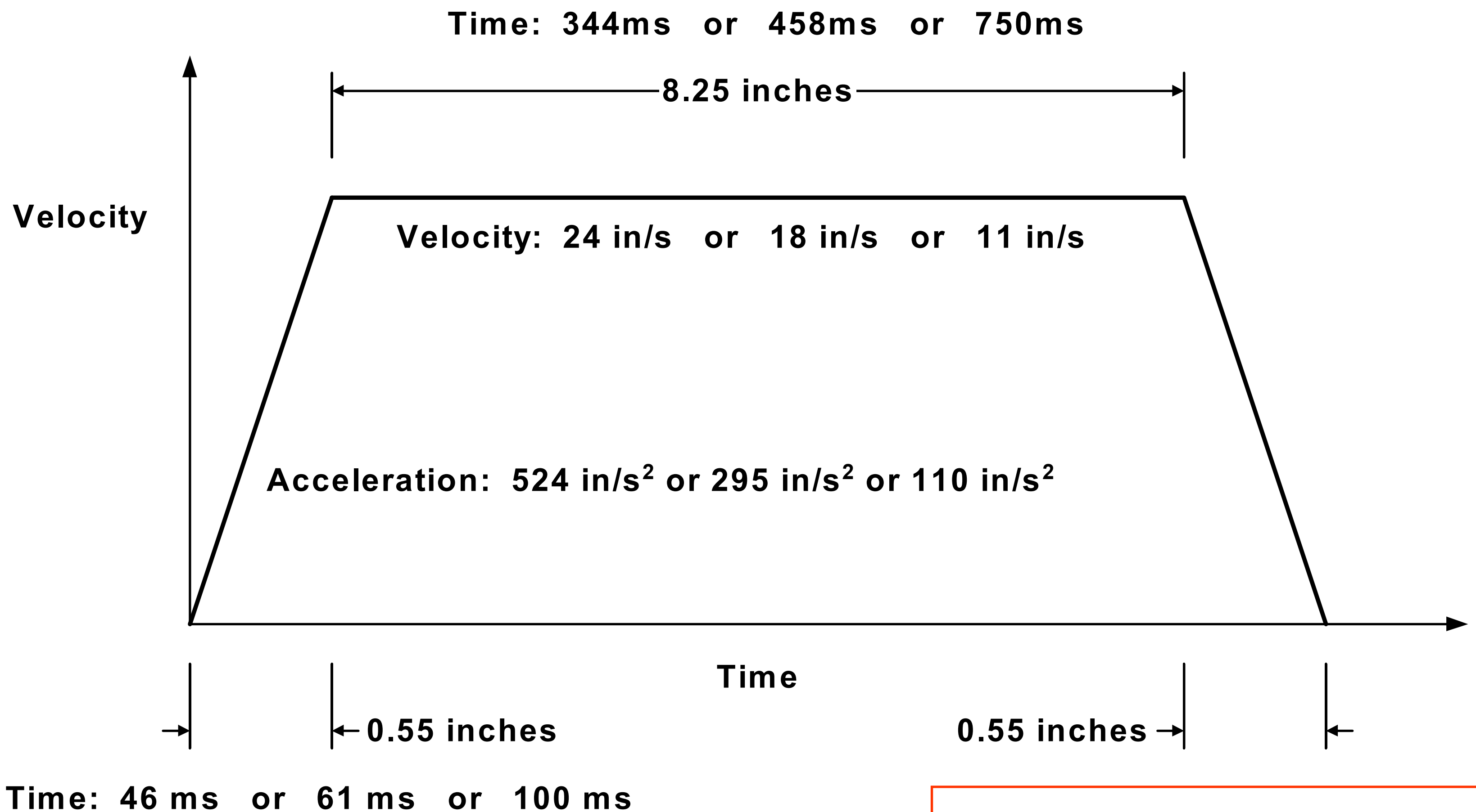
Printer Description

Problem Statement

- Current inkjet printer scan system exhibits undesirable noise and motion quality variations at certain velocities or scan positions.
- Potential Causes:
  - Step tables
  - Carriage vibration
  - Carriage-to-rail interface

- Potential Countermeasures:
  - Optimize scan motor step tables
  - Optimize for cost and performance the system stiffness
  - Optimize rail-to-carriage interface
- Goal
  - Develop an analytical and empirical understanding of the relationship between input parameters and output responses

Step Tables System Stiffness System Damping	⟷	Position and Velocity Noise Level Print Quality



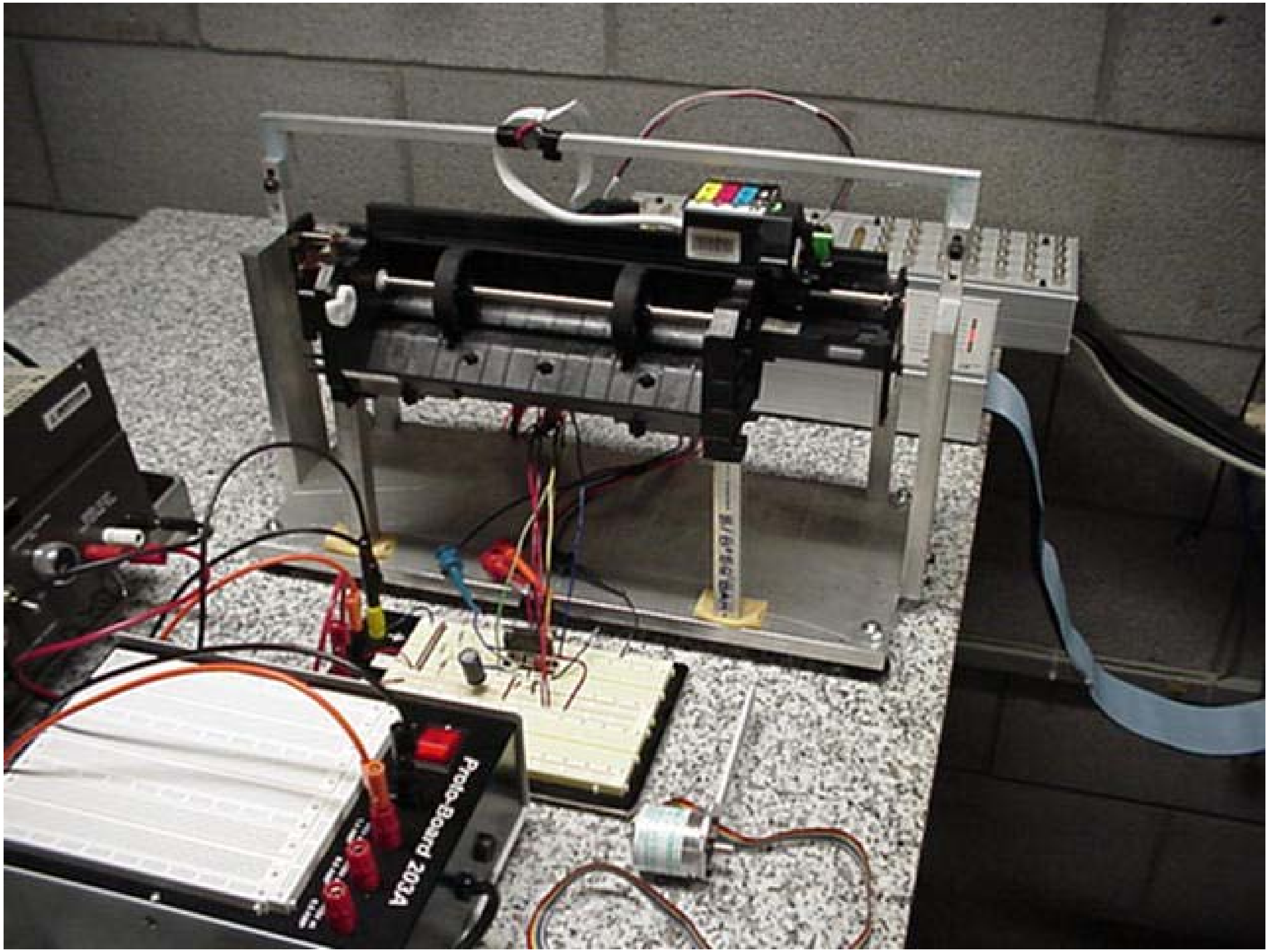
Typical Nominal Motion Trajectory: Velocity vs. Time

0.55 in = 33 half steps
8.25 in = 495 half steps
96 half steps / rev
0.25 in/rad

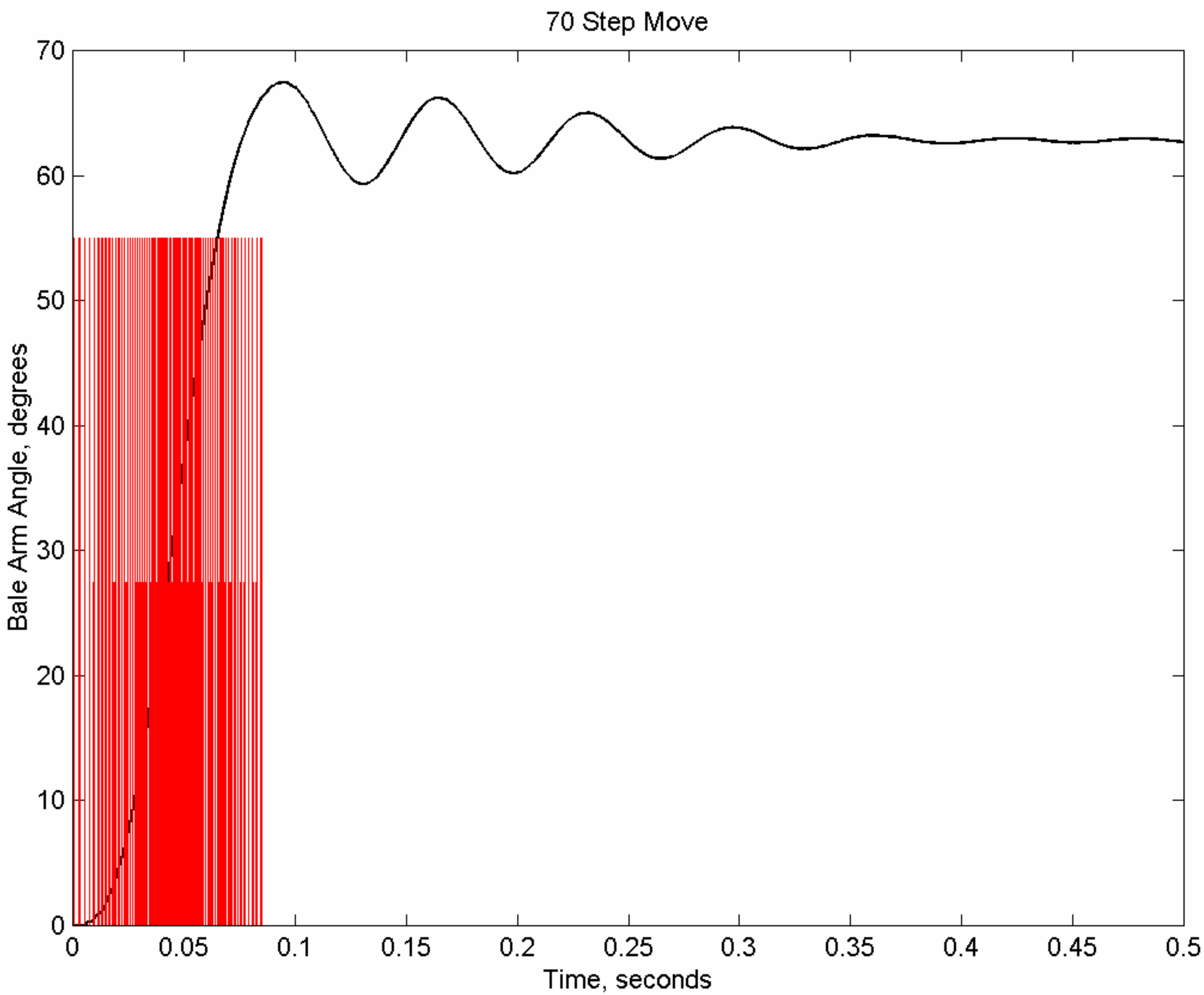
Inkjet Printer Testbed

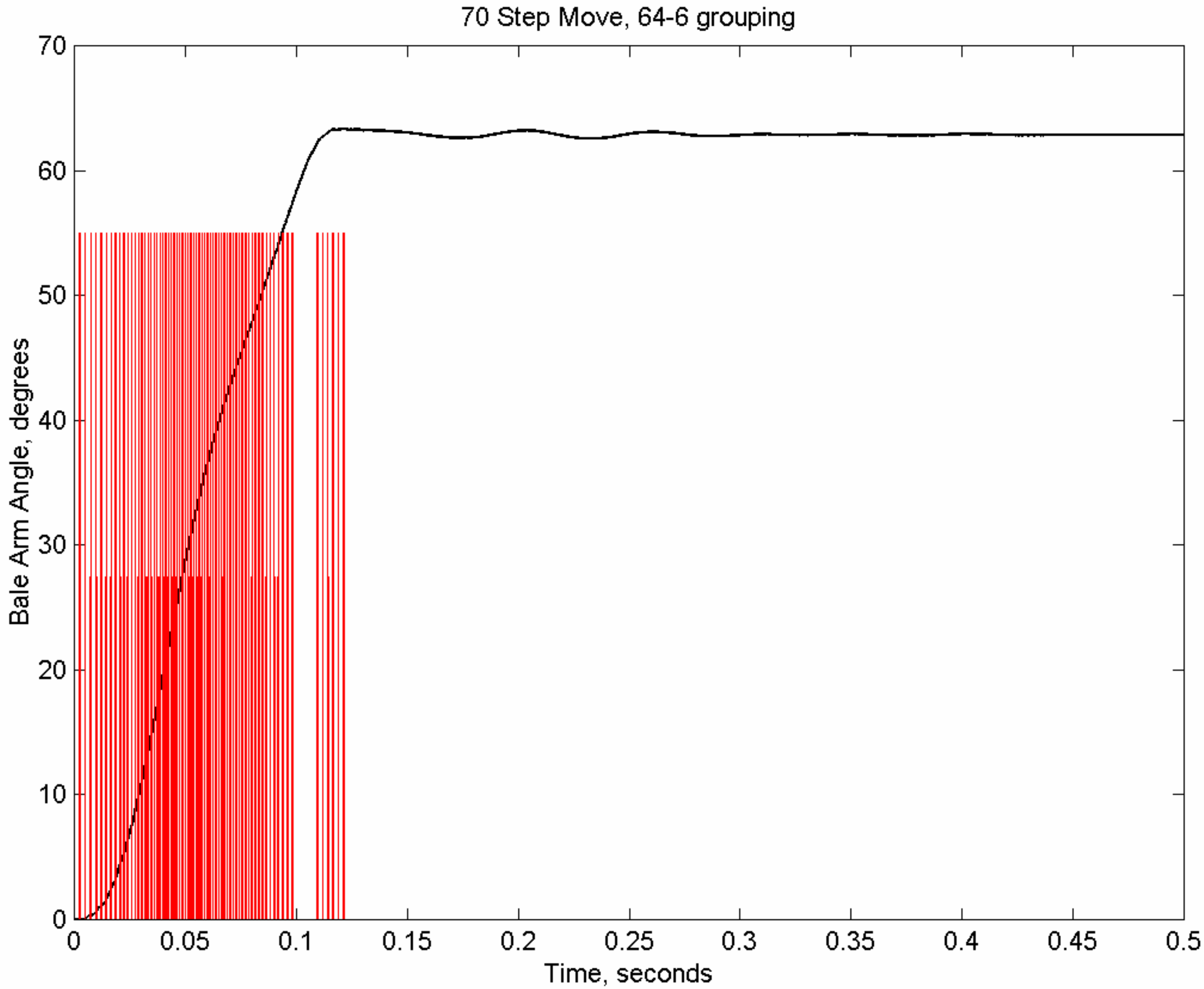
- Physical System Description
  - Printer-Carriage, Belt-Drive System
  - Motors: Bipolar and Unipolar
  - Encoders: Linear and Rotary
  - Driver Chips: Bipolar PWM and Unipolar PWM
  - dSpace / MatLab / Simulink Implementation
- System Capabilities
  - Inkjet Printer Applications
  - General Stepper Motor System Design Studies

Inkjet Printer Testbed



Xerox Application


70 Step Move
Bale Arm Angle, degrees
Time, seconds
0
10
20
30
40
50
60
70
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5


70 Step Move, 64-6 grouping
Bale Arm Angle, degrees
Time, seconds
0
10
20
30
40
50
60
70
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5

Stepper Motor Basics

- Introduction
- DC Motors vs. Stepper Motors
- Advantages of Stepper Motors
- Disadvantages of Stepper Motors
- Stepper Motor Basics
  - Variable Reluctance (VR) Stepper Motor
  - Hybrid Stepper Motor
- Bifilar vs. Unifilar Windings

Introduction

- Proper selection of actuators is of *utmost importance* in the design of mechatronic systems.
- Factors such as power, motion resolution, repeatability, and operating bandwidth requirements for an actuator can differ significantly, depending on the particular mechatronic system and the specific function of the actuator within the system.
- AC motors, DC motors and hydraulic/pneumatic actuators are *continuous-drive actuators*.
- A stepper motor is an electromagnetic actuator, like a DC motor, however, it is an *incremental-drive (digital) actuator* and is driven in fixed angular steps.

- Each step of rotation is the response of the motor to an input pulse (or digital command).
- Step-wise rotation of the rotor can be synchronized with pulses in a command-pulse train, assuming that no steps are missed, thereby making the motor respond faithfully to the pulse signal in an open-loop manner.
- Stepper motors have emerged as cost-effective alternatives for DC servomotors in high-speed, motion-control applications (except the high torque-speed range) with the improvements in permanent magnets and the incorporation of solid-state circuitry and logic devices in their drive systems.
- Today stepper motors can be found in computer peripherals, machine tools, medical equipment, automotive devices, and small business machines, to name a few applications.

DC Motors vs. Stepper Motors

- Stepper motors are operated open loop, while most DC motors are operated closed loop.
- Stepper motors are easily controlled with microprocessors, however logic and drive electronics are more complex.
- Stepper motors are brushless and brushes contribute several problems, e.g., wear, sparks, electrical transients.
- DC motors have a continuous displacement and can be accurately positioned, whereas stepper motor motion is incremental and its resolution is limited to the step size.
- Stepper motors can slip if overloaded and the error can go undetected. (A few stepper motors use closed-loop control.)
- Feedback control with DC motors gives a much faster response time compared to stepper motors.

Advantages of Stepper Motors

- Position error is noncumulative. A high accuracy of motion is possible, even under open-loop control.
- Large savings in sensor (measurement system) and controller costs are possible when the open-loop mode is used.
- Because of the incremental nature of command and motion, step motors are easily adaptable to digital control applications.
- No serious stability problems exist, even under open-loop control.
- Torque capacity and power requirements can be optimized and the response can be controlled by electronic switching.
- Brushless construction has obvious advantages.

Disadvantages of Stepper Motors

- They have low torque capacity (typically less than 2,000 oz-in) compared to DC motors.
- They have limited speed (limited by torque capacity and by pulse-missing problems due to faulty switching systems and drive circuits).
- They have high vibration levels due to stepwise motion.
- Large errors and oscillations can result when a pulse is missed under open-loop control.

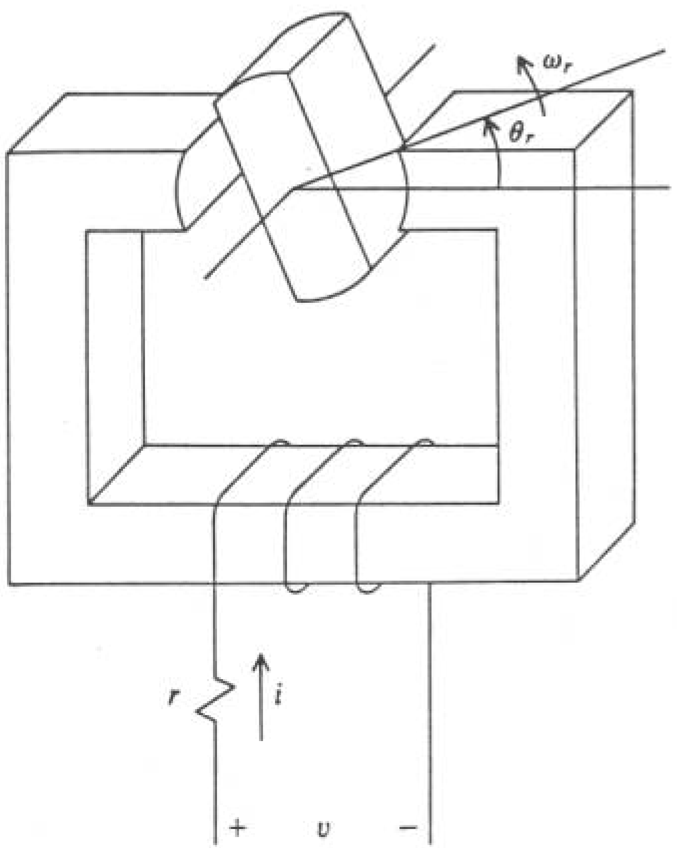
Stepper Motor Basics

- To produce a significant torque from a reasonable volume, both the stationary and rotating components must have large numbers of iron teeth, which must be able to carry a substantial magnetic flux.
- Performance of the stepper motor depends on the strength of the magnetic field. High flux leads to high torque.
- Only two basic types need to be considered:
  - Variable-Reluctance
  - Hybrid

- Essential Property
  - Ability to translate switched excitation changes into precisely defined increments of rotor position (steps).
  - Accurate positioning of the rotor is generally achieved by magnetic alignment of the iron teeth of the stationary and rotating parts of the motor.
- Hybrid Motor
  - Main source of magnetic flux is a permanent magnet; dc currents flowing in one or more windings direct the flux along alternative paths.
- Variable Reluctance (VR)
  - There are two configurations; in both cases the magnetic field is produced solely by the winding currents.

Review: Single-Phase Reluctance Machine

- The machine consists of:
  - stationary core with a winding of N turns
  - moveable member which rotates



$\theta_r =$ angular displacement

$\omega_r =$ angular velocity

$$\theta_r = \int_0^t \omega_r(\xi)\,d\xi + \theta_r(0)$$

$$v = ri + \frac{d\lambda}{dt}$$

voltage equation

$$\phi = \phi_\ell + \phi_m$$

$\phi_\ell =$ leakage flux

$\phi_m =$ magnetizing flux

$$\lambda = \left(L_\ell + L_m\right)i$$

It is convenient to express the flux linkages as the product of the sum of the leakage inductance and the magnetizing inductance and the current in the winding.

$L_\ell =$ constant (independent of θ_r)

$L_m =$ periodic function of θ_r

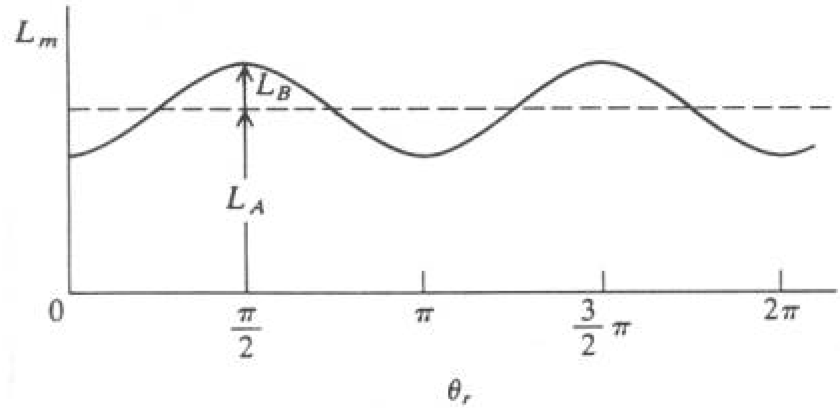
$$L_m = L_m(\theta_r)$$

$$L_m(0) = \frac{N^2}{\Re_m(0)} \Rightarrow \begin{cases} \Re_m \text{ is maximum} \\ L_m \text{ is minimum} \end{cases}$$

$$L_m\left(\frac{\pi}{2}\right) = \frac{N^2}{\Re_m\left(\frac{\pi}{2}\right)} \Rightarrow \begin{cases} \Re_m \text{ is minimum} \\ L_m \text{ is maximum} \end{cases}$$

The magnetizing inductance varies between maximum and minimum positive values twice per revolution of the rotating member.

Assume that this variation may be adequately approximated by a sinusoidal function.



$$L_m(\theta_r) = L_A - L_B \cos(2\theta_r)$$

$$L_m(0) = L_A - L_B$$

$$L_m\left(\frac{\pi}{2}\right) = L_A + L_B$$

$$L_A > L_B$$

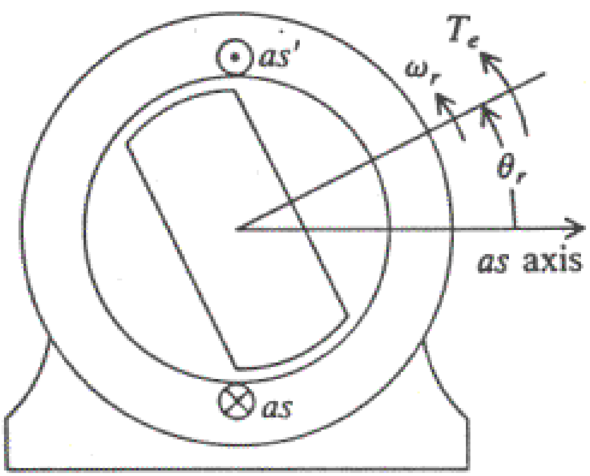
$$L_A = \text{ average value}$$

$$\begin{aligned} L(\theta_r) &= L_\ell + L_m(\theta_r) \\ &= L_\ell + L_A - L_B \cos(2\theta_r) \end{aligned}$$

$$v = ri + \left[L_\ell + L_m(\theta_r)\right]\frac{di}{dt} + i\frac{dL_m(\theta_r)}{d\theta_r}\frac{d\theta_r}{dt}$$

voltage equation

- This elementary two-pole single-phase reluctance machine is shown in a slightly different form. Winding 1 is now winding *as* and the stator has been changed to depict more accurately the configuration common for this device.



$$v_{as} = r_s i_{as} + \frac{d\lambda_{as}}{dt}$$

$$\lambda_{as} = L_{asas} i_{as}$$

$$L_{asas} = L_{\ell s} + L_A - L_B \cos(2\theta_r)$$

$$\theta_r = \int_0^t \omega_r(\xi) d\xi + \theta_r(0)$$

r_s = resistance of *as* winding

L_{asas} = self-inductance of *as* winding

$L_{\ell s}$ = leakage inductance

- Electromagnetic torque:
  - Magnetic system is linear, hence $W_f = W_c$.

$$W_c\left(i_{as},\theta_r\right) = \frac{1}{2}\left(L_{\ell s} + L_A - L_B\cos\left(2\theta_r\right)\right)i_{as}^2$$

$$T_e\left(\vec{i},\theta\right) = \sum_{j=1}^{J}\left[i_j\frac{\partial\lambda_j\left(\vec{i},\theta\right)}{\partial\theta}\right] - \frac{\partial W_f\left(\vec{i},\theta\right)}{\partial\theta}$$

$$T_e\left(\vec{i},\theta\right) = \frac{\partial W_c\left(\vec{i},\theta\right)}{\partial\theta}$$

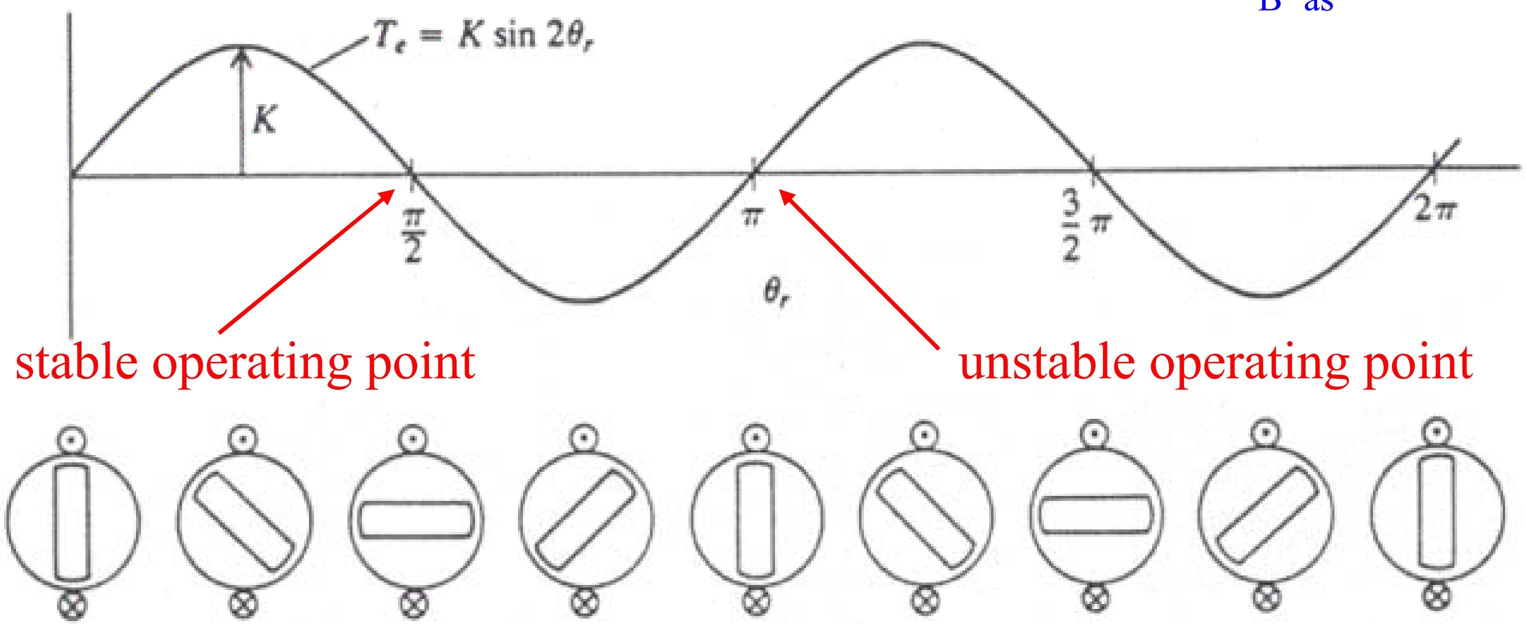
$$T_e\left(i_{as},\theta_r\right) = L_B i_{as}^2\sin\left(2\theta_r\right)$$

Valid for both transient and steady-state operation

- Consider steady-state operation: i_{as} is constant

$$T_e = K\sin(2\theta_r)$$

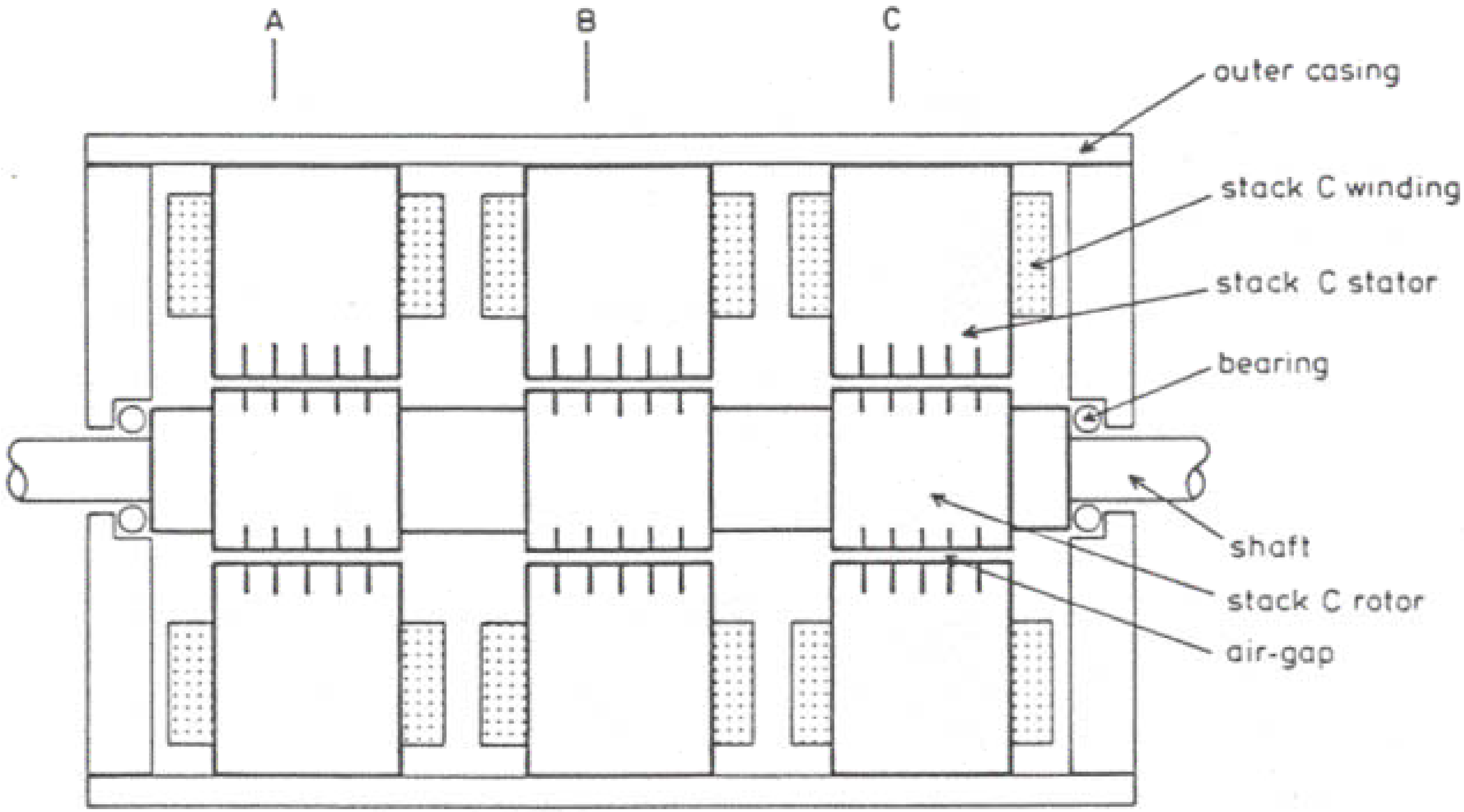
$$K = L_B i_{as}^2$$



Electromagnetic torque versus angular displacement of a single-phase reluctance machine with constant stator current

- Although the operation of a single-phase reluctance machine with a constant current is impracticable, it provides a basic understanding of reluctance torque, which is the operating principle of variable-reluctance stepper motors.
- In its simplest form, a variable-reluctance stepper motor consists of three cascaded, single-phase reluctance motors with rotors on a common shaft and arranged so that their minimum reluctance paths are displaced from each other.

- Multi-Stack Variable Reluctance Stepping Motor

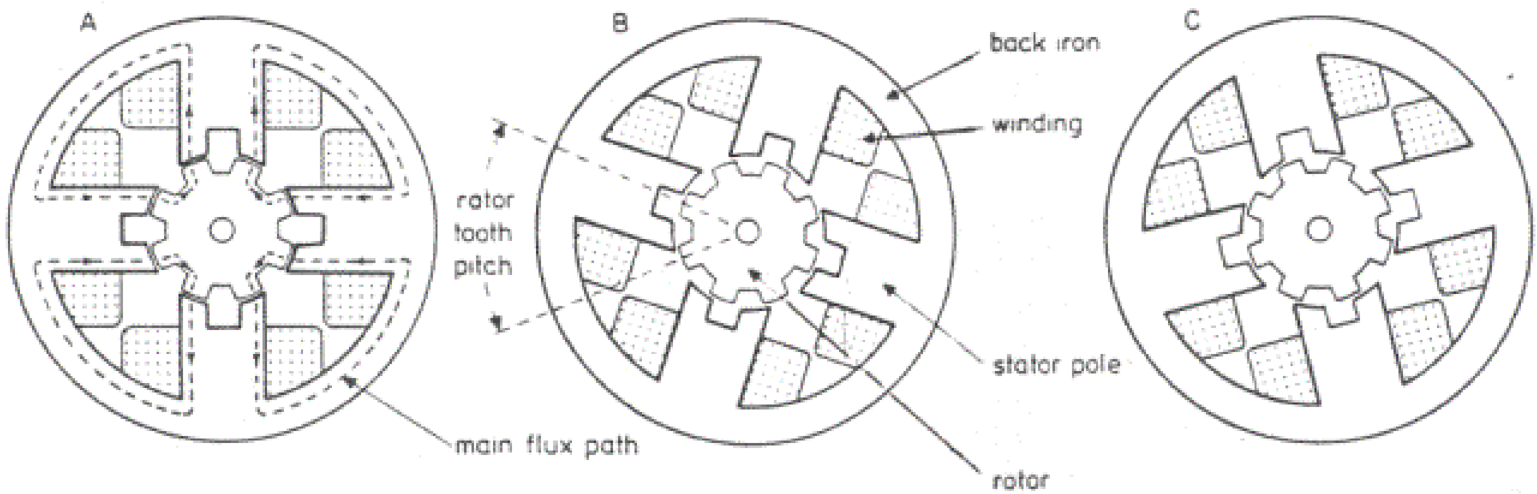


Cross–section of a three-stack variable-reluctance stepping motor parallel to the shaft

Step Length

$$\frac{360^\circ}{Np} = \frac{360^\circ}{(3)(8)} = 15^\circ$$

4 poles, 8 stator / rotor teeth, 3 stacks, 3 phases



Cross–sections of a three-stack, variable-reluctance stepping motor perpendicular to the shaft

- Magnetically isolated sections (stacks), each of which has a stationary stator and a one-piece rotor, both made of laminated iron.
- Each stator has a number of wound poles, with adjacent poles wound in the opposite sense.
- Magnetic circuit for each pair of adjacent poles is from one stator pole, across the air-gap into the rotor, through the rotor, across the air-gap into an adjacent pole, through this pole, returning to the original pole via the back-iron.
- Rotor and stator have equal numbers of teeth.

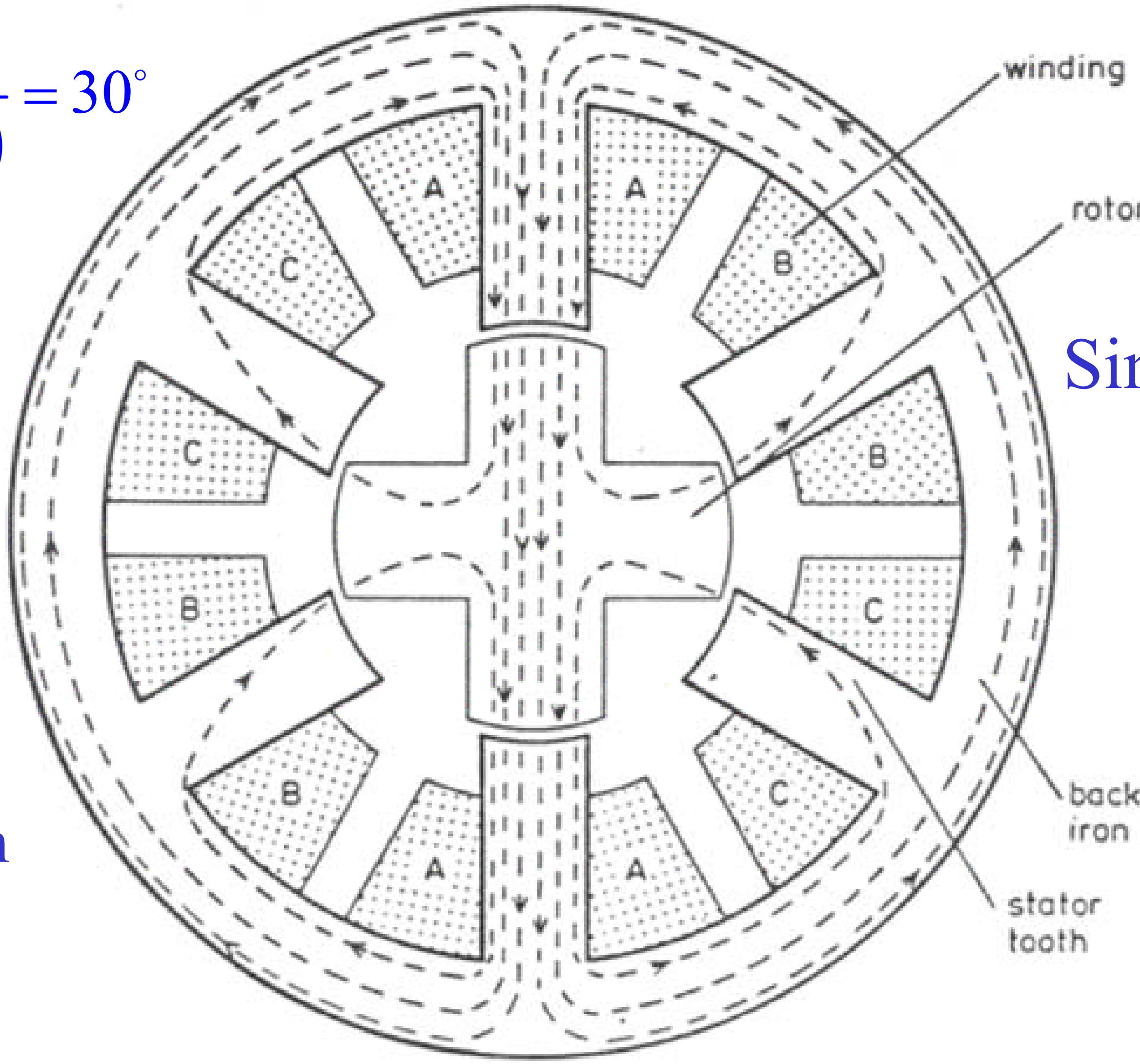
- When stator and rotor teeth are fully aligned, the circuit reluctance is minimized and the magnetic flux is at its maximum value.
- Rotor teeth in each stack are aligned; stator teeth have different relative orientations between stacks.
- One tooth pitch is $360°/p$ where p is the number of rotor teeth. If N is the number of stacks (and phases), then the step length = $360°/Np$. Typical step lengths are 2 to 15 degrees.
- Motors with higher stack numbers have no real performance advantages over a 3-stack motor.

- Continuous CW rotation can be produced by repeating the sequence: A, B, C, A, B, C, A, ….
- Continuous CCW rotation can be produced by repeating the sequence: A, C, B, A, C, B, A, ….
- If bi-directional operation is required from a multi-stack motor, it must have at least three stacks so that two distinct excitation sequences are available.

- Design Limitations
  - Pole flux density (magnetic saturation)
  - Winding temperature rise
  - The stator / rotor should reach magnetic saturation at the rated winding current
- Winding Interconnections Vary
  - Low-voltage, high-current drive with parallel winding connection
  - High-voltage, low-current drive with series winding connection
  - In either case, there is no difference in power supplied to the phase

$$\frac{360^\circ}{Np} = \frac{360^\circ}{(3)(4)} = 30^\circ$$

Step Length



Single-Stack VR Stepper

6 stator teeth
4 rotor teeth
3 phases

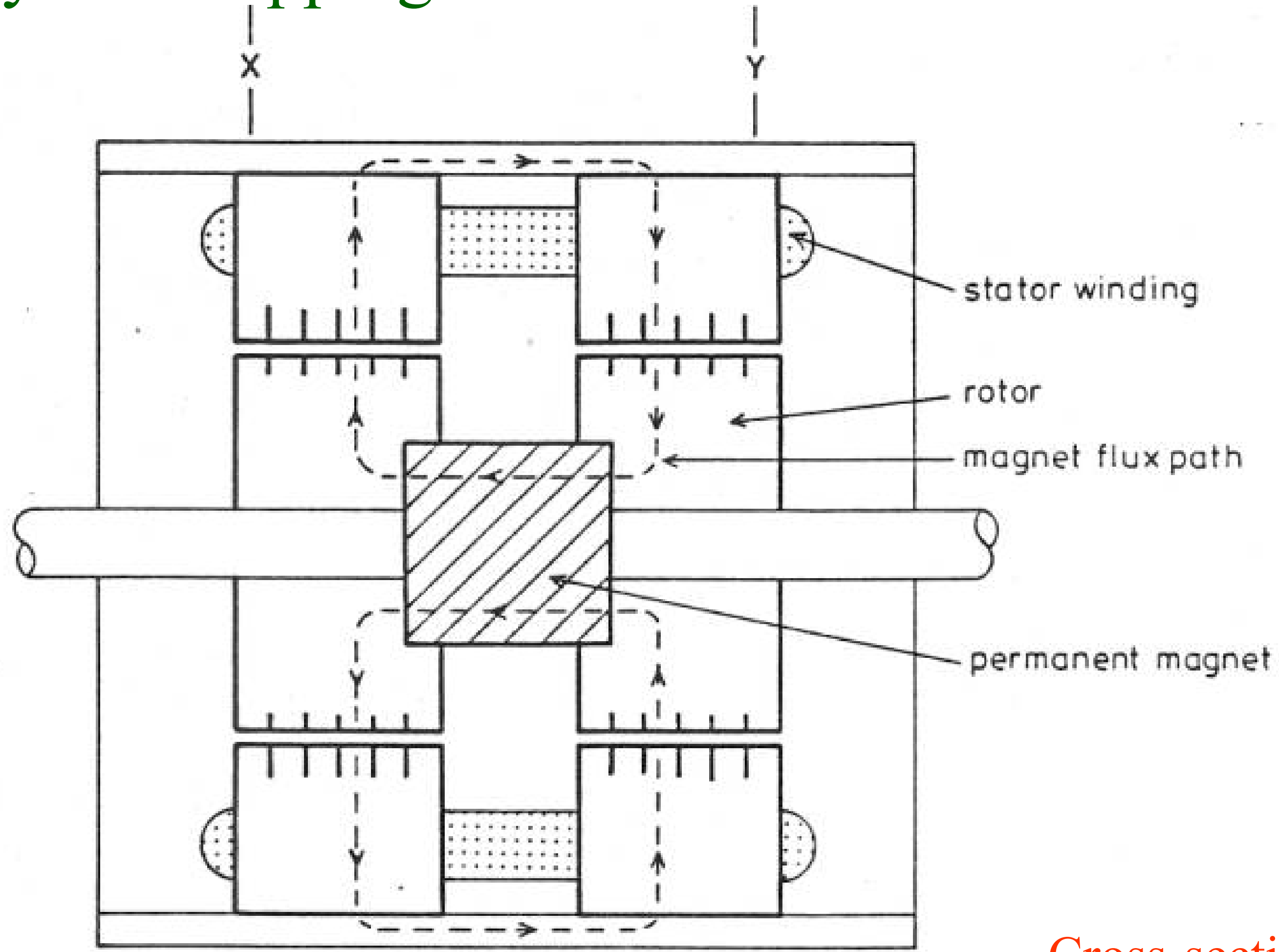
Cross–section of a single-stack variable-reluctance stepping motor perpendicular to the shaft

- There are essential differences between the single- and multi-stack types:
  - Each tooth has a separate winding.
  - Windings on opposite teeth are connected together to form one phase and are in opposing senses.
  - With one phase excited, the main flux path lies from one stator tooth, across the air-gap into a rotor tooth, directly across the rotor to another rotor-tooth / air-gap / stator-tooth combination and returns via the back-iron.
  - Secondary flux paths produce mutual coupling between the phase windings.

- Rotor and stator have different numbers of teeth.
- With one phase excited, only two of the rotor teeth carry the main flux. The rotor moves to a position that minimizes the main flux path reluctance.
- It is interesting to note that the rotor movement is in the opposite direction to the stepped rotation of the stator magnetic field, e.g., for continuous CCW rotation the excitation sequence is: A, B, C, A, B, C, A, …; similarly, CW rotation can be produced using the excitation sequence: A, C, B, A, C, B, A, ….

- If N is the number of phases and p is the number of rotor teeth, then the step length = $360°/Np$.
- The number of stator teeth has to be an even multiple of the number of phases.
- For satisfactory stepping action, the number of stator teeth must be near (but not equal) to the number of rotor teeth.

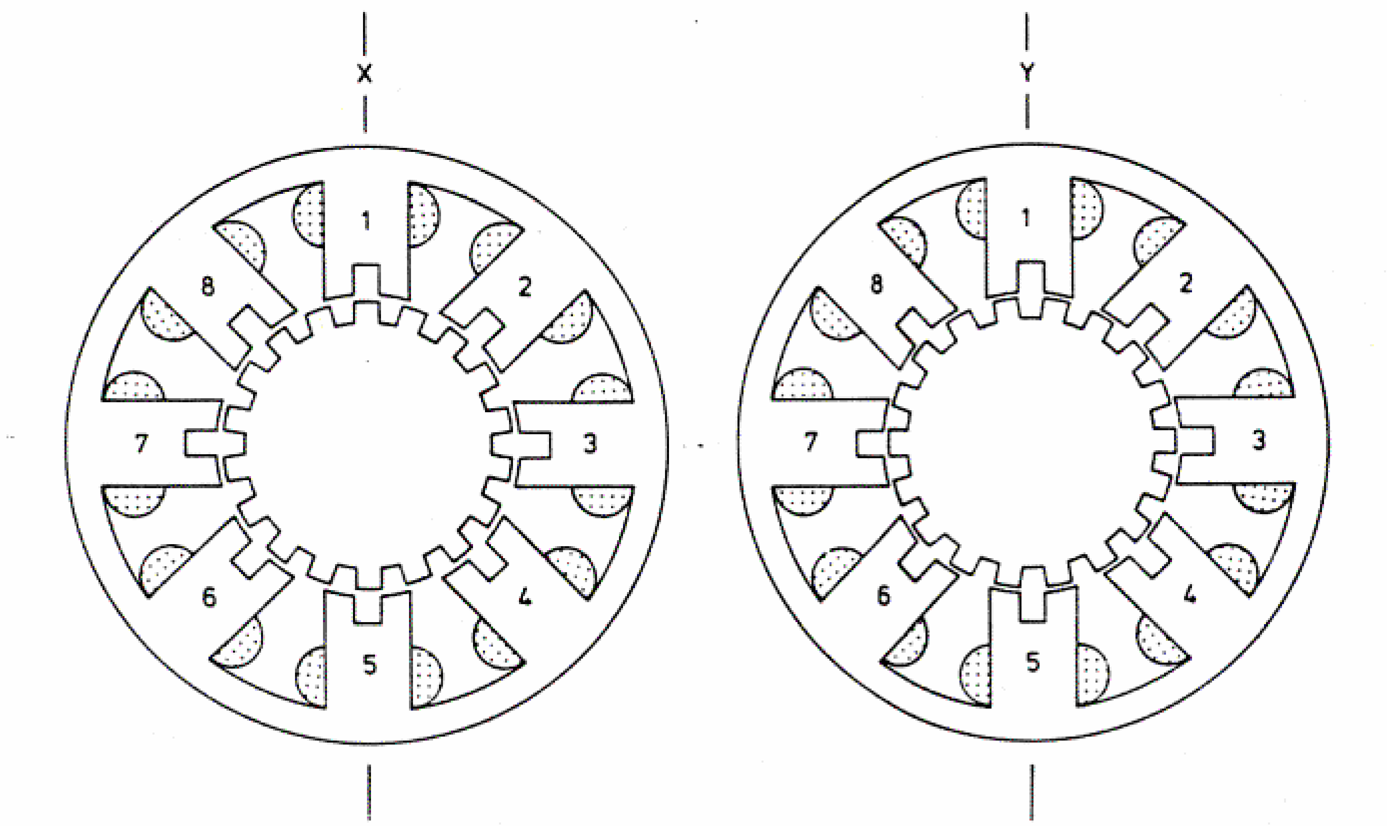
- # Hybrid Stepping Motors



Cross-section parallel to the shaft

Step Length

$$\frac{90^\circ}{p} = \frac{90^\circ}{18} = 5^\circ$$



Cross-section of hybrid motor perpendicular to the shaft

- Permanent magnet is mounted on the rotor.
- Main flux path lies from the magnet N-pole, into a soft-iron end-cap, radially through the end-cap, across the air-gap, through the stator poles of section X, axially along the stator back-iron, through the stator poles of section Y, across the air-gap and back to the magnet S-pole via the end-cap.
- There are typically 8 stator poles and each pole has between 2 and 6 teeth. There are two windings (phases) and each winding is situated on 4 of the 8 stator poles.

– Winding A is placed on poles 1, 3, 5, 7, and winding B is placed on poles 2, 4, 6, 8.
– Successive poles of each phase are wound in the opposite sense.

Winding	Current Direction	Radially Outward	Radially Inward
A	Positive	3, 7	1, 5
A	Negative	1, 5	3, 7
B	Positive	4, 8	2, 6
B	Negative	2, 6	4, 8

- Both stator poles and rotor end-caps are toothed.
- For the motor shown, there are 16 stator teeth and 18 rotor teeth. The stator teeth in sections X and Y are fully aligned; the rotor teeth are completely misaligned between the two sections.
- The rotor tends to align itself so that the air-gap reluctance of the flux path is minimized.
- Continuous CW rotation is produced by sequential excitation of the phase windings: A+, B+, A-, B-, A+, B+, …. CCW rotation would result from the excitation sequence: A+, B-, A-, B+, A+, B-, ….

- A complete cycle of excitation consists of 4 states (4 steps) and corresponds to a rotor movement of one tooth pitch (360°/p where p is the number of rotor teeth).
- The step length is therefore 90°/p.

- <u>Comparison of Motor Types</u>
- Hybrid Motors
  - Small step length (typically 1.8°).
  - Greater torque-producing capability for a given motor volume.
  - Natural choice for applications requiring a small step length and high torque in a restricted working space.
  - Detent torque retains rotor at step position when windings are unexcited.

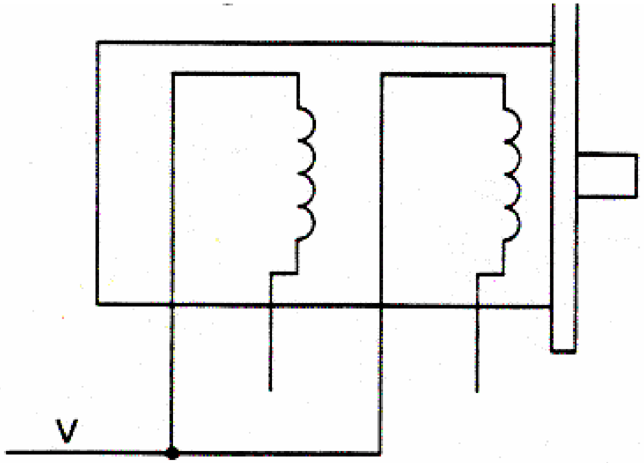
- Variable-Reluctance Motors
  - Typical step lengths (15°) are longer than in the hybrid, so less steps are required to move a given distance.
  - Fewer steps implies less excitation changes and it is the speed with which excitation changes can take place which ultimately limits the time taken to move the required distance.
  - Another advantage is the lower rotor mechanical inertia because of the absence of the permanent magnet.

Bifilar vs. Unifilar Windings

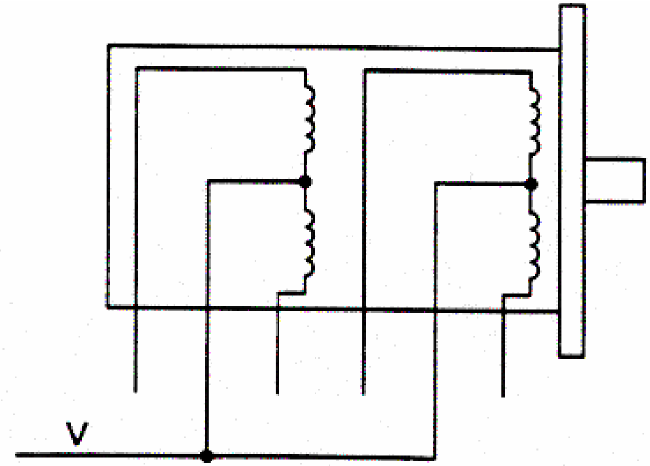
- A common feature in any stepper motor is that the stator of the motor contains several pairs of field windings that can be switched on to produce the electromagnetic pole pairs (N & S).
- The polarities can be reversed in two ways:
  - By reversing the direction of current in the winding (*unifilar windings*).
  - By using two pairs of windings (*bifilar windings*) for each pole pair, one pair giving one set of poles when energized and the other pair giving the opposite polarities.

- Simple on/off switching is adequate for bifilar windings, while current reversal circuitry is needed for unifilar windings.
- Twice the normal number of windings are needed for bifilar windings which increases the motor size for a given torque rating. Decreasing wire diameter helps; this also increases resistance, which increases damping and decreases electrical τ, giving fast but less oscillatory single-step response.

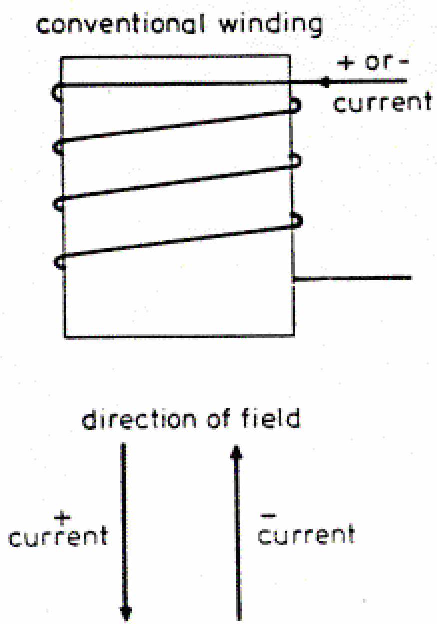
- Because current reversals are absent in bifilar windings, there are smaller levels of induced voltages by self induction and mutual induction. For this reason, the dynamic torque at a given stepping rate is usually larger for bifilar steppers, particularly at high speeds.
- At low speeds, dissipation effects will dominate induced-voltage effects, thereby giving higher torques at low speeds with unifilar windings.

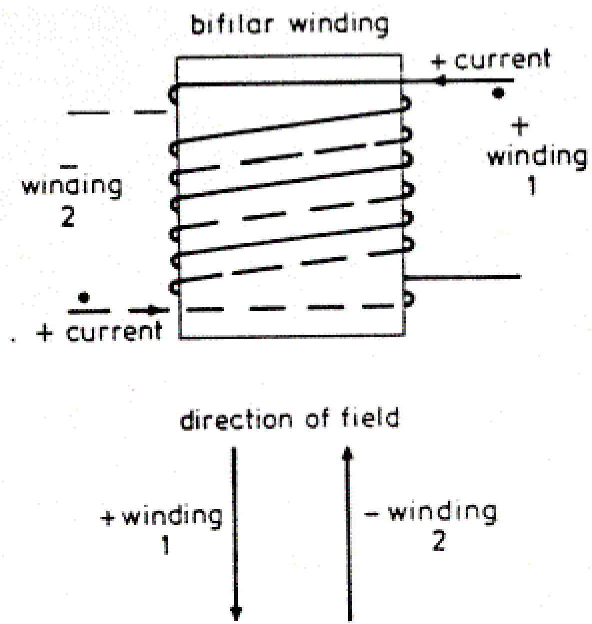


4-Wire Bipolar (unifilar)



6-Wire Unipolar (bifilar)





Two-Phase Step Motors

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?

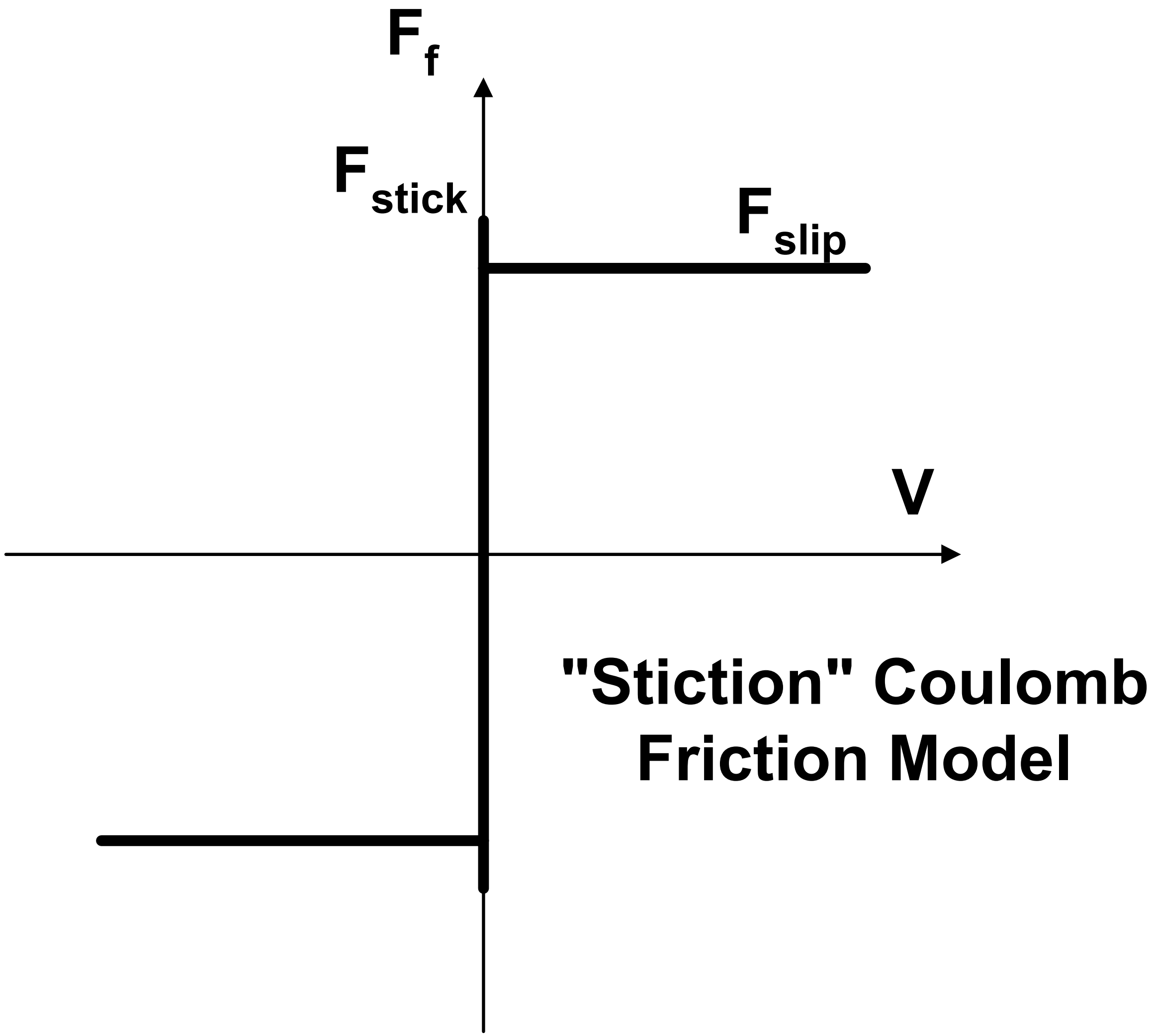
- General remedies include:
  - Alter the design to reduce the effective loop gain of the controller, especially at high frequencies where the effects of parasitics are often predominant. This generally entails sacrifice in performance.
  - Techniques specifically intended to enhance robustness of the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

Coulomb Friction: Modeling and Simulation

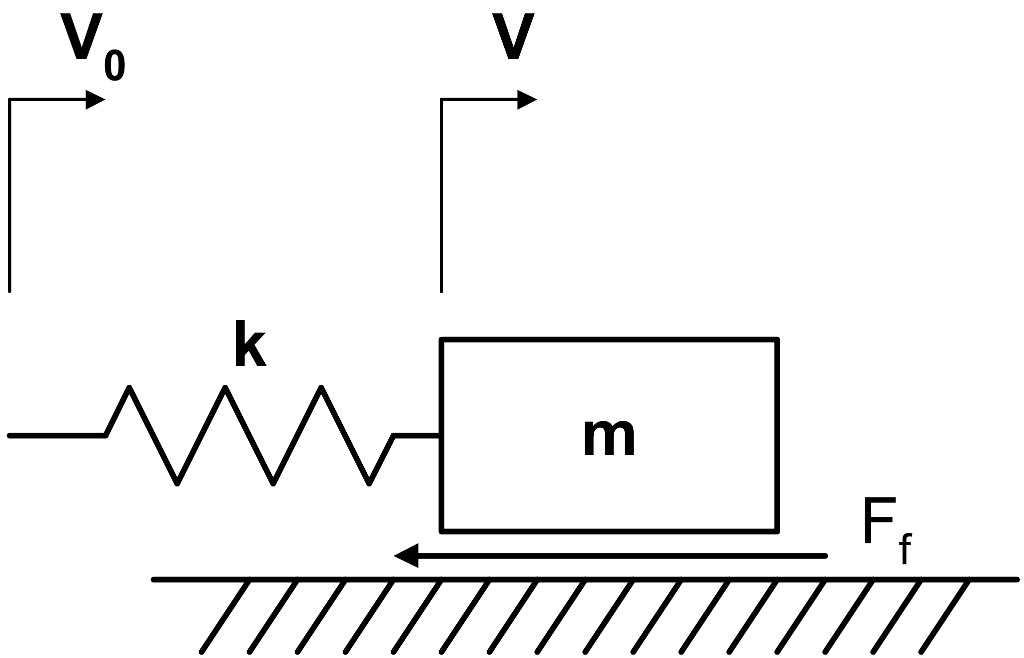
- In most control systems, Coulomb friction is a nuisance.
- Coulomb friction is difficult to model and troublesome to deal with in control system design.
- It is a nonlinear phenomenon in which a force is produced that tends to oppose the motion of bodies in contact in a mechanical system.
- Undesirable 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

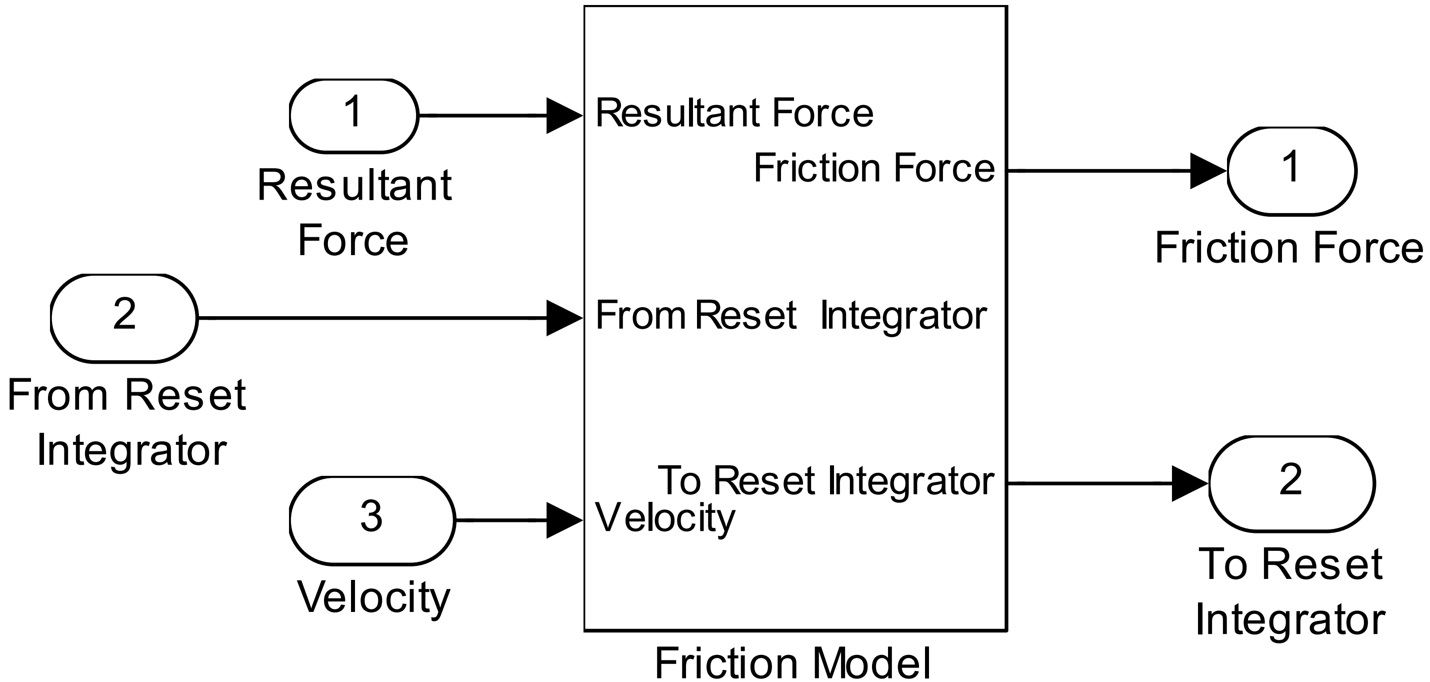
$k = 100$ N/m

$F_{stick} = 0.25$ N

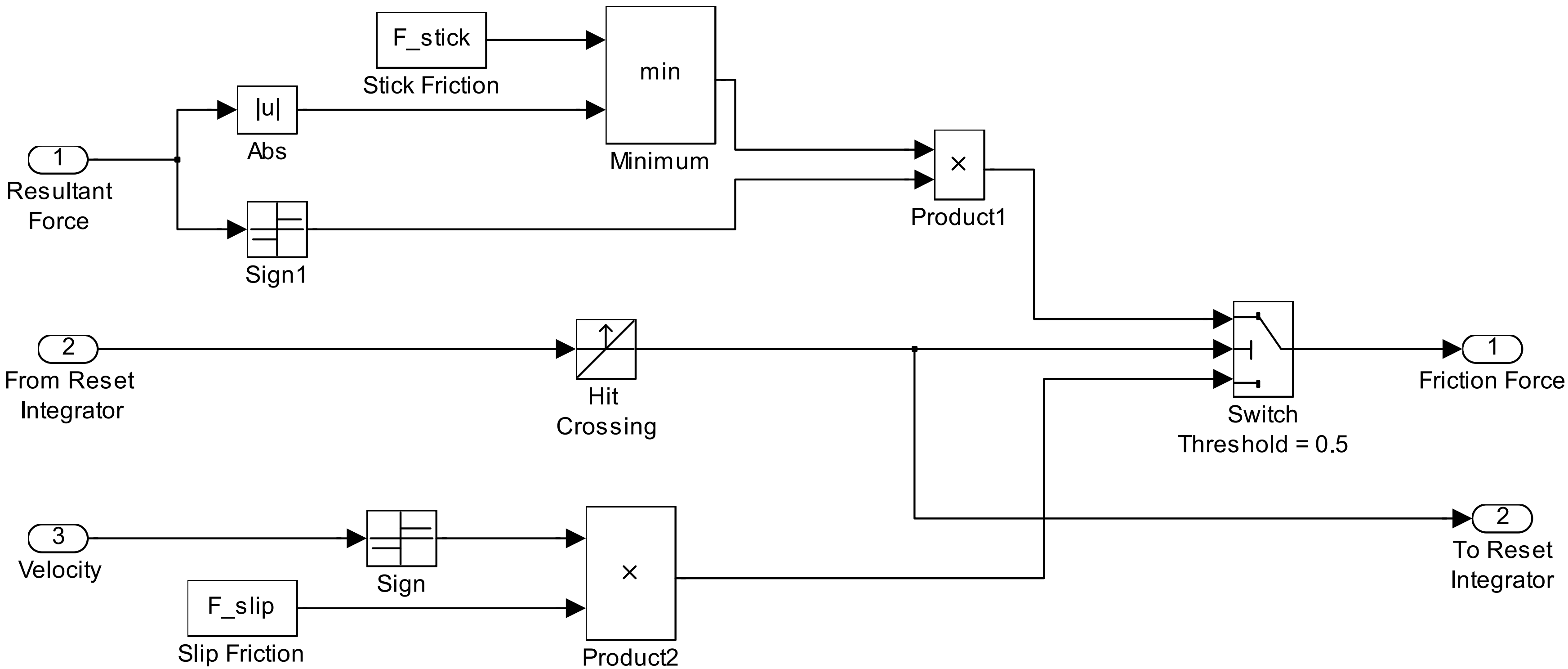
$F_{slip} = 0.20$ N (assumed independent of velocity)

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

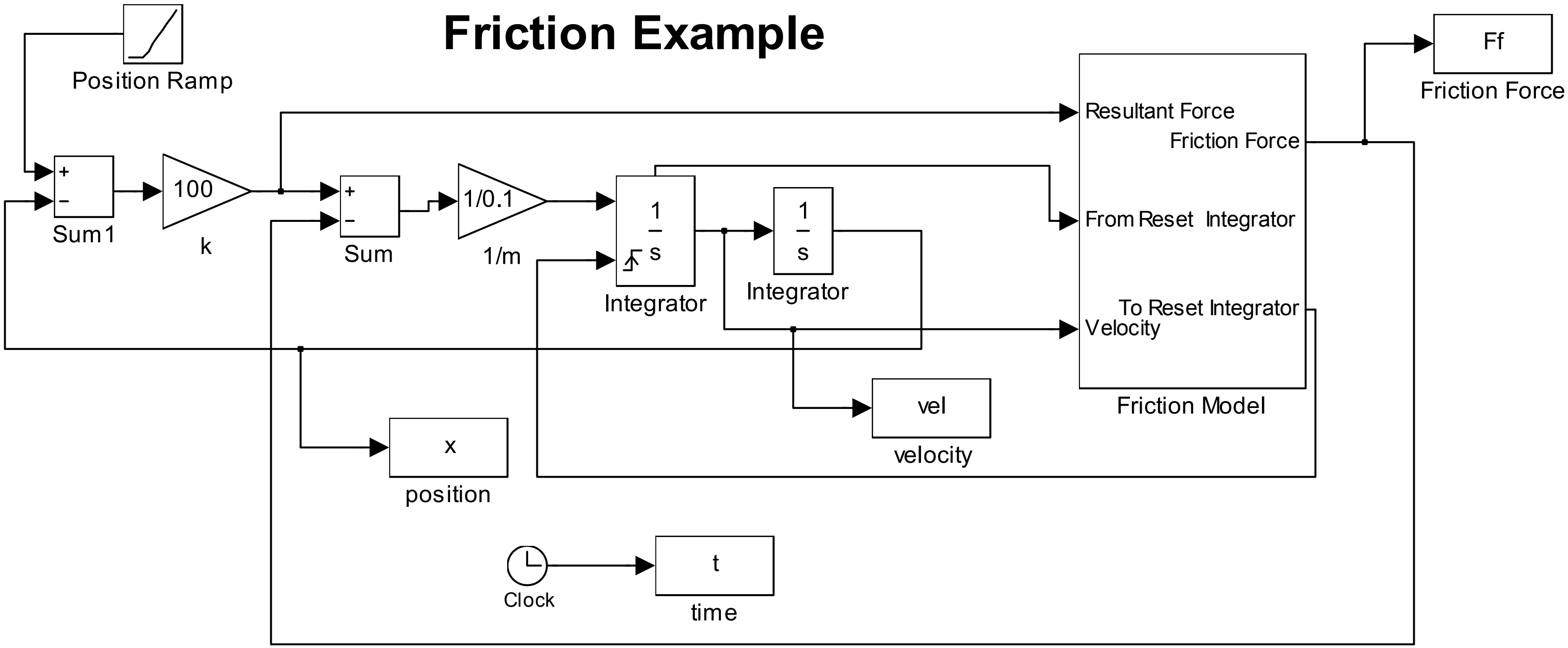
Friction Model in Simulink



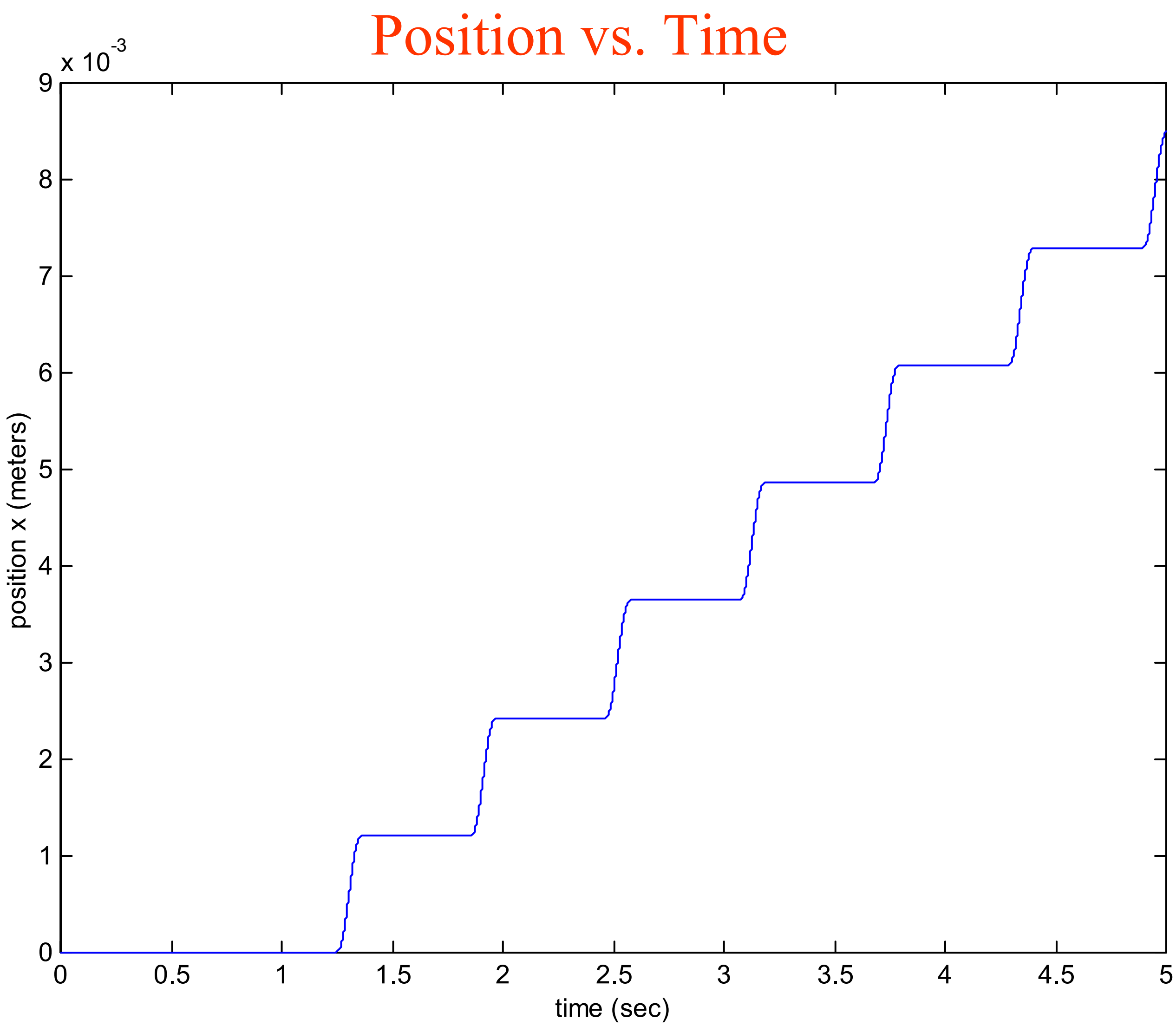
Simulink Block Diagram



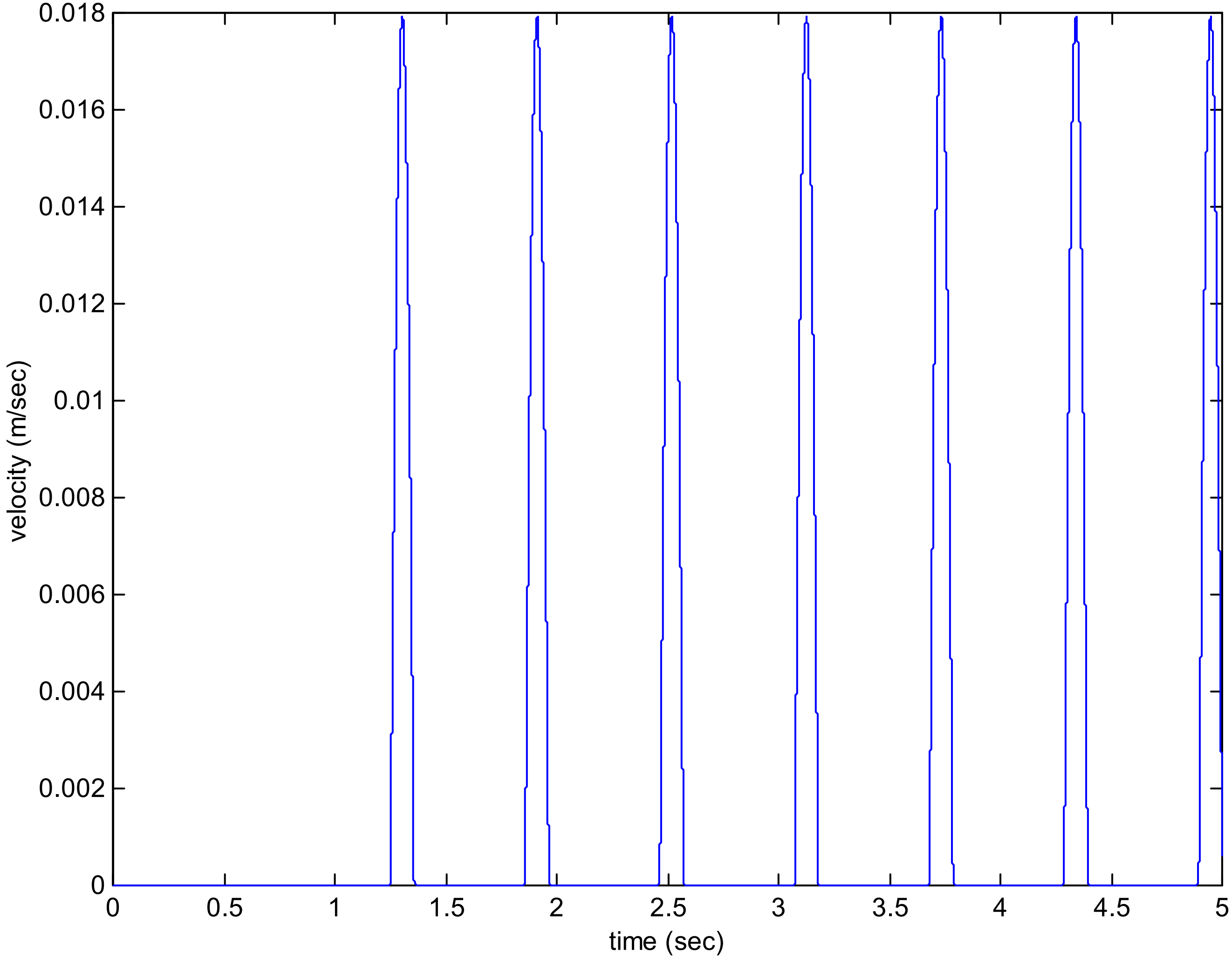
Example with Friction Model



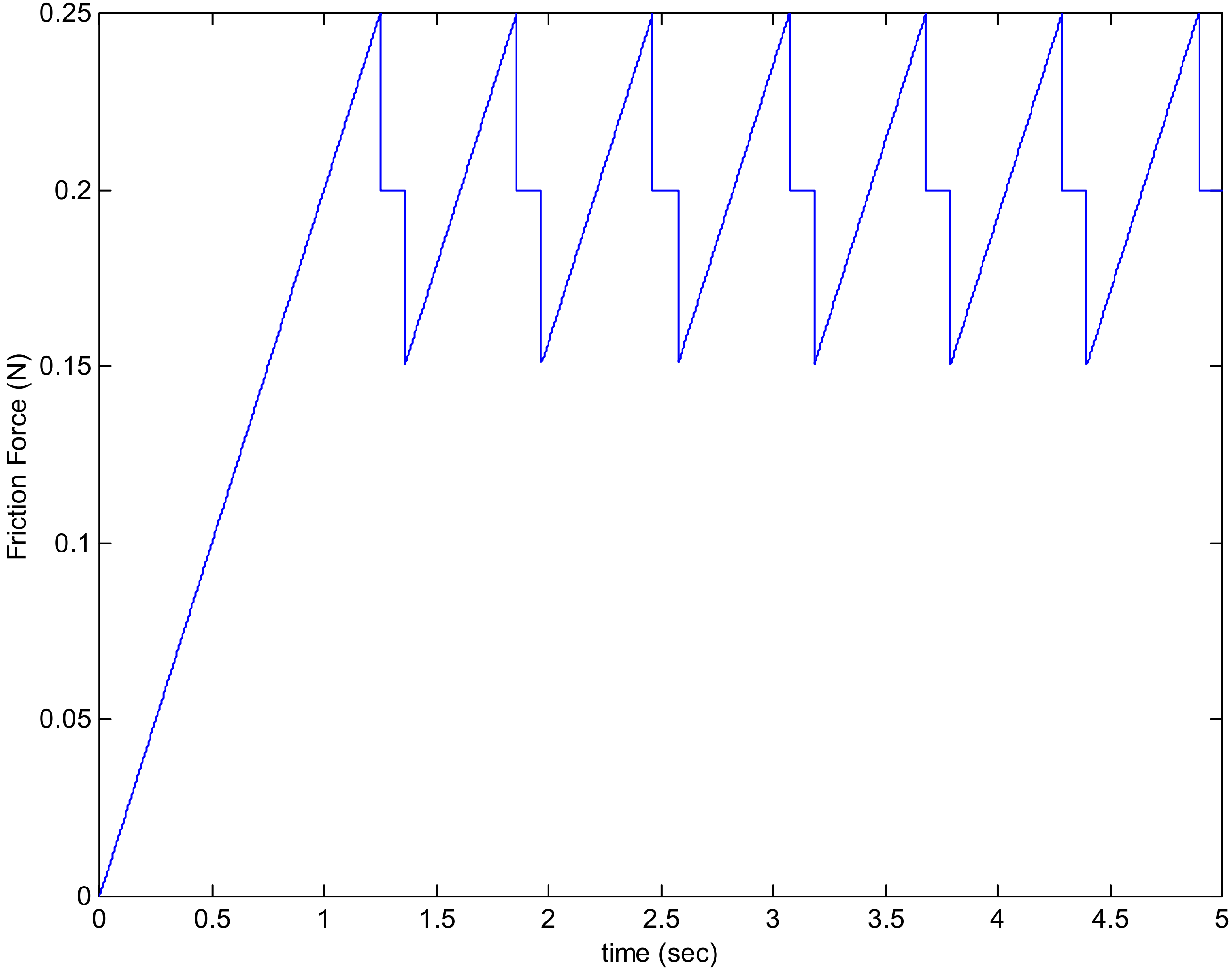
Model Simulation Results



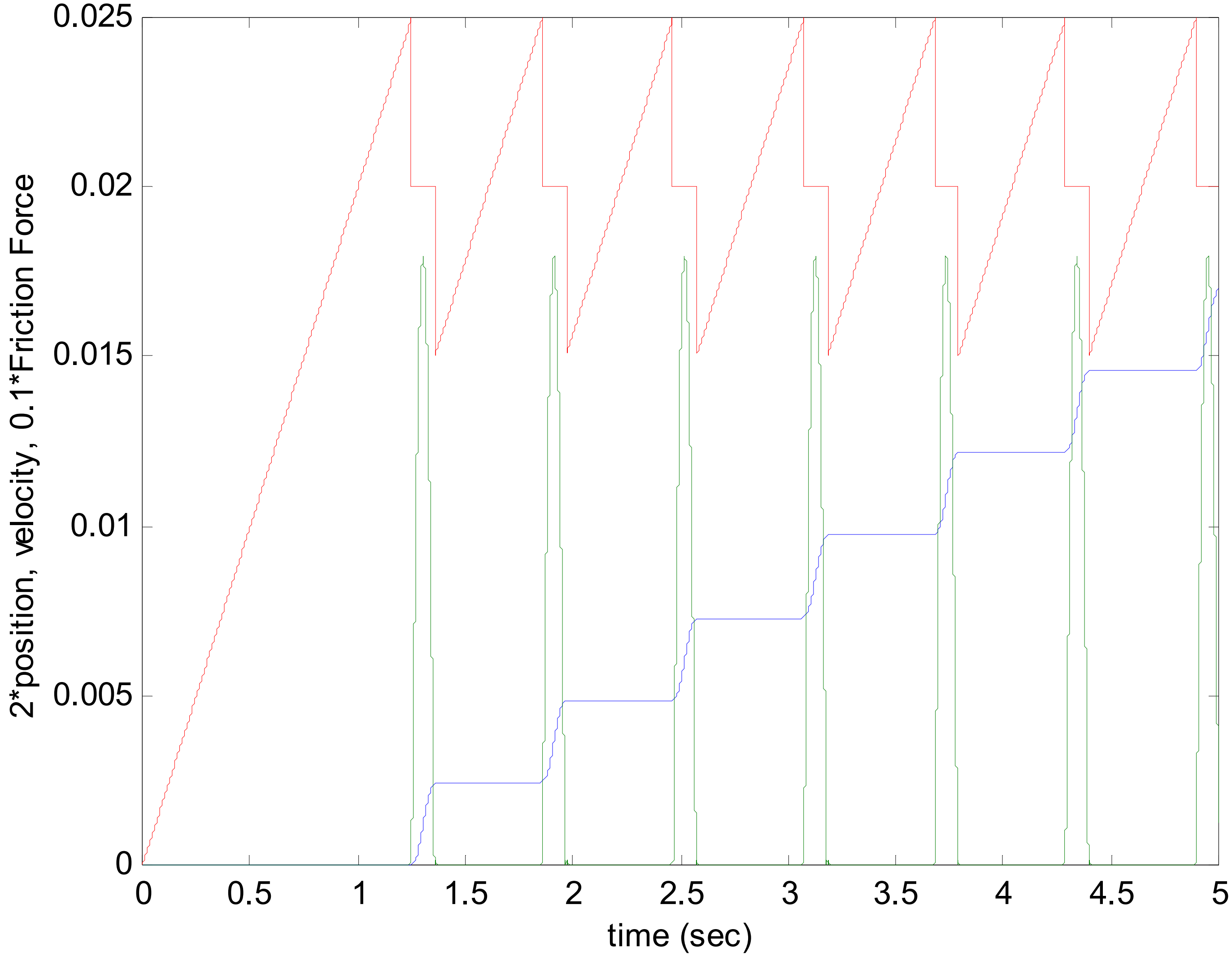
Velocity vs. Time



Friction Force vs. Time

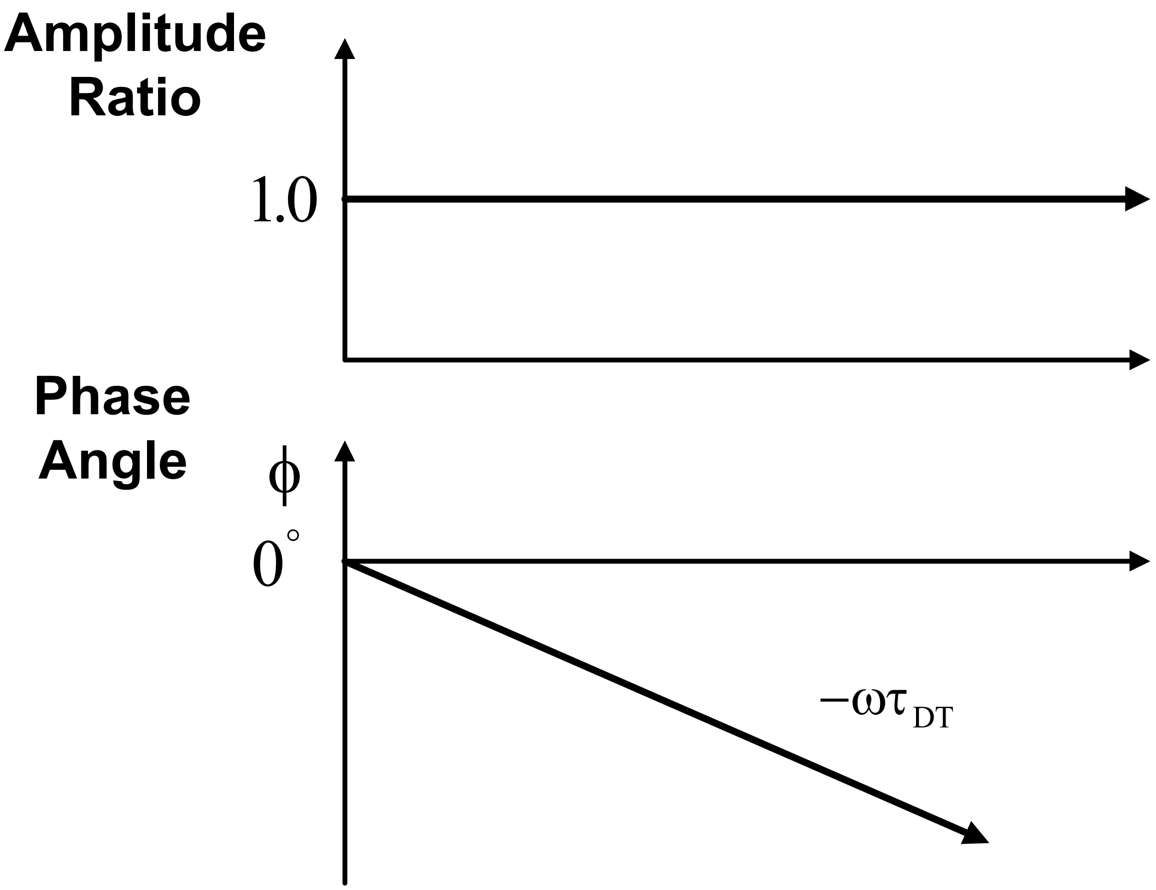


Position, Velocity, Friction Force vs. Time



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):

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

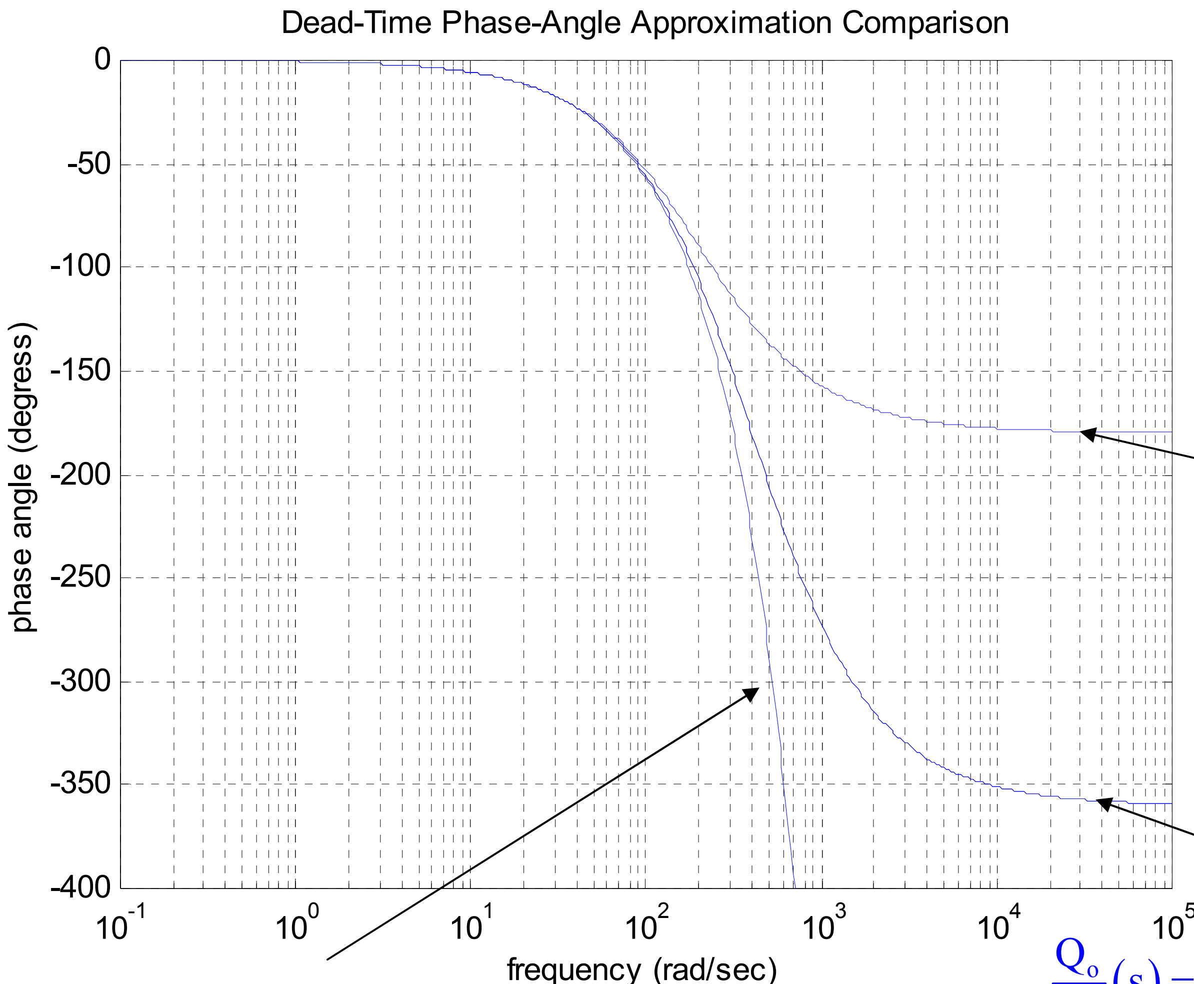
– In some cases, a very crude approximation given by a first-order lag is acceptable:

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

- 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:



$$\tau_{dt} = 0.01$$

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

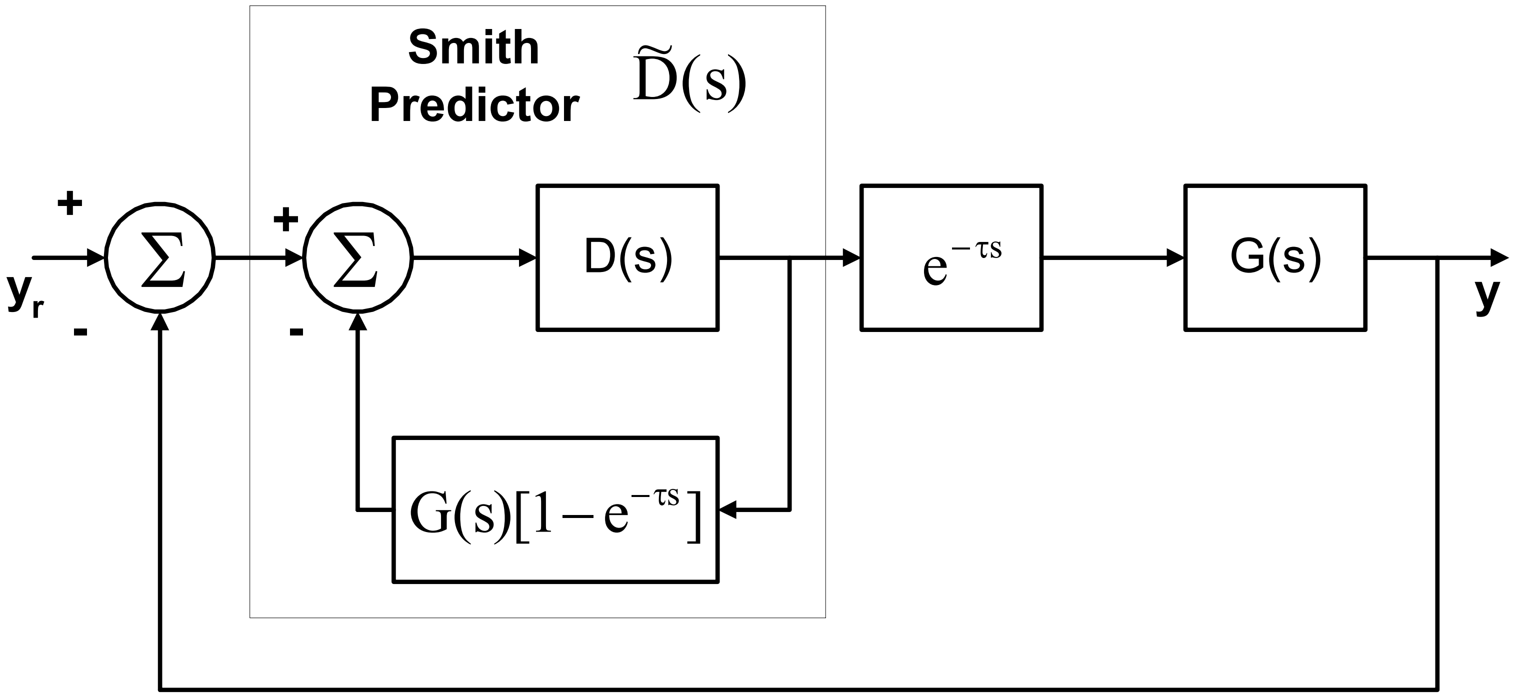
$$e^{-\tau_{dt} s} = 1 \angle - \omega \tau_{dt}$$

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

- 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



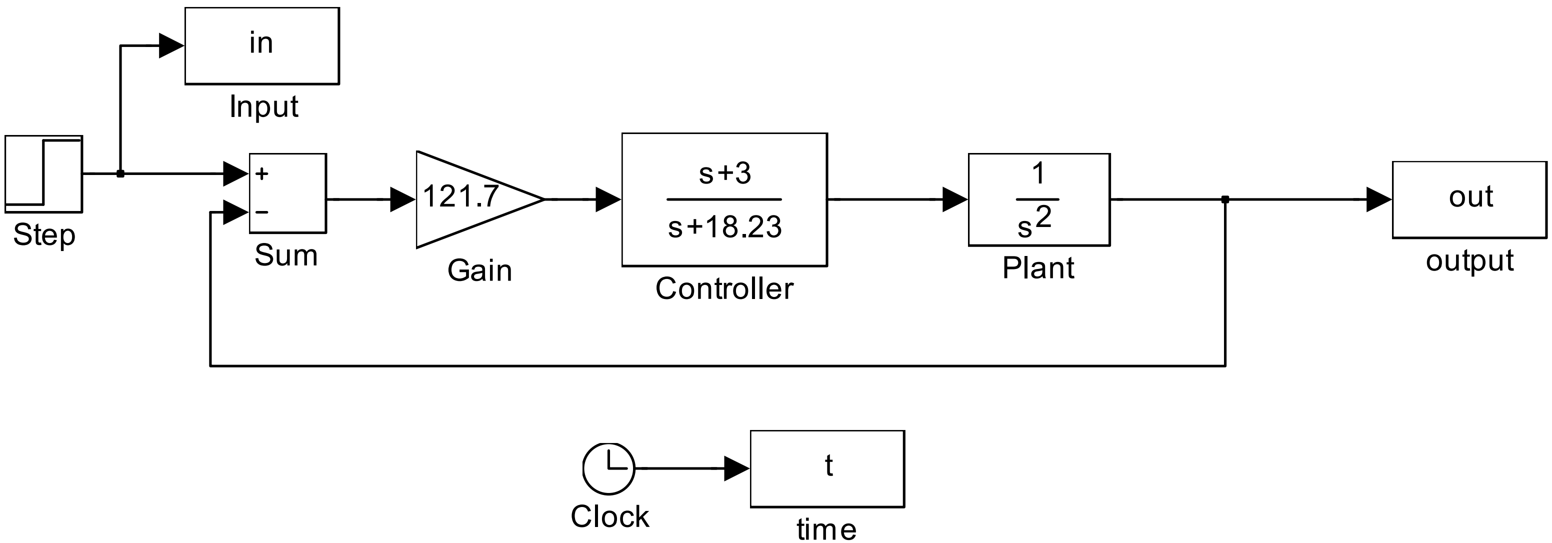
$$\tilde{D}(s) = \frac{D(s)}{1+(1-e^{-\tau s})D(s)G(s)} \qquad \frac{y}{y_r} = \frac{\tilde{D}(s)G(s)e^{-\tau s}}{1+\tilde{D}(s)G(s)e^{-\tau s}} = \frac{D(s)G(s)}{1+D(s)G(s)}e^{-\tau s}$$

- 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.

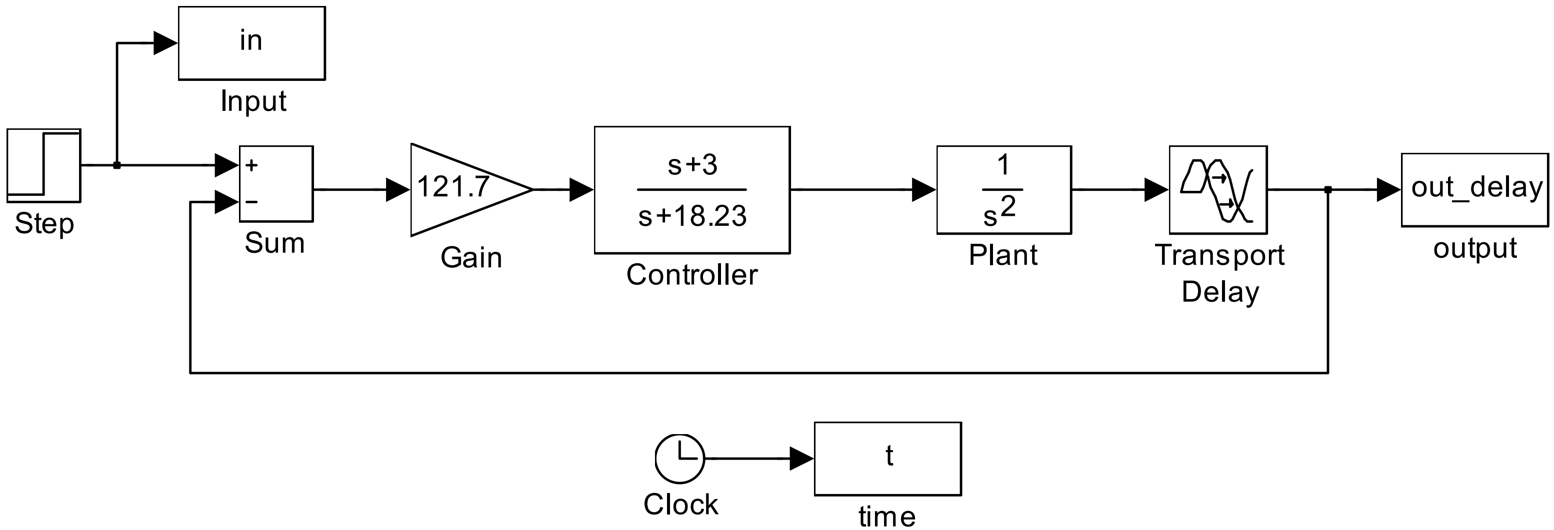
- Implementation Issues
  - You must know the plant transfer function and the time delay with reasonable accuracy.
  - You need a method of realizing the pure time delay that appears in the feedback loop, e.g., Pade approximation:

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

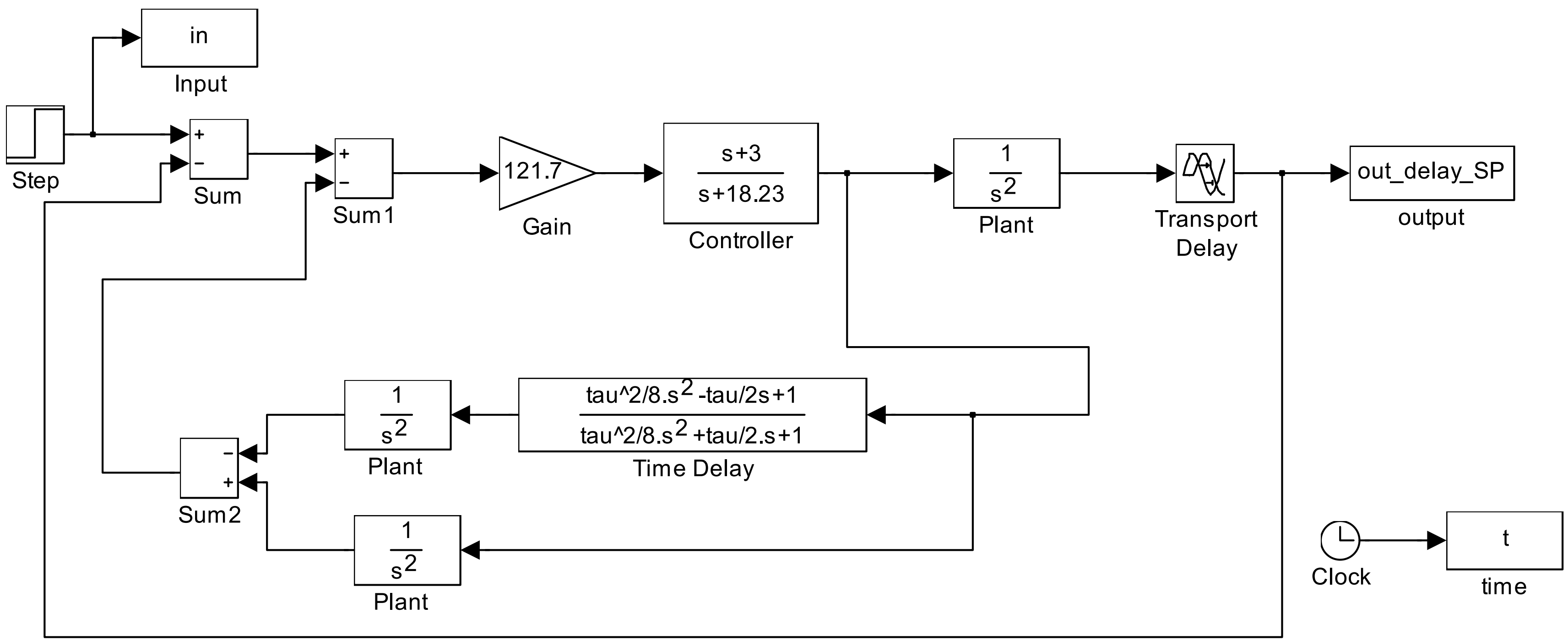
Example Problem



Basic Feedback Control System with Lead Compensator

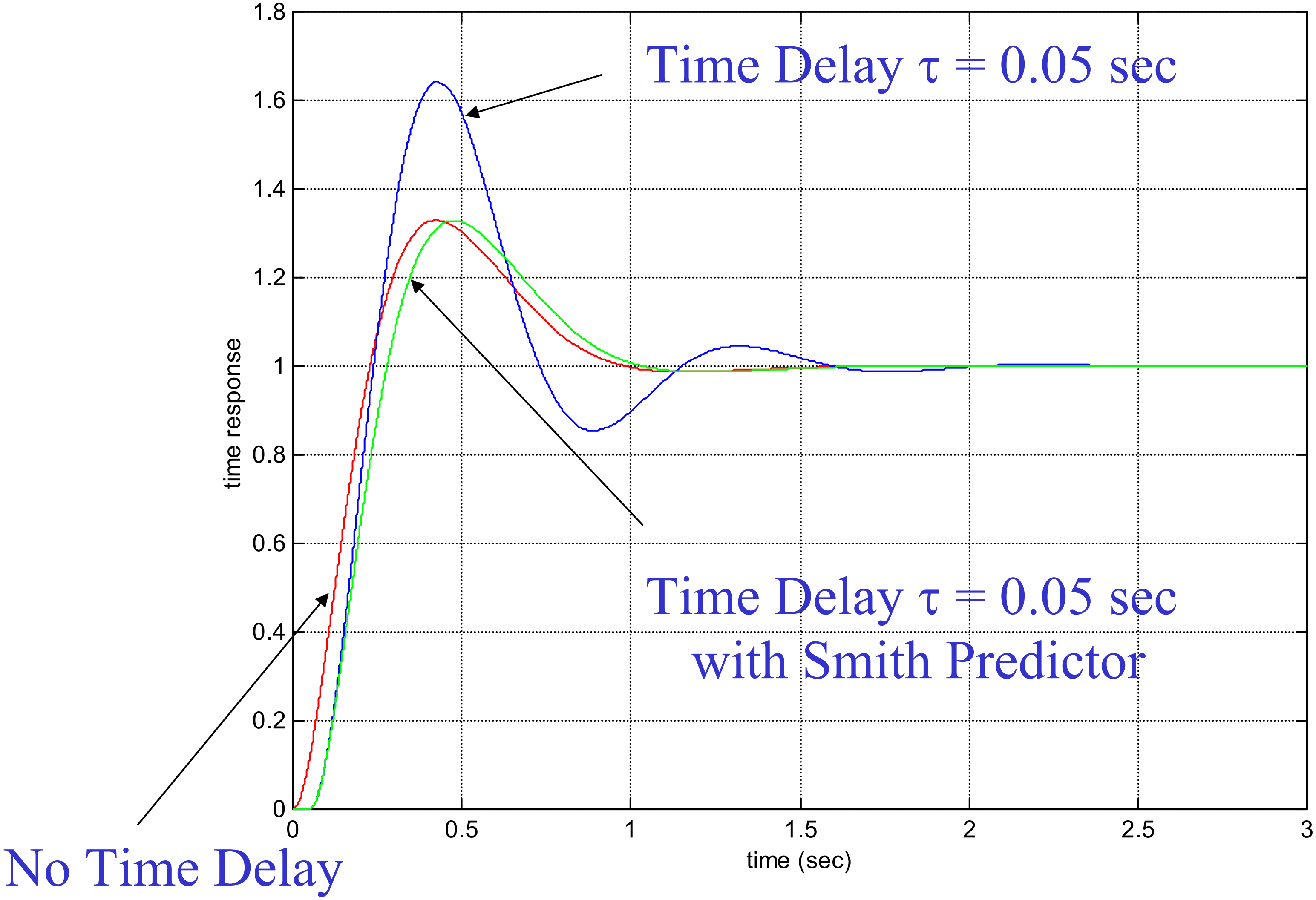


Basic Feedback Control System with Lead Compensator BUT with Time Delay $\tau = 0.05$ sec



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

System Step Responses



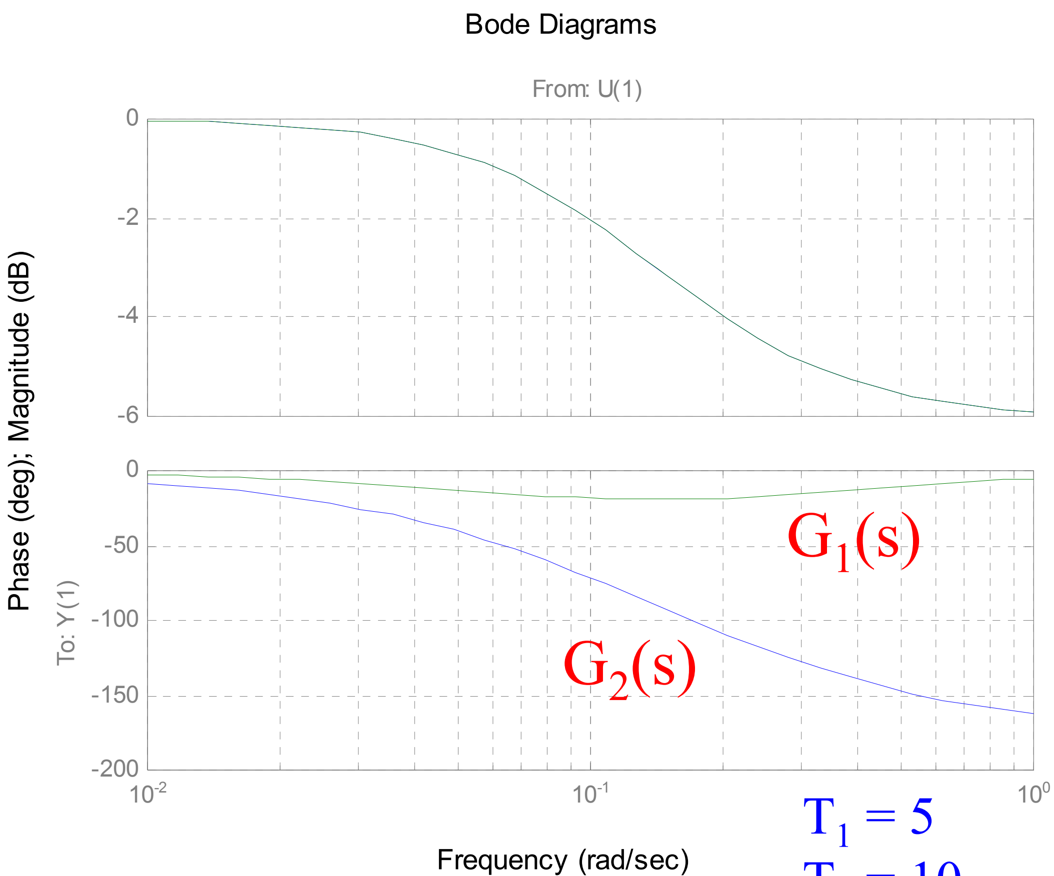
- Comments
  - The system with the Smith Predictor tracks reference variations with a time delay.
  - The Smith Predictor minimizes the effect of the DT on stability as model mismatching is bound to exist. This however still allows tighter control to be used.
  - What is the effect of a disturbance?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

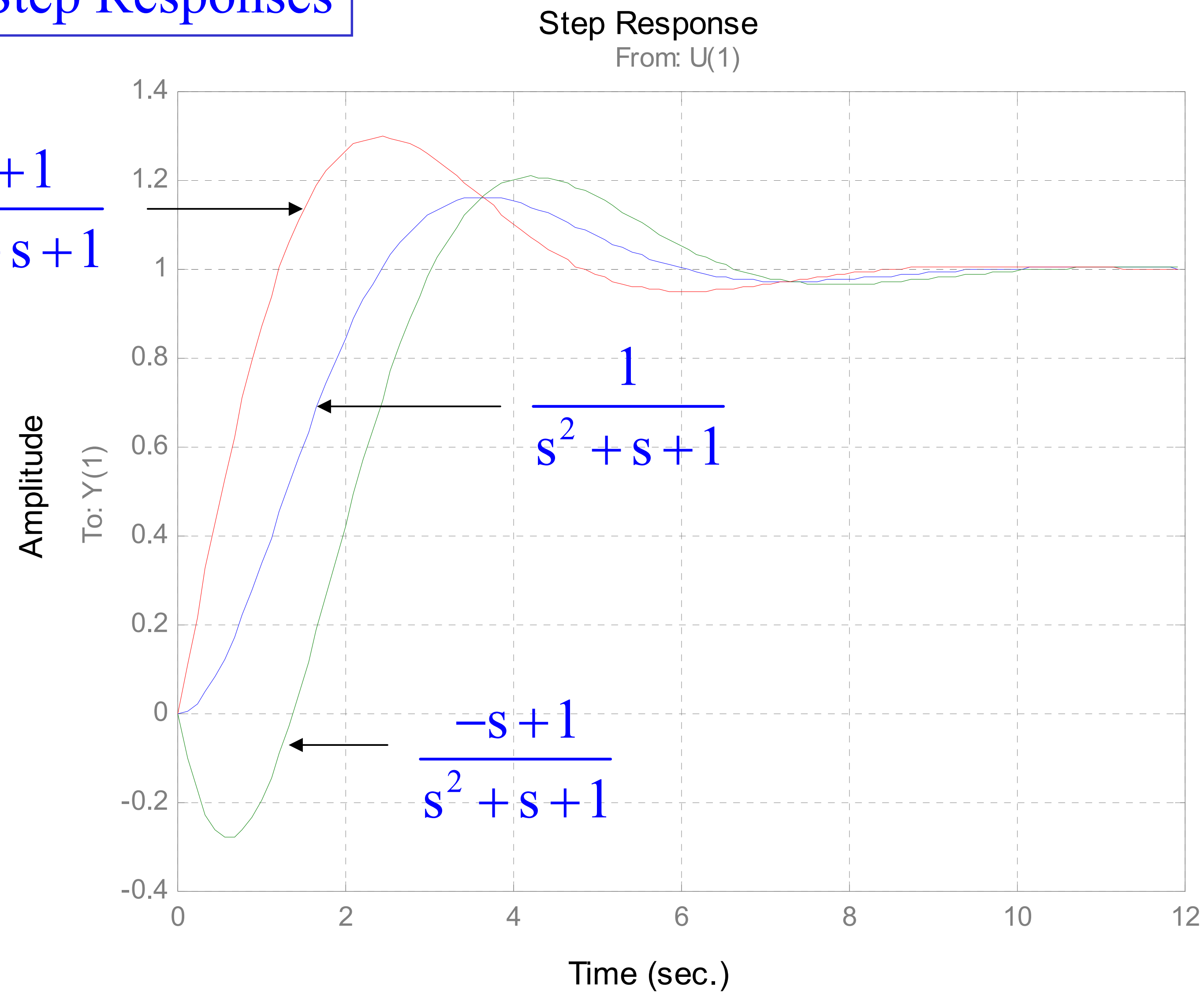
$$G(s) = \frac{1 - T_1 s}{1 + T_1 s}$$

- 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

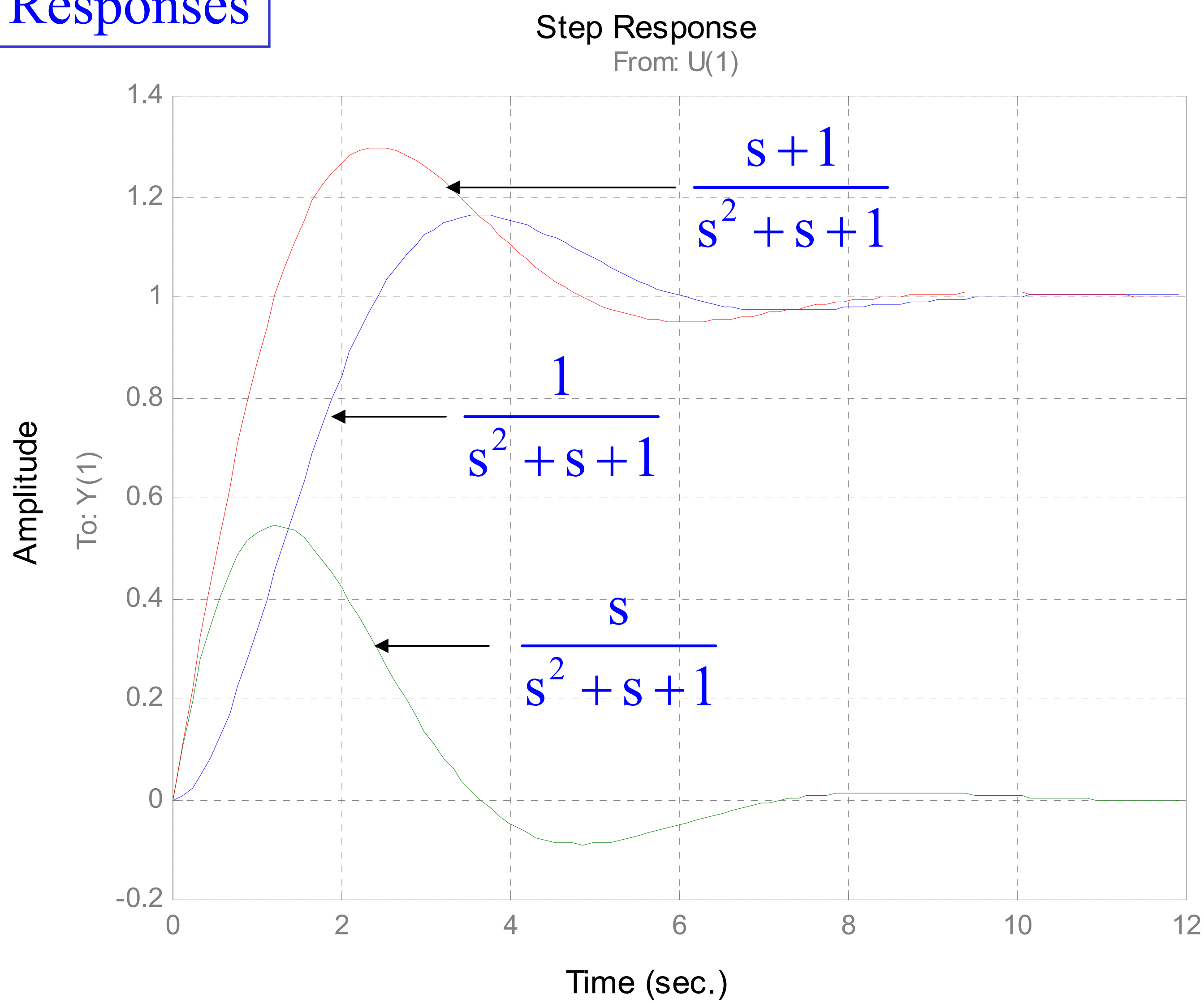


$$\frac{s+1}{s^2+s+1}$$

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

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

Unit Step Responses

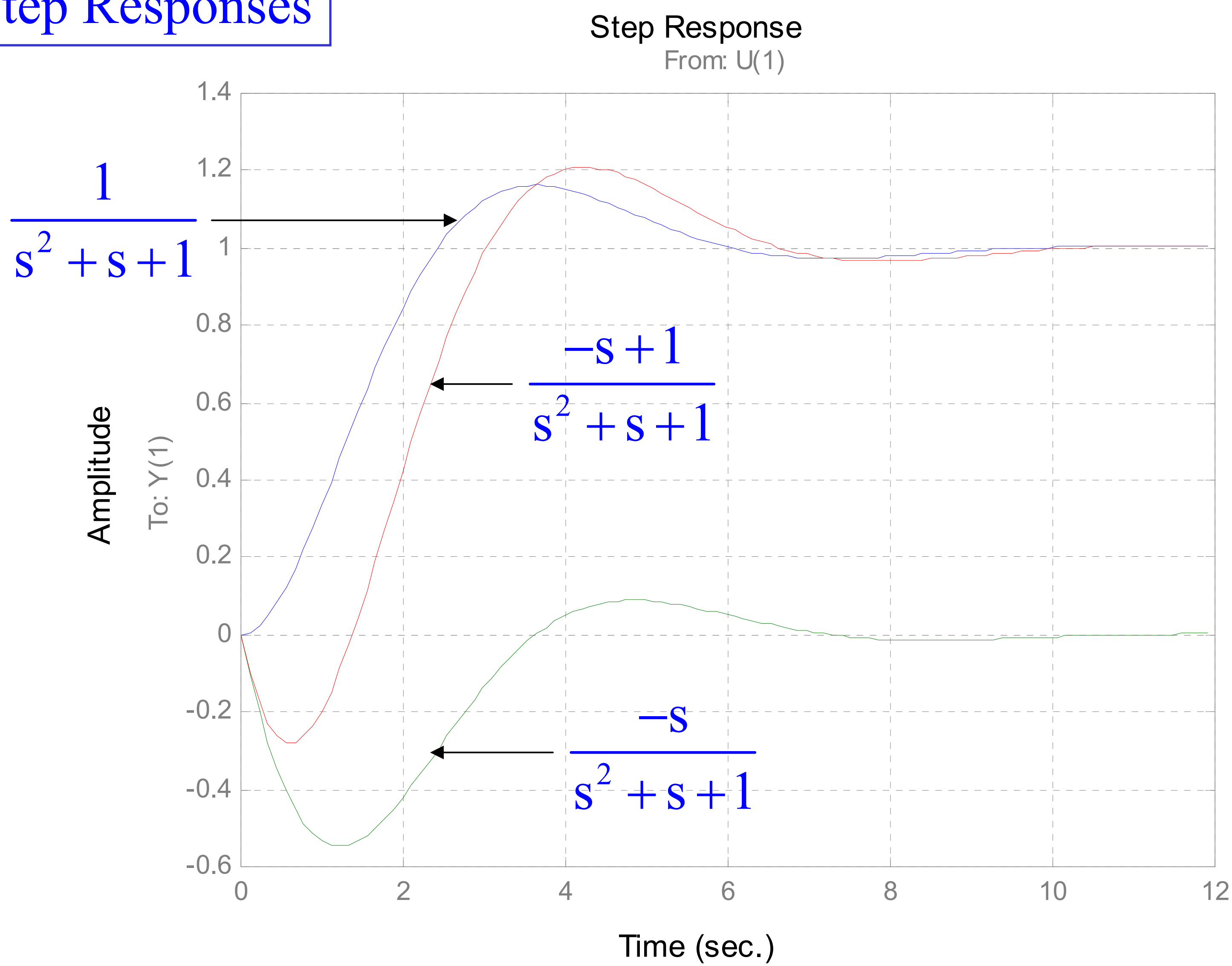


$$\frac{s+1}{s^2+s+1}$$

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

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

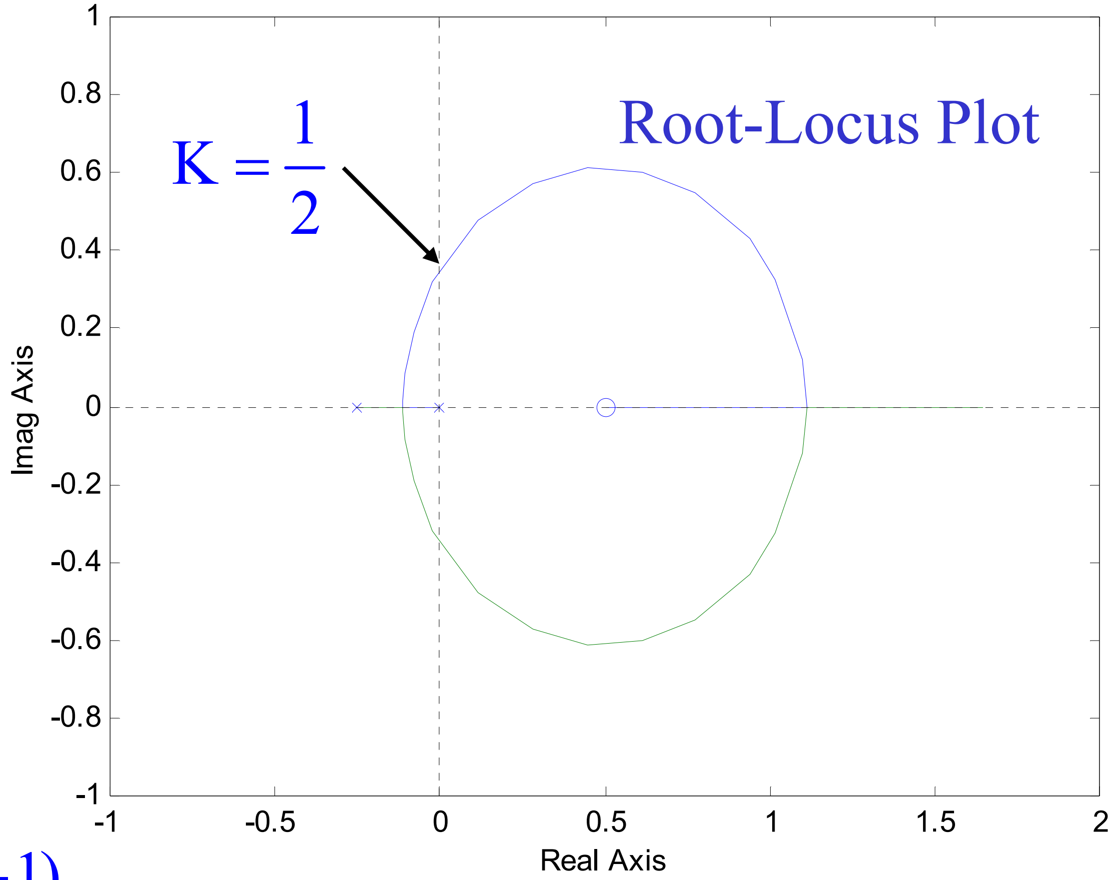
Unit Step Responses



- 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:

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

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



Angle Condition:

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

$$=\angle\frac{K(2s-1)}{s(4s+1)}+180°=\pm180°\left(2k+1\right)$$ or $$\angle\frac{K(2s-1)}{s(4s+1)}=0°$$

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.

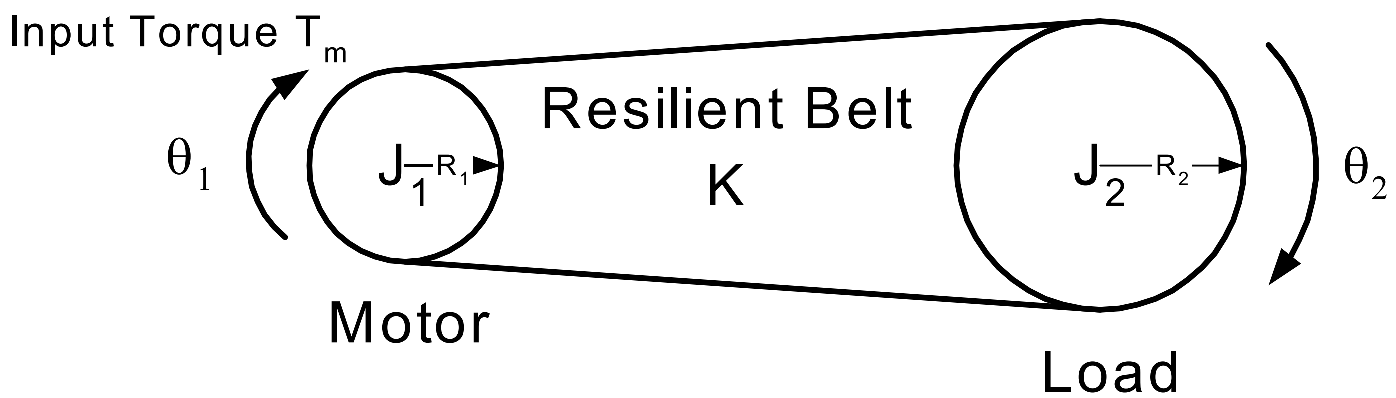
- If the precise nature of the resonances are known, they can be modeled and included in the design model.
- However, in many applications the frequencies of the poles (and neighboring zeros) of the resonances are not known with precision or may shift during the operation of the system. A small error 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

<u>Rigid Belt Case</u>

$$R_1\theta_1 = R_2\theta_2$$

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

– 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.

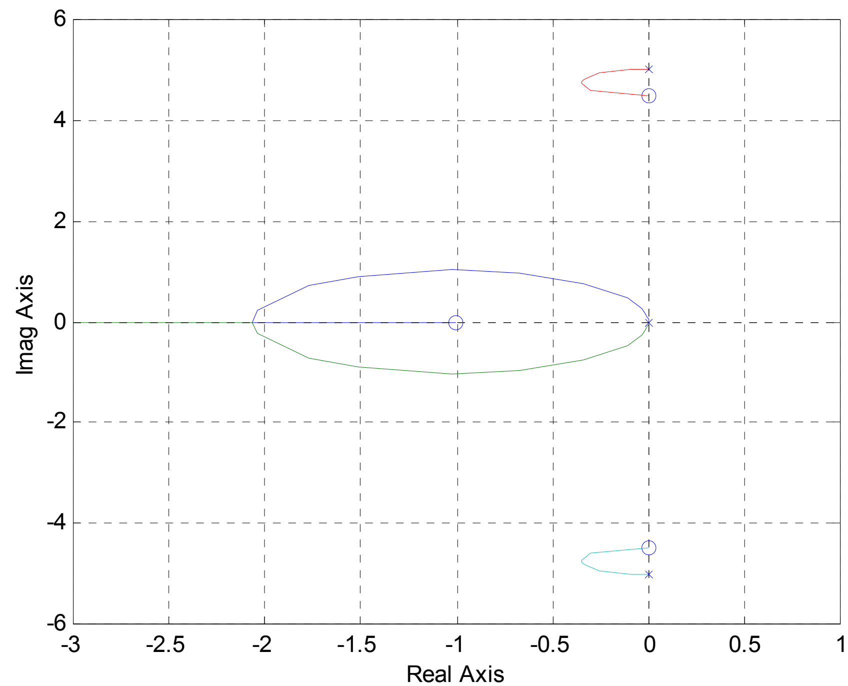
- Complex Zeros correspond to the frequencies where the system behaves as an energy sink.
- Complex Zeros represent frequencies at which energy being applied by the input is completely trapped in the energy storage elements of a sub-portion of the original system such that no output can ever be detected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

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

Root-Locus Plot with PD Controller



Poles:

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

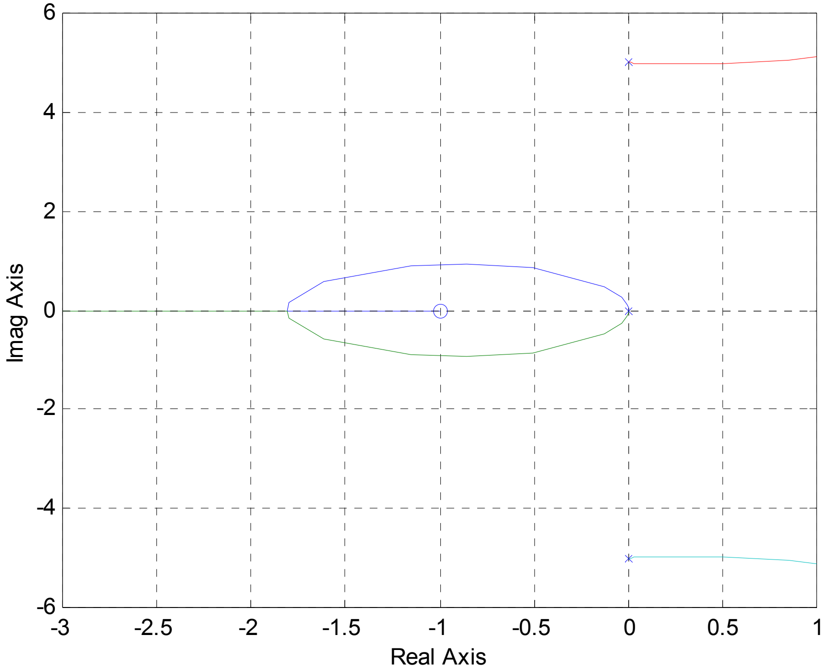
Zeros: $\pm i\sqrt{\frac{2KR_2^2}{J_2}}$

• 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.

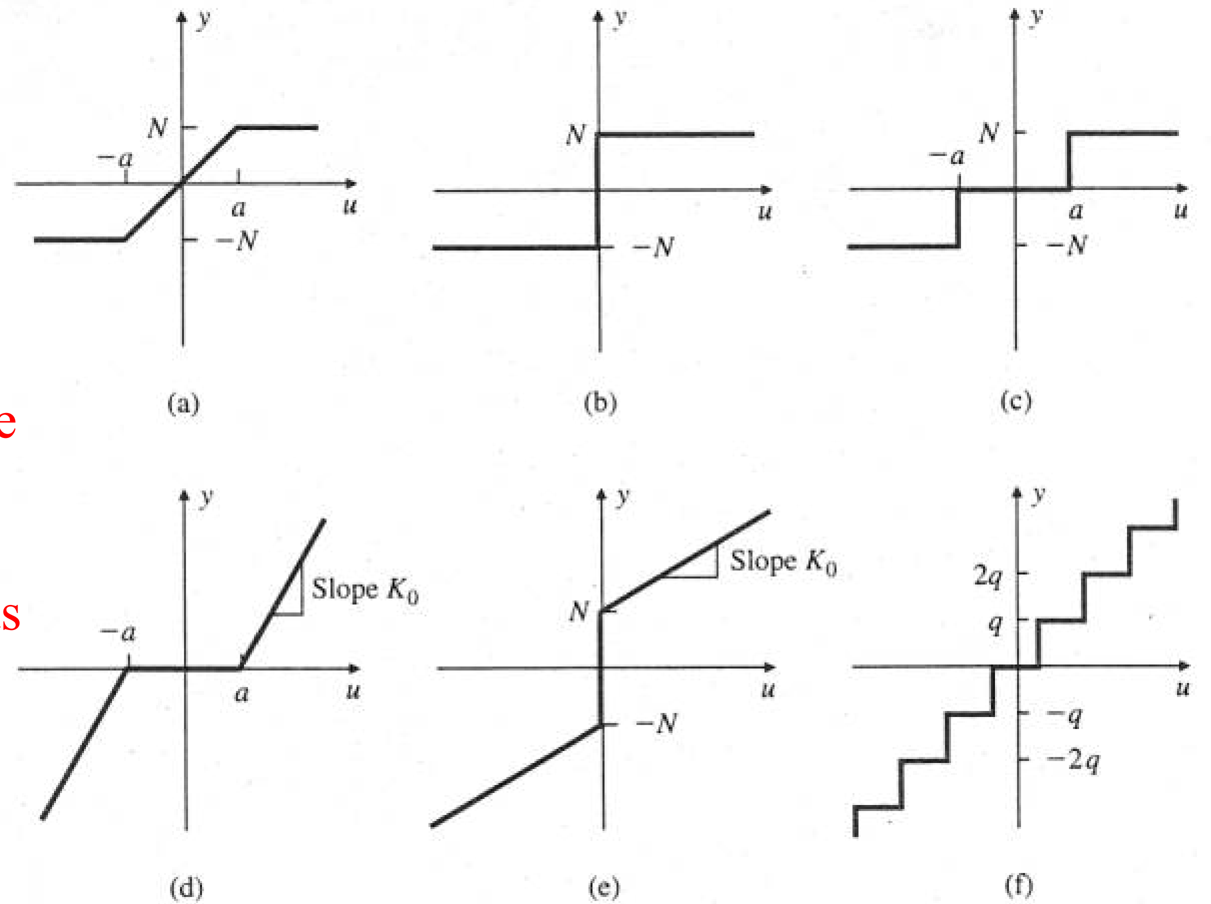
- If you do not want to rely on the uncertain characteristics of the hardware to stabilize the system, you must attend to the stabilization in your compensator design. A simple low-pass filter between the the control output and the input to the motor might do the job.
- Remember, if you want the system 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

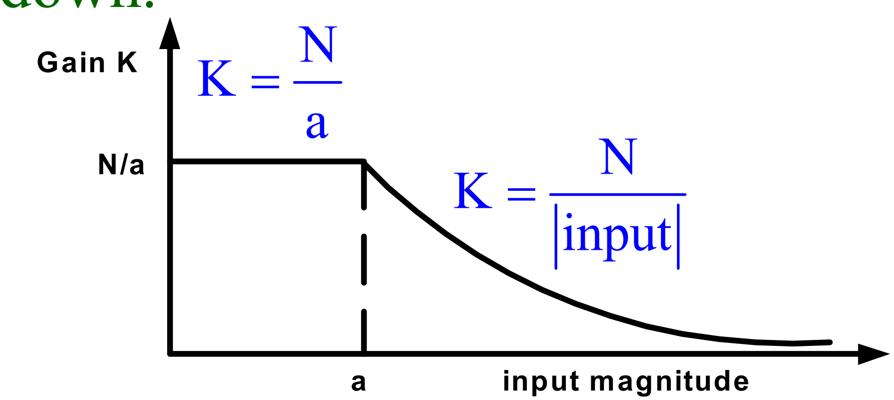
(a) Saturation
(b) Relay
(c) Relay with Dead Zone
(d) Gain with Dead Zone
(e) Pre-loaded Spring or Coulomb plus Viscous Friction
(f) Quantization



– 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.

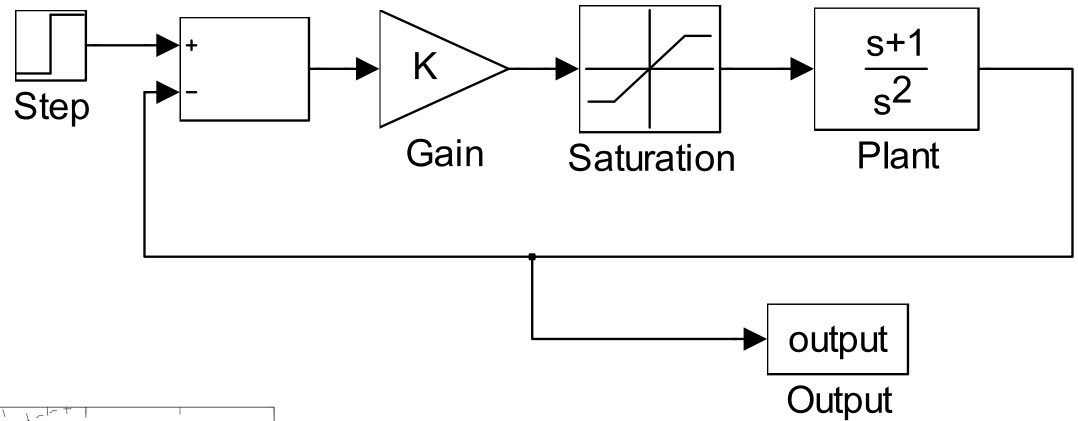
General Shape of the Effective Gain of Saturation

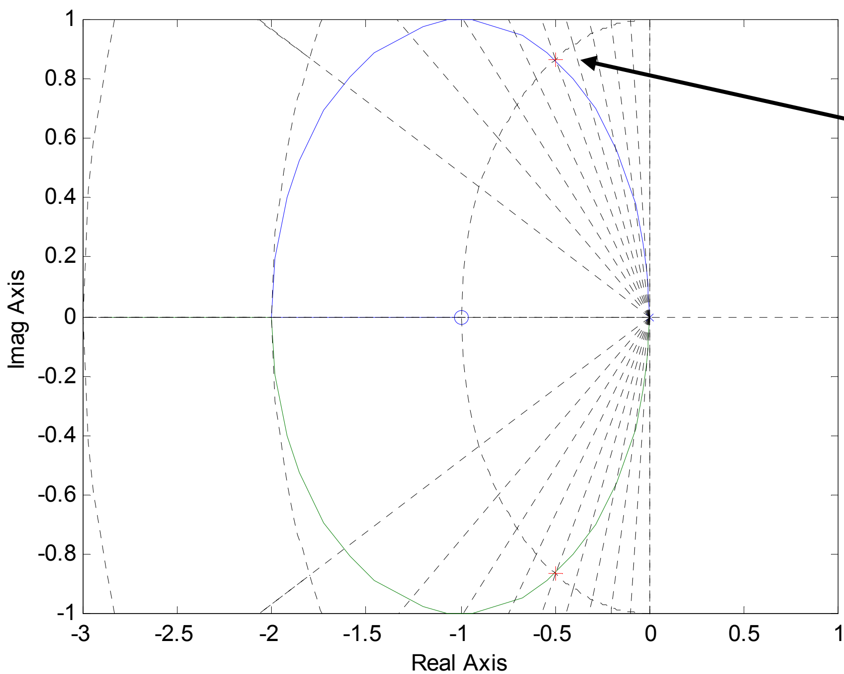


- An important aspect of control system design is *sizing the actuator*, which means picking the size, weight, power required, cost, and saturation level of the device.
- Generally, higher saturation levels require bigger, heavier, and more costly actuators.
- The key factor that enters into the sizing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

Saturation levels: ± 0.4

Dynamic System With Saturation





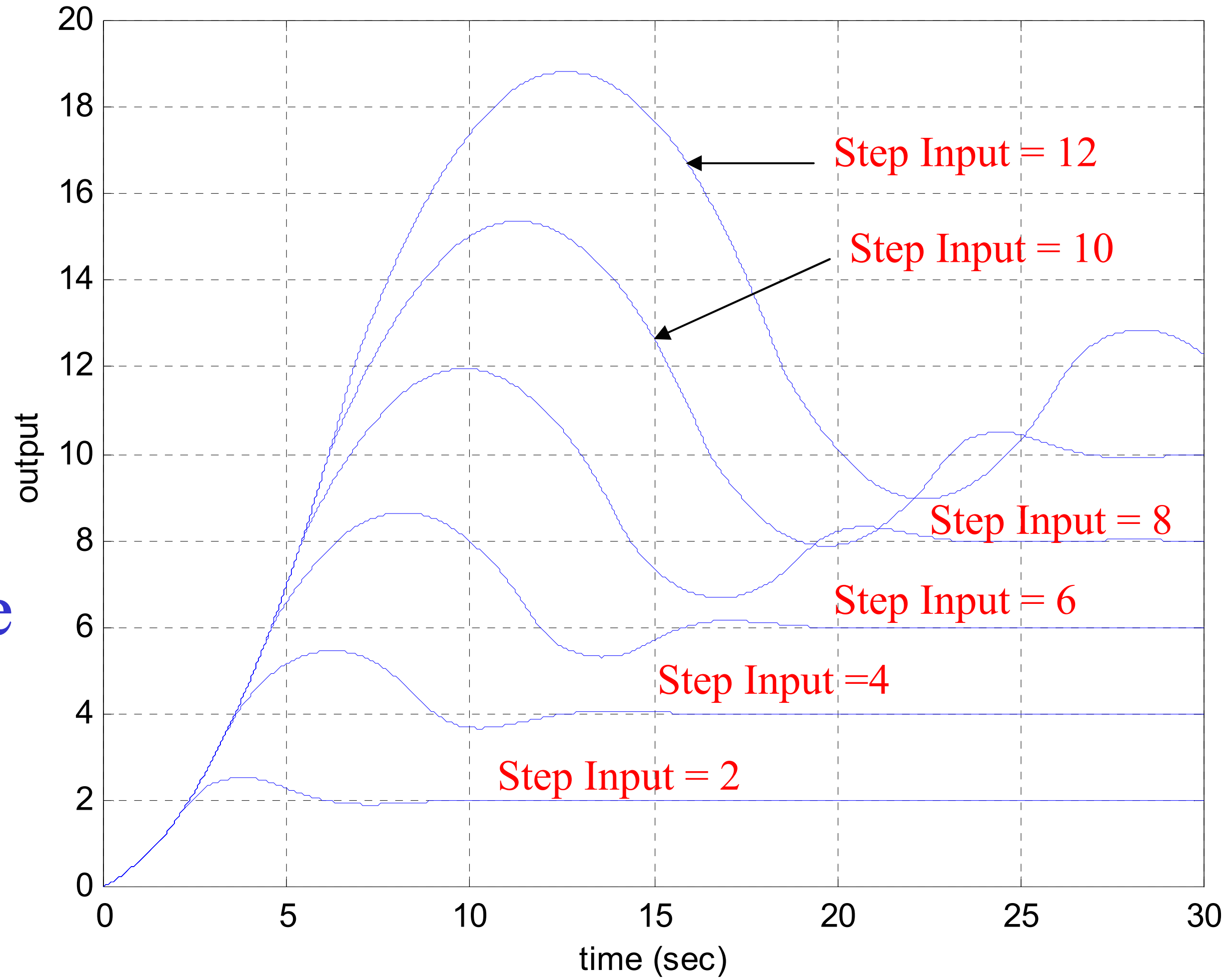
$\zeta = 0.5, K = 1$

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

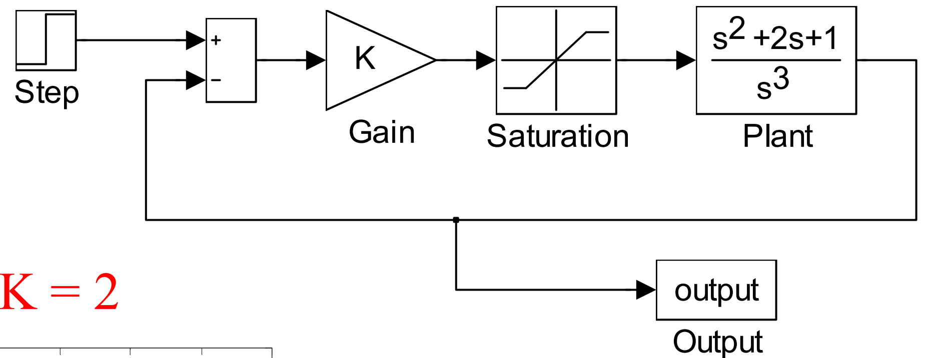
$\zeta = 0.5, K = 1$



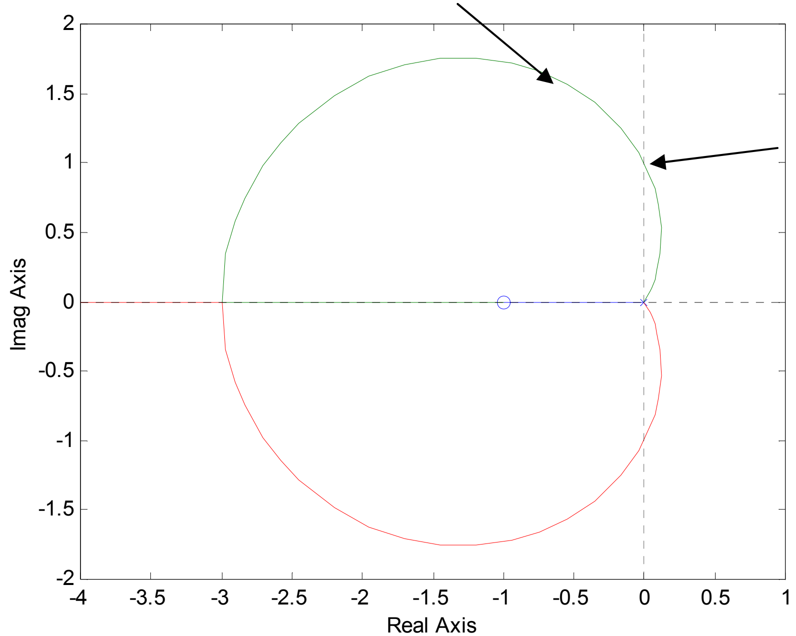
– Observations
  - As long as the signal entering the saturation remains less than 0.4, the system will be linear and should behave according to the roots at $\zeta = 0.5$.
  - However, notice that as the input gets larger, the response has more and more overshoot and slower and slower recovery.
  - 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



$\zeta = 0.5$, K = 2



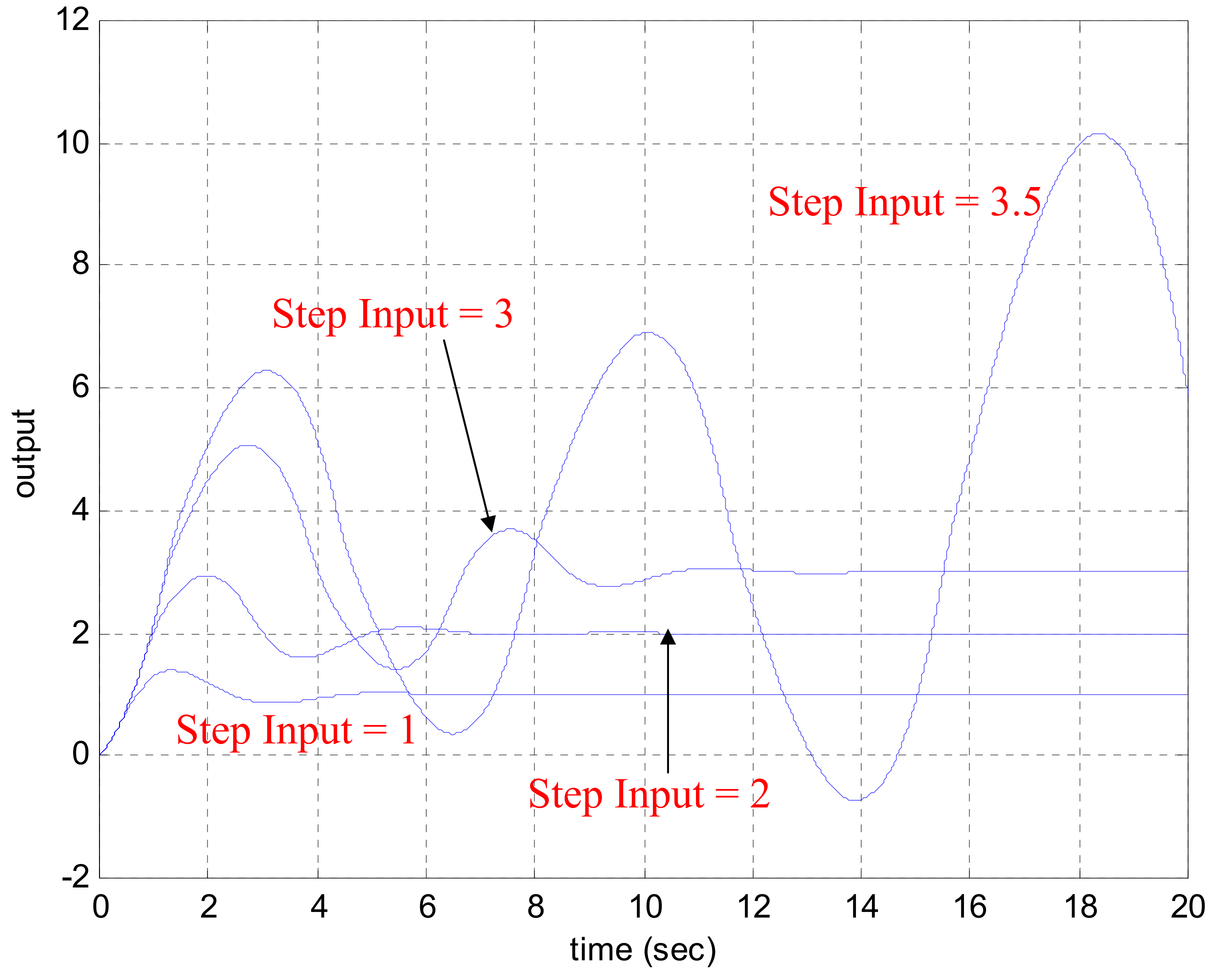
K = 1/2

Root-Locus Plot Without Saturation

System is stable for large gains but unstable for smaller gains

Step-Response Results

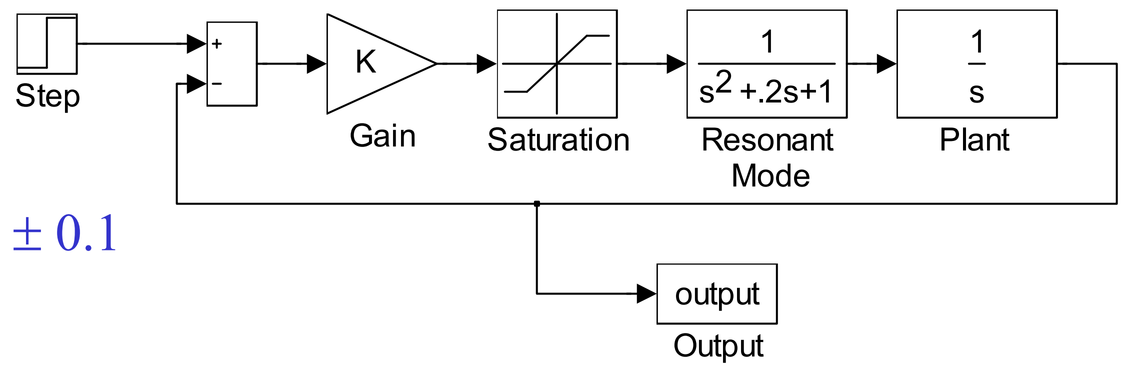
$\zeta = 0.5$, K = 2



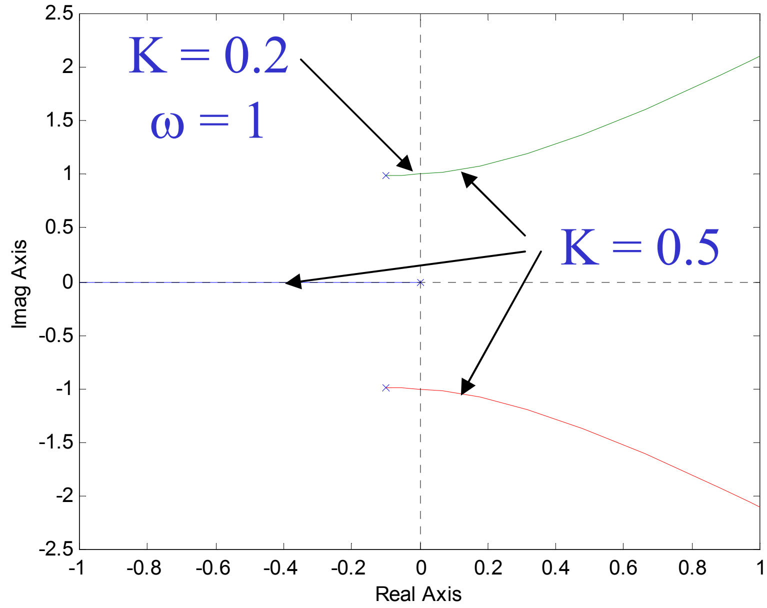
– Observations

- 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.

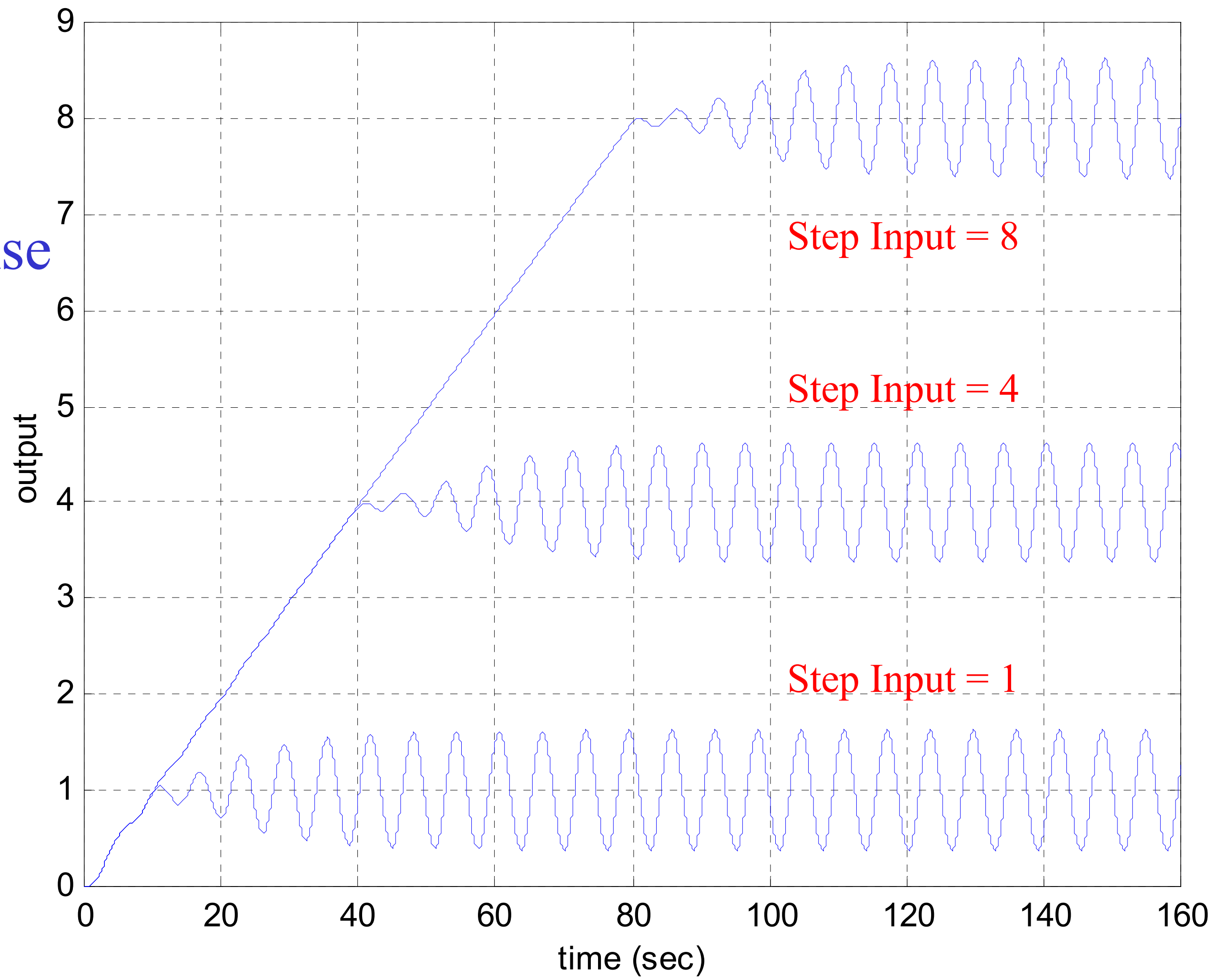


Saturation levels: ± 0.1



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.

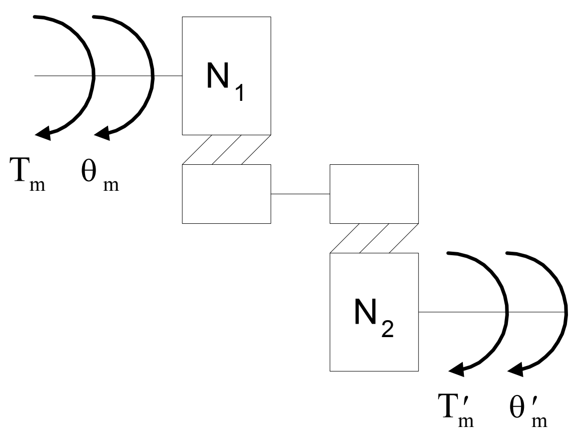
- As an example, consider a motor connected to a drive gear and that the driven gear is connected to an inertial load. The equations of motion are:

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

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

  – τ is the torque transmitted through the gears

Gear Train Relations:



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

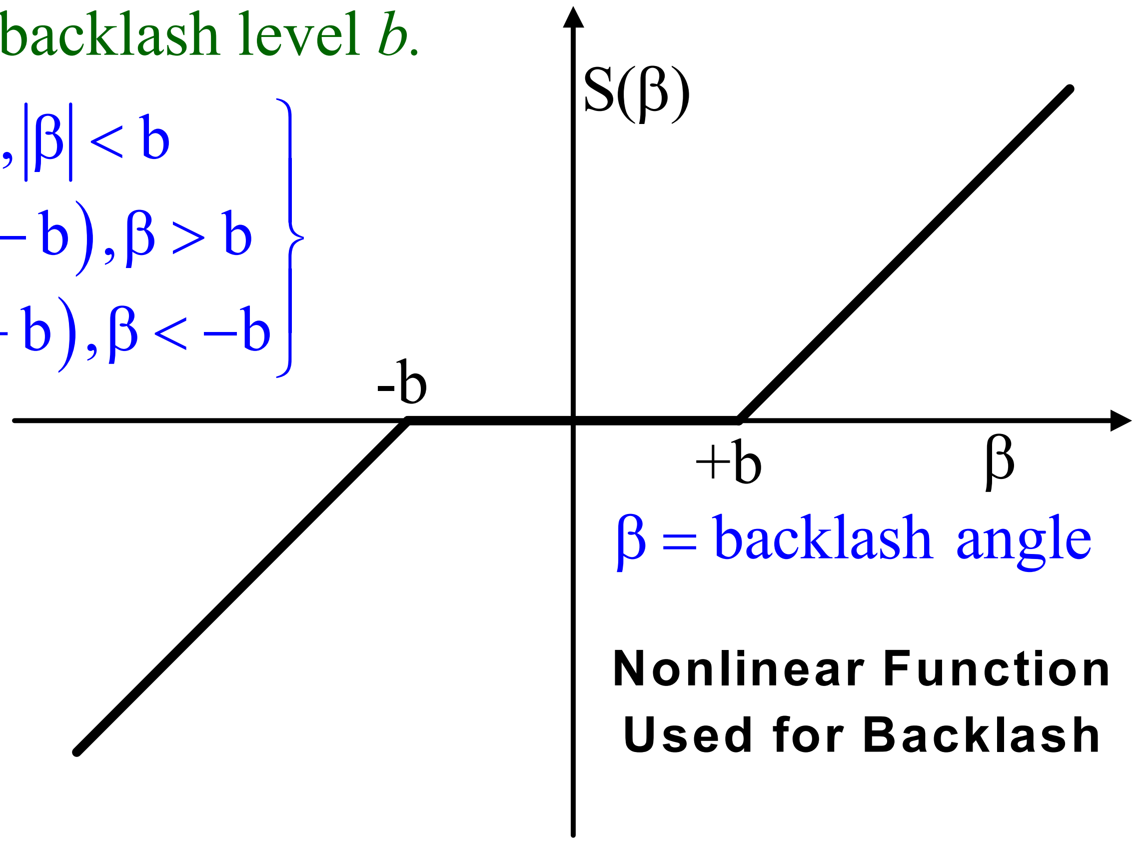
$$\frac{T_m}{T'_m} = \frac{N_1}{N_2} \equiv \frac{1}{N}$$

- 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 4th order; if they are engaged, the system has only one degree of freedom and the system is 2nd 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*.

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

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



β = backlash angle

Nonlinear Function Used for Backlash

– The differential equations can now be written as:

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

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

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

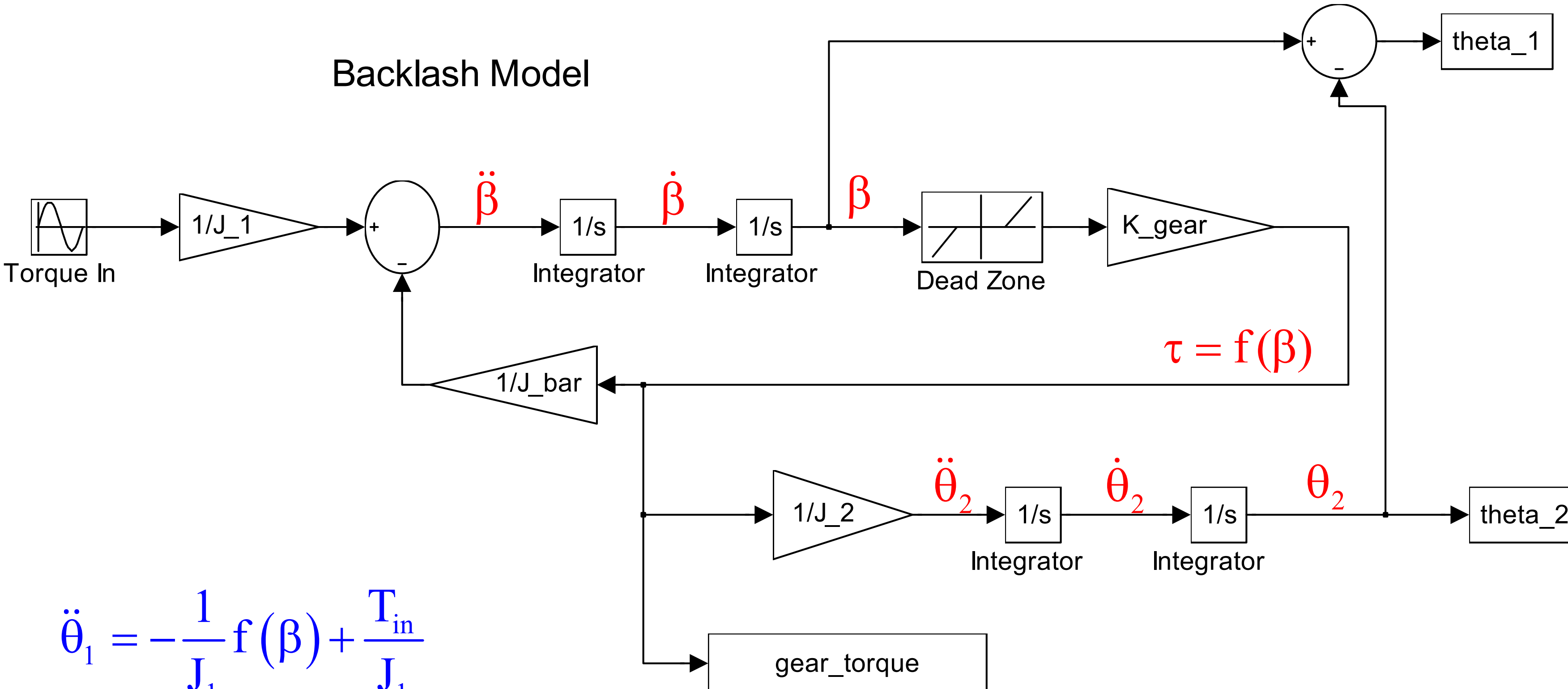
– The differential equation for β is obtained from the above equations noting that $\ddot{\beta} = \ddot{\theta}_1 - \ddot{\theta}_2$

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

– This equation defines the dynamics of a nonlinear oscillator.

$$\ddot{\beta} = \left(\frac{-1}{J_1} + \frac{-1}{J_2} \right) f(\beta) + \frac{T_{in}}{J_1} = -\frac{1}{\bar{J}} f(\beta) + \frac{T_{in}}{J_1} \qquad \bar{J} = \frac{J_1 J_2}{J_1 + J_1} \qquad \beta = \theta_1 - \theta_2$$

Backlash Model



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

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

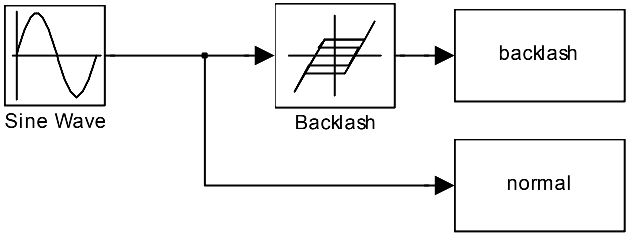
- MatLab Simulation of Backlash
  - The backlash block implements a system in which a change in input causes an equal change in output. However, when the input changes direction, an initial change in input has no effect on the output. The amount of side-to-side play in the system is referred to as the 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.

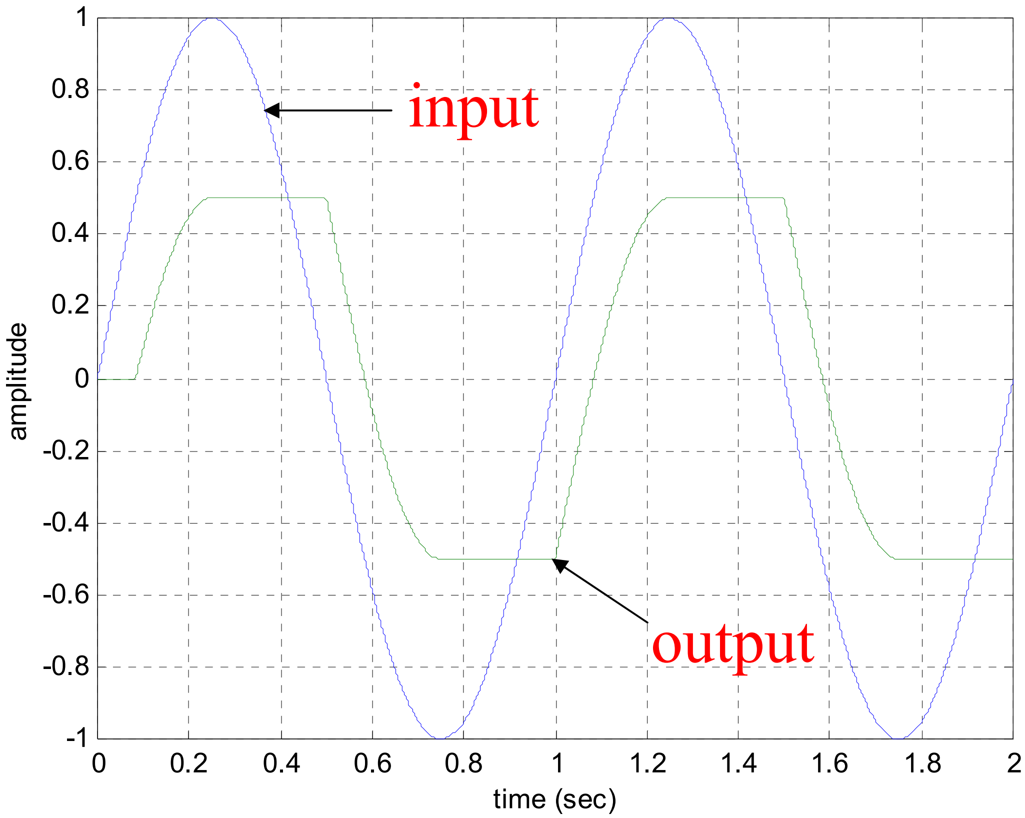
- Engaged in a negative direction: In this mode, the input is decreasing (has a negative slope) and the output is equal to the input plus half the deadband width.

– 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

Sensors & Actuators in Mechatronics

MEAE 6960
Summer 2002

Assignment # 3

- **Problem # 1**
  - A load is driven at constant power under steady-state operating conditions using a separately-excited DC motor with constant supply voltages to the field and armature windings. Show that, in theory, two operating points are possible. Also show that one of the operating points is stable and the other one is unstable.

- Problem # 2
  - Consider the shunt-connected DC motor and the permanent-magnet DC motor. Derive the equations for the steady-state torque-speed characteristics. Sketch the corresponding characteristic curves and discuss the behavior of the motors.

- Problem # 3
  - In this problem you will examine the starting characteristics of the permanent-magnet DC motor as well as its dynamic response to a load change. The motor is rated at 6V, has a no-load speed of approximately 3350 rpm, and the no-load armature current is approximately 0.15 A. The motor has the following parameters:
    - $r_A = 7\ \Omega$
    - $L_{AA} = 120$ mH
    - $k_T = 2$ oz-in/A
    - $J = 150\ \mu$oz-in-s^2

- With no load on the motor ($T_L = 0$) and the motor at rest, a step input in armature voltage of 6V is applied. Plot the armature current and rotor speed vs. time.
- With the motor operating at the no-load condition, a step increase in load torque, T_L, of 0.5 oz-in is applied. Plot the armature current and rotor speed vs. time.
- Use MatLab/Simulink to predict the behavior and generate the required plots. Discuss your results.

Hydraulic & Pneumatic Actuators

- References
- Introduction
- Fluid System Fundamentals
- Electrohydraulic Valve-Controlled Servomechanism Case Study
- Pneumatic System Closed-Loop, Computer-Controlled Positioning Experiment and Case Study

References

- *Control of Fluid Power*, D. McCloy & H.R. Martin, 2nd Edition, John Wiley, New York, 1980.
- *Hydraulic Control Systems*, H. Merritt, John Wiley, New York, 1967.
- *System Dynamics*, E. Doebelin, Marcel Dekker, New York, 1998.
- *Modeling and Simulation of Dynamic Systems*, R. Woods & K. Lawrence, Prentice Hall, New Jersey, 1997.

- *Introduction to Fluid Mechanics*, R. Fox & A. McDonald, John Wiley, New York, 1985.
- *Control System Principles & Design*, E. Doebelin, John Wiley, New York, 1995.
- *Fluid Power Control*, J.R. Blackburn, G. Reethof, and J.L. Shearer, The MIT Press, Cambridge, MA, 1960.
- *The Analysis and Design of Pneumatic Systems*, B. Andersen, John Wiley, New York, 1967.

Introduction

- Applications of Hydraulic and Pneumatic Actuators
- Hydraulic / Pneumatic Systems vs. Electromechanical Systems
- Hydraulic Systems vs. Pneumatic Systems

Applications of Hydraulic & Pneumatic Actuators

- *Hydraulic and Pneumatic Control System* components include pumps, pressure regulators, control valves, actuators, and servo-controls.
- *Industrial Applications* include automation, logic and sequence control, holding fixtures, and high-power motion control.
- *Automotive Applications* include power steering, power brakes, hydraulic brakes, and ventilation controls.
- *Aerospace Applications* include flight-control systems, steering-control systems, air conditioning, and brake-control systems.

Hydraulic / Pneumatic Systems vs. Electromechanical Systems

- Power Density Capability
  - Electromagnetic motors, generators, and actuators are limited by magnetic field saturation and can produce up to about 200 pounds per square inch of actuator.
  - In hydraulic systems, 3000 to 8000 pounds per square inch of actuator is common in aircraft applications and 1000 pounds per square inch is common in industrial applications.
  - Hydraulic systems, both actuators and generators, benefit from a greater ratio of force per unit volume.

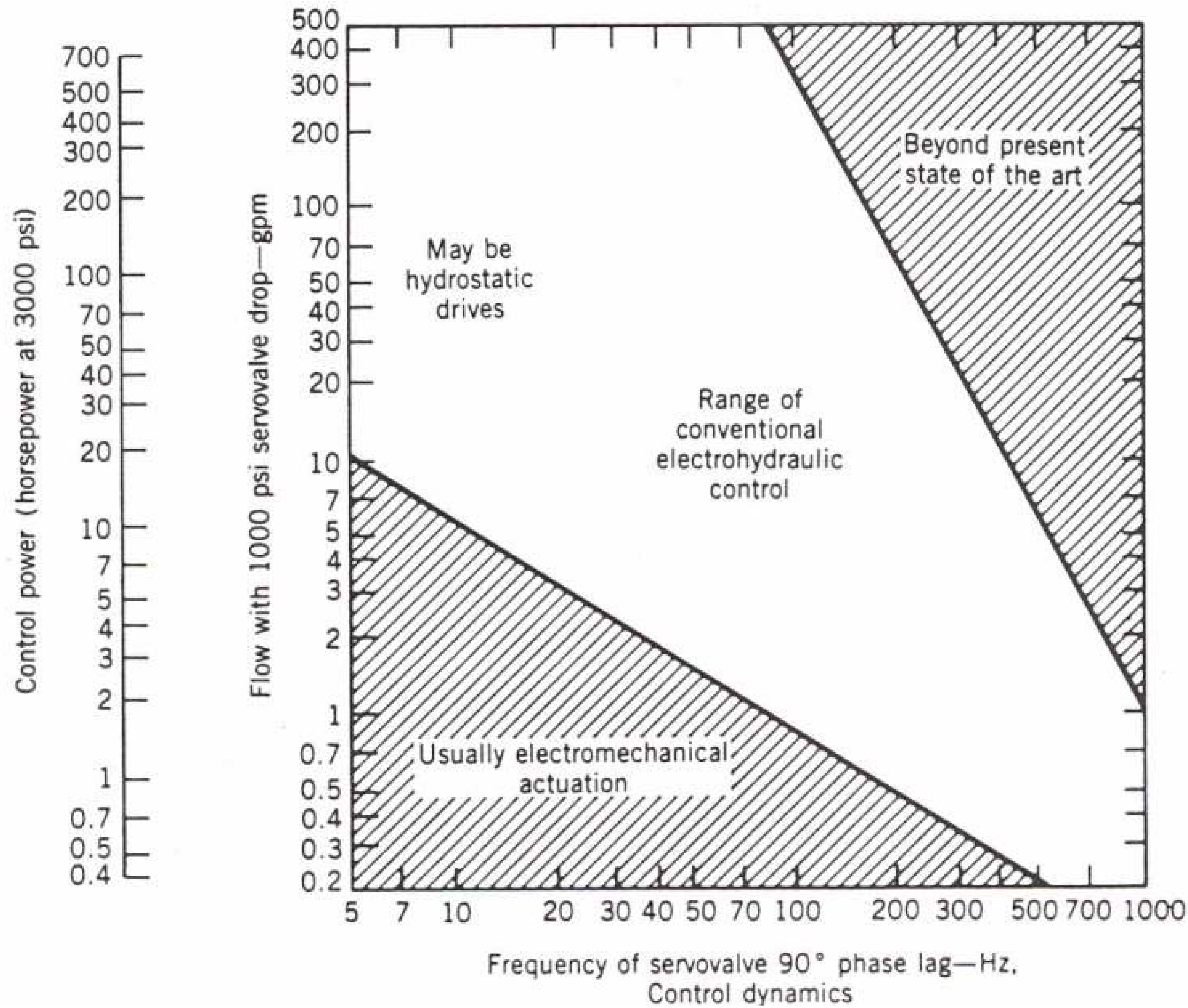
- Responsiveness and Bandwidth of Operation
  - Electromagnetic actuators have a large inertia associated with their motion, so they cannot accelerate quickly.
  - Hydraulic and pneumatic systems are more responsive and have a greater bandwidth of operation at the same power output levels.
- Heat Dissipation
  - Fluid circulating to and from an actuator removes heat generated by the actuator that is doing work.
  - Electromechanical actuators and motors have limited ability to dissipate heat generated inside the device and rely on free or forced convection to the surrounding environment.

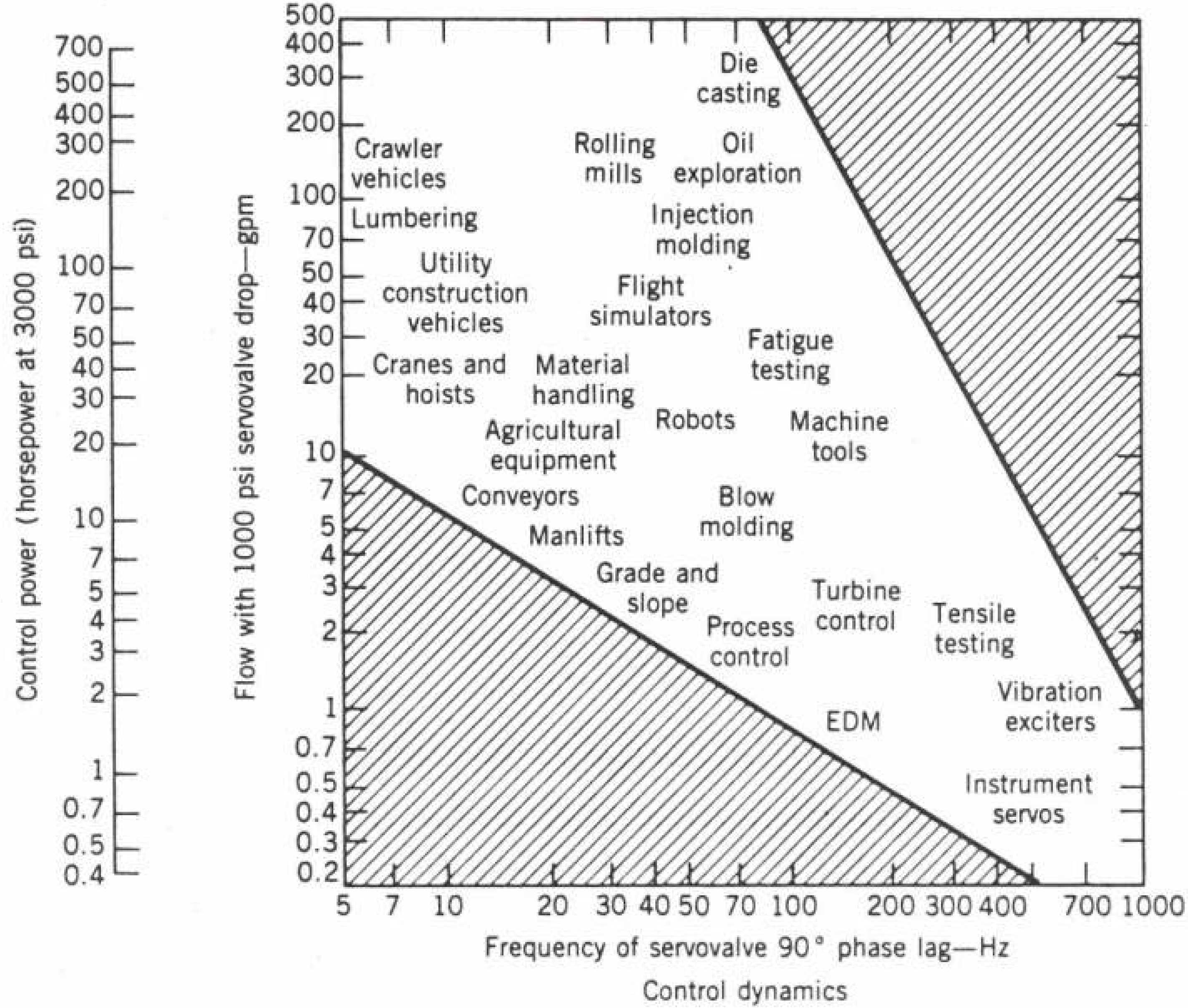
- Heat is the predominant damaging mechanism in electric and electronic systems.
- Reliability of electromagnetic devices is limited compared to that of hydraulic and pneumatic devices.

- <u>Modeling and Simulation</u>
  - Hydraulic and pneumatic systems generally have more significant nonlinearities than do electric or mechanical systems.
- <u>Miscellaneous</u>
  - Electric power is more readily available, cleaner and quieter, and easier to transmit, but may create electrical interference with low-level data signals and can cause overheating problems at low speeds.

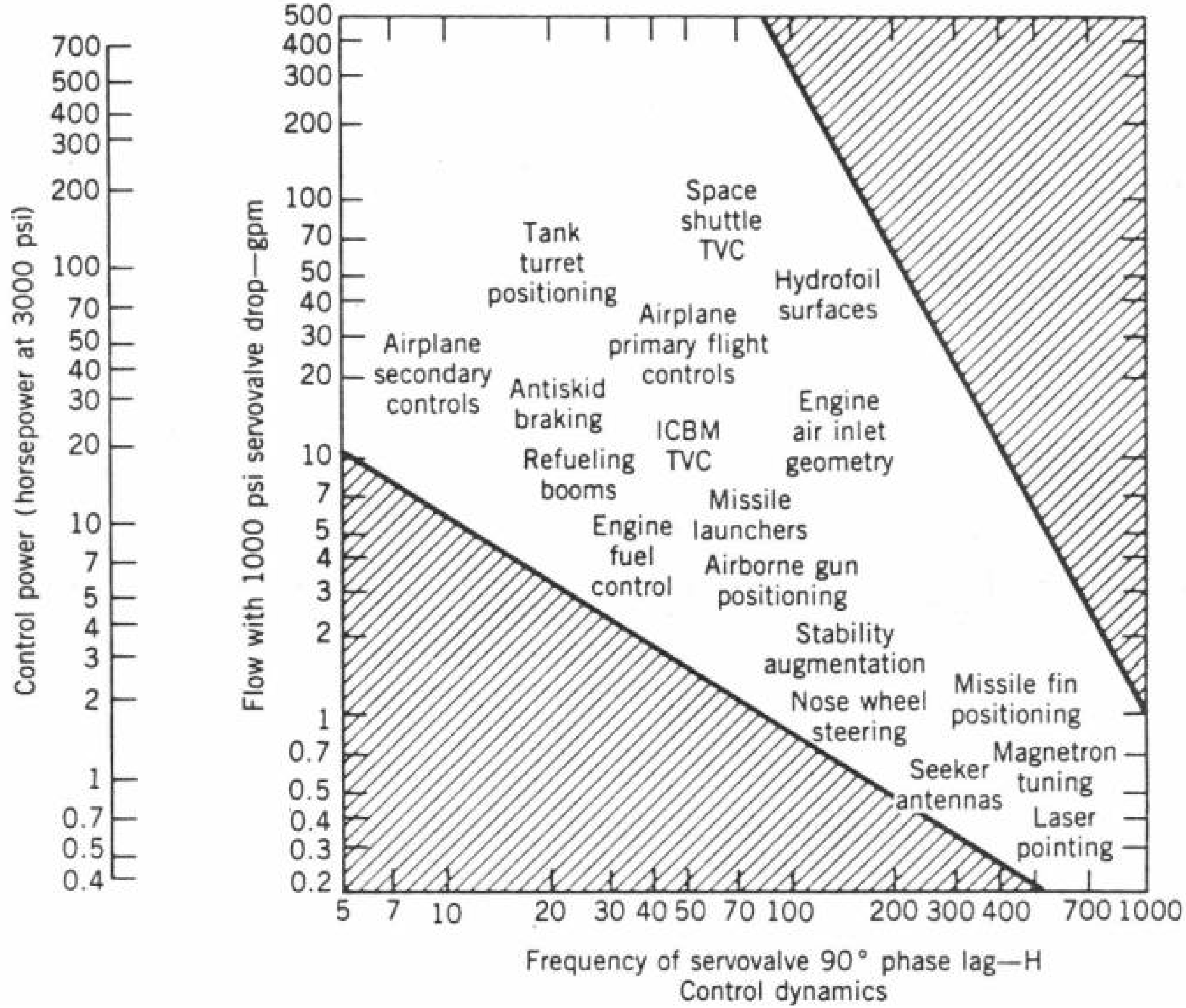
– Hydraulic drives tend to be stiffer with respect to load disturbances; intermittent and stalled operation without damage is possible; but oil leakage, flammability, and fluid contamination may pose problems.

When both high speed and high power are required, hydraulic systems my be mandatory.





Control power (horsepower at 3000 psi)
700
500
400
300
200
100
70
50
40
30
20
10
7
5
4
3
2
1
0.7
0.5
0.4
Flow with 1000 psi servovalve drop—gpm
500
400
300
200
100
70
50
40
30
20
10
7
5
4
3
2
1
0.7
0.5
0.4
0.3
0.2
Die casting
Crawler vehicles
Rolling mills
Oil exploration
Lumbering
Injection molding
Utility construction vehicles
Flight simulators
Fatigue testing
Cranes and hoists
Material handling
Robots
Machine tools
Agricultural equipment
Conveyors
Blow molding
Manlifts
Grade and slope
Turbine control
Tensile testing
Process control
Vibration exciters
EDM
Instrument servos
5
7
10
20
30
40
50
70
100
200
300
400
700
1000
Frequency of servovalve 90° phase lag—Hz
Control dynamics


Control power (horsepower at 3000 psi)
700
500
400
300
200
100
70
50
40
30
20
10
7
5
4
3
2
1
0.7
0.5
0.4
Flow with 1000 psi servovalve drop—gpm
500
400
300
200
100
70
50
40
30
20
10
7
5
4
3
2
1
0.7
0.5
0.4
0.3
0.2
Space shuttle TVC
Tank turret positioning
Hydrofoil surfaces
Airplane primary flight controls
Airplane secondary controls
Antiskid braking
Engine air inlet geometry
ICBM TVC
Refueling booms
Missile launchers
Engine fuel control
Airborne gun positioning
Stability augmentation
Missile fin positioning
Nose wheel steering
Seeker antennas
Magnetron tuning
Laser pointing
5
7
10
20
30
40
50
70
100
200
300
400
700
1000
Frequency of servovalve 90° phase lag—H
Control dynamics

Hydraulic Systems vs. Pneumatic Systems

- Compressibility of Fluid
  - Speed of response: hydraulic systems have a rapid initial response while pneumatic systems exhibit a time delay.
  - Stiffness to external load disturbances: pneumatic systems exhibit a lack of stiffness, especially to external load disturbances.
- Efficiency
  - The efficiency of pneumatic systems is much lower than that of hydraulic systems. Losses are caused by:
    - Cooling of the fluid which dissipates energy in the form of heat.

- Use of positive-displacement actuators causes loss of energy by expansion.
- Leakage – it is difficult to achieve satisfactory sealing. This also puts a limit on system pressure.
- Lubrication – The poor lubricating properties of gases compared to oils also leads to energy losses.

- High-Temperature Applications
  - The pneumatic system can work above 500°C where both electrical and hydraulic systems would fail.
- Mathematical Model Complexity
  - The basic equations of pneumatic systems are nonlinear and tend to be more complex than those of incompressible flow systems.

Fluid System Fundamentals

- Classification of Materials
- Fundamental Concepts & Properties of Fluids
- Basic Equations of Fluid Dynamics
- Lumped-Parameter Approach
- Passive Elements
  - Fluid Resistance
  - Fluid Capacitance
  - Fluid Inertance
- Lumped vs. Distributed Fluid System Models

- Fluid Impedance
- Fluid Sources: Pressure and Flow Rate

Classification of Materials

- Materials may be classified *rheologically* with respect to their shear stress – shear strain behavior in simple shear.
- *Rheology* is the science of the deformation and flow of matter.
- In general, a fluid will undergo a continuous deformation without rupture when subjected to a constant anisotropic stress, where as a solid will generally assume a static equilibrium configuration under such conditions.

- An *isotropic* quantity is the same in any direction from a given point in a system. *Anisotropy* implies a dependence on direction or orientation.
- This type of behavior is relative and depends upon the characteristic time required for the material to respond to a change in stress or strain relative to the time scale of observation, as well as the magnitude of the stress or strain.
- What is silly putty? What is granite?
  - Silly putty will fracture cleanly as a solid if subjected to a large suddenly-applied stress, while it will flow freely as a liquid when subjected to a constant stress of low or moderate magnitude.

- Granite, normally considered a solid, will flow measurably in large formations over a period of geologic time under the influence of gravity alone.

- Basic axiom of rheology is that "everything flows."

- Classification of Materials: Idealized Models
  - Spectrum of Material Classification in Simple Shear Deformations

Solid

- Rigid Solid $\gamma = 0$
- Linear Elastic Solid (Hookean) $\tau = G\gamma \quad (G = \text{constant})$
- Nonlinear Elastic Solid $\tau = G(\lambda)\lambda$
- Viscoelastic $\tau = f(\gamma, \dot{\gamma}, t, \ldots)$

Fluid

- Nonlinear Viscous Fluid (Non-Newtonian) $\tau = \eta(\dot{\gamma})\dot{\gamma}$
- Linear Viscous Fluid (Newtonian) $\tau = \mu\dot{\gamma} \quad (\mu = \text{constant})$
- Inviscid Fluid $\tau = 0$

$\gamma = \text{shear strain}$

$\dot{\gamma} = \text{shear rate}$

$\tau = \text{shear stress}$

Fundamental Concepts & Properties of Fluids

- Basic Concepts
- Definition of a Fluid: Liquids and Gases
- Density
- Equation of State: Liquids and Gases
- Viscosity
- Propagation Speed
- Thermal Properties
- Reynolds Number Effects
- Classification of Fluid Motions

Basic Concepts

- Continuum
  - Fluid is a continuum, an infinitely-divisible substance. As a consequence, each fluid property is assumed to have a definite value at each point in space. Fluid properties are considered to be functions of position and time, e.g., density scalar field $\rho = \rho\,(x, y, z, t)$ and velocity vector field $\vec{v} = \vec{v}(x, y, z, t)$.
- Velocity Field $\vec{v} = \vec{v}(x, y, z, t)$
  - Steady flow – all properties remain constant with time at each point.

$$\frac{\partial \eta}{\partial t} = 0 \quad \eta \text{ is any fluid property}$$

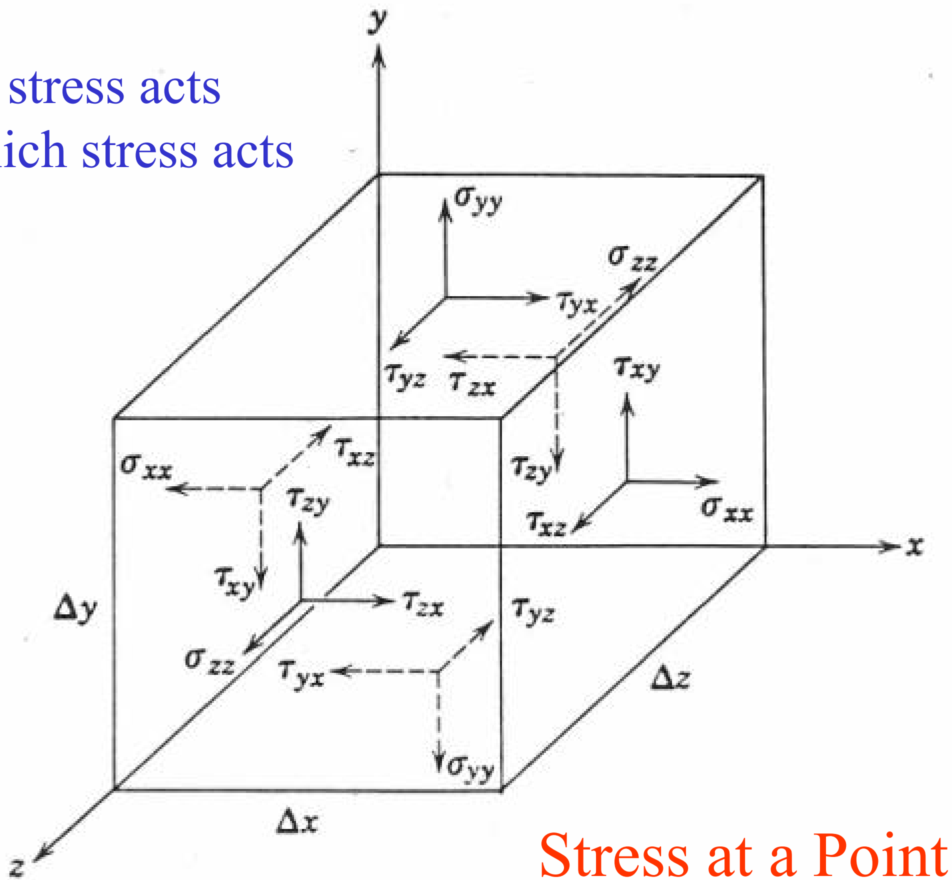
- One-, two-, three-dimensional flows – depends on the number of space coordinates required to specify the velocity field.
- Uniform flow – velocity is constant across any cross section normal to the flow. Other properties may be assumed uniform at a section.
- Timelines, pathlines, streaklines, and streamlines provide visual representation of a flow. In steady flow, pathlines, streaklines, and streamlines coincide.
- All fluids satisfying the continuum assumption must have zero relative velocity at a solid surface (no-slip condition) and so most flows are inherently two or three dimensional. For many problems in engineering, a one-dimensional analysis is adequate to provide approximate solutions of engineering accuracy.

- Stress Field
  - Surface forces: all forces acting on the boundaries of a medium through direct contact.
  - Body forces: forces developed without physical contact and distributed over the volume of the fluid, e.g., electromagnetic and gravitational forces.
  - Stresses in a medium result from forces acting on some portion of the medium. The concept of stress provides a convenient means to describe the manner in which forces acting on the boundaries of the medium are transmitted through a medium.
  - State of stress at a point can be described completely by specifying the stresses acting on three mutually perpendicular planes through the point.

– A stress component is considered positive when the direction of the stress component and the plane on which it acts are both positive or both negative.

1st subscript: plane on which stress acts
2nd subscript: direction in which stress acts

$$\begin{bmatrix} \sigma_{xx} & \tau_{xy} & \tau_{xz} \\ \tau_{yx} & \sigma_{yy} & \tau_{yz} \\ \tau_{zx} & \tau_{zy} & \sigma_{zz} \end{bmatrix}$$



Notation for Stress

Stress at a Point

Definition of a Fluid

- A fluid is a substance (a continuum) that deforms *continuously* under the application of a shear stress no matter how small the shear stress may be and includes both liquids and gases. A fluid cannot sustain a shear stress when at rest.
- Liquids are nearly incompressible.
- Gases are highly compressible.
- Liquids are distinguished from gases by orders of magnitude differences in their density, absolute viscosity, and bulk modulus.

Density

- The density ρ of a fluid is defined as the mass m per unit volume V under specified conditions of pressure P_0 and temperature T_0:

$$\rho = \left.\frac{m}{V}\right|_{P_0, T_0}$$

Equation of State: Liquids

- An equation of state is used to relate the density, pressure, and temperature of a fluid.
- A relation derived from a Taylor series expansion is valid over limited ranges of pressure and temperature:

$$\rho(P,T) = \rho\left(P_0, T_0\right) + \left.\frac{\partial \rho}{\partial P}\right|_{P_0, T_0} \left(P - P_0\right) + \left.\frac{\partial \rho}{\partial T}\right|_{P_0, T_0} \left(T - T_0\right)$$

$$= \rho_0 \left[1 + \frac{1}{\beta}\left(P - P_0\right) - \alpha\left(T - T_0\right) \right]$$

- $\beta \Rightarrow$ Bulk Modulus (inverse of compressibility)

$$\beta = \rho_0 \left.\frac{\partial P}{\partial \rho}\right|_{P_0, T_0} = \left.\frac{\partial P}{\partial \rho / \rho_0}\right|_{P_0, T_0}$$

  – The bulk modulus can be measured by noting the change in pressure with a fractional change in volume of a fixed mass of fluid:

$$\beta = -\frac{\partial P}{\partial V / V_0}$$

  – Isothermal Bulk Modulus (or merely bulk modulus) β can be used when the pressure changes occur at slow enough rates during heat transfer to maintain constant temperature.

- Adiabatic Bulk Modulus β_a can be used when the rate of pressure change is rapid enough to prevent significant heat transfer.
- C_p/C_v, the ratio of specific heats, is only slightly greater than 1.0 for liquids.

$$\beta = -\frac{\partial P}{\partial V / V_0} \qquad \beta_a = \frac{C_p}{C_v}\beta$$

- $\alpha \Rightarrow$ Thermal Expansion Coefficient

$$\alpha = -\frac{1}{\rho_0}\frac{\partial \rho}{\partial T}\bigg|_{P_0, T_0}$$

  - Thermal Expansion Coefficient α relates the incremental change in volume with changes in temperature and can be expressed in terms of volumes for a fixed mass of fluid:

$$\alpha = \frac{\partial V / V_0}{\partial T}\bigg|_{P_0, T_0}$$

$\alpha = 0.5 \times 10^{-3}/°F$ for most liquids

- The compressibility of power-transmitting fluids is a major factor in determining systems resonant frequency. This in turn puts limitations on the speed of response of the output.
- The actual value of the bulk modulus of oil is strongly dependent on the amount of air present in the form of bubbles. Some level of entrained air is impossible to avoid even in the best circuit designs.
- An additional factor which will further reduce the effective bulk modulus will be the type of pipework used to contain the oil.

Equation of State: Gases

- The equation of state for an ideal gas is

$$\rho = \frac{P}{RT}$$

- P and T are in absolute terms and R is the gas constant. Most gases follow this ideal behavior for considerable ranges of pressures and temperatures.

- A gas undergoing a polytropic process follows the relationship:

$$\frac{P}{\rho^n} = C = \text{ constant}$$

  - n = 1.0 for an isothermal or very slow process
  - n = k (ratio of specific heats) for an adiabatic or very fast process
  - n = 0.0 for an isobaric process
  - n = ∞ for an isovolumetric process
- Bulk Modulus for a gas is:

$$\beta = \rho_0 \left.\frac{\partial P}{\partial \rho}\right|_{P_0, T_0} = \rho_0 \left[nC\rho^{n-1} \right]\Big|_{P_0, T_0} = \rho_0 n \left.\frac{C\rho^n}{\rho}\right|_{P_0, T_0} = nP_0$$

- The bulk modulus of a gas is thus related to the absolute pressure of the gas.
- The bulk modulus of a liquid is on the order of 5000 to 15000 bar (1 bar = 1.0E5 N/m^2) compared to 1 to 10 bar for a gas.

Viscosity

- Absolute Viscosity (or dynamic viscosity) μ (dimensions Ft/L^2 or M/Lt) of a fluid represents the ability of the fluid to support a shear stress τ between a relative velocity *u* of the fluid and a solid boundary.

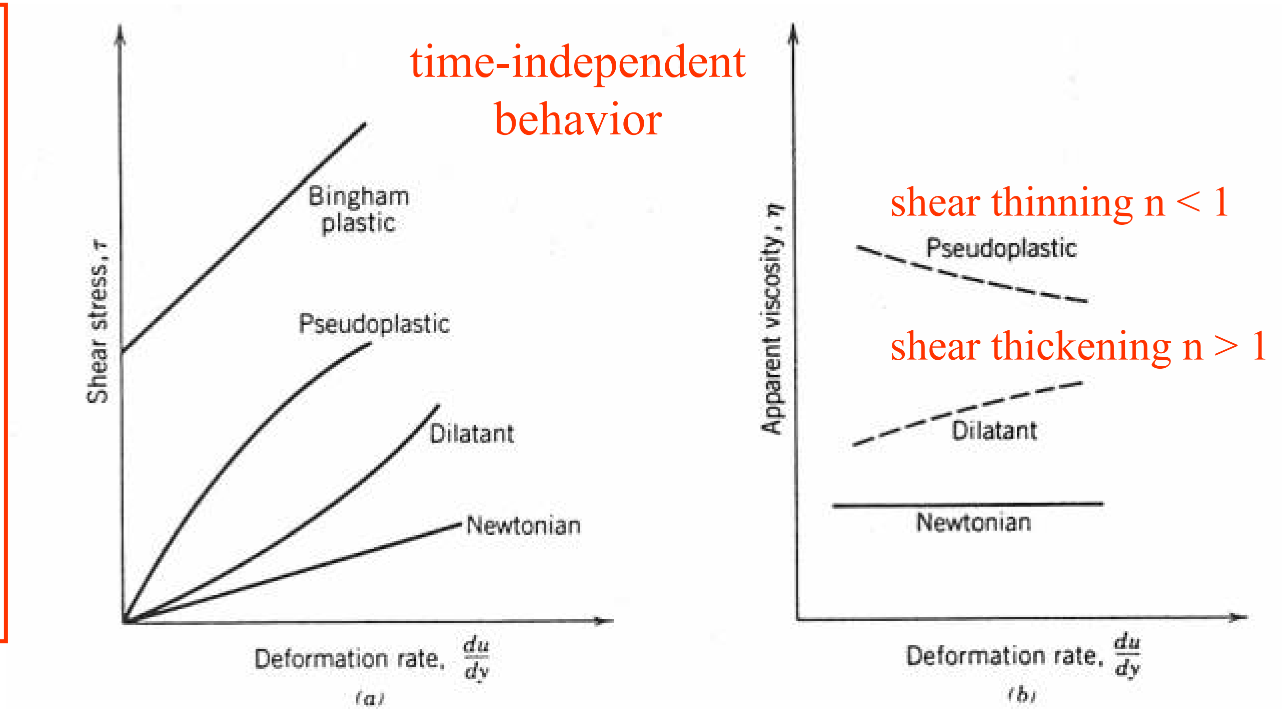
$$\mu = \frac{\tau}{\partial u / \partial y} = \frac{\text{shear stress}}{\text{shear rate}}$$

- A Newtonian fluid is a fluid in which shear stress is directly proportional to shear rate. This is not true for a non-Newtonian fluid.
- Kinematic Viscosity $\nu = \mu/\rho$ (dimensions L^2/t)

- Absolute viscosity of a liquid decreases markedly with temperature: $\mu = \mu_0 e^{-\lambda_L (T - T_0)}$
  - μ_0 and T_0 are values at the reference conditions.
  - λ_L is a constant that depends upon the liquid.
  - In a liquid, resistance to deformation is primarily controlled by cohesive forces among molecules.
- Absolute viscosity of a gas increases with temperature $\mu = \mu_0 + \lambda_G (T - T_0)$
  - μ_0 and T_0 are values at the reference conditions.
  - λ_G is a constant that depends upon the liquid.
  - In a gas, resistance to deformation is primarily due to the transfer of molecular momentum.

- The absolute viscosity of most gases is almost independent of pressure (from 1 to 30 bar).
- The viscosities of most liquids are not affected by moderate pressures, but large increases have been found at very high pressures.

Shear stress τ and apparent viscosity η as a function of deformation rate for one-dimensional flow of various non-Newtonian fluids



Power-Law Model:

$$\tau_{yx} = k\left(\frac{du}{dy}\right)^n = k\left|\frac{du}{dy}\right|^{n-1}\frac{du}{dy} = \eta\frac{du}{dy}$$

Bingham-Plastic Model:

$$\tau_{yx} = \tau_y + \mu_p\frac{du}{dy}$$

Other Categories:

time-dependent behavior { Thixotropic, Rheopectic

Viscoelastic

Under a constant applied shear stress:
Thixotropic – decrease in η with time
Rheopectic – increase in η with time

Propagation Speed

- Speed of propagation, c_0, of a pressure signal in a fluid depends on the bulk modulus and the density:

$$c_0 = \sqrt{\frac{\beta}{\rho}}$$

- The bulk modulus of a gas being perturbed at a high speed is the ratio of specific heats, k, times the absolute pressure of the gas, i.e., $\beta = kP$. Thus the ratio β/ρ reduces to

$$c_0 = \sqrt{kRT}$$

- The speed of sound in air at 25°C is 347 m/s and in a typical hydraulic fluid at 25°C is 1370 m/s.

Thermal Properties

- Specific heat of a fluid is the amount of heat required to raise the temperature of a unit mass of the fluid by 1 degree.
- The specific heat at constant pressure is C_p.
- The specific heat at constant volume is C_v.
- The specific heat ratio $k = C_p/C_v$.
- For some liquids, the specific heat ratio is approximately 1.04. For gases, k is larger, e.g., $k = 1.4$ for air.

Reynolds Number Effects

- Inertial flow forces $\propto \rho A v^2$

$$F = ma = \rho V a \propto \rho d^3 a$$

$$v = \frac{ds}{dt} \text{ and } a = \frac{dv}{dt} \quad \Rightarrow \quad a = v\frac{dv}{ds}$$

$$F \propto \frac{\rho d^3 v^2}{d} = \rho d^2 v^2 = \rho A v^2$$

- Viscous flow forces $\propto \mu d v$

$$\tau A = \mu \frac{du}{dy} A \propto \mu \frac{v}{d} d^2 = \mu v d$$

  - v is the velocity of the fluid
  - d is a characteristic dimension associated with the physical situation
- Flow conditions are often defined by means of a dimensionless ratio called the Reynolds Number, the ratio of inertial to viscous forces:

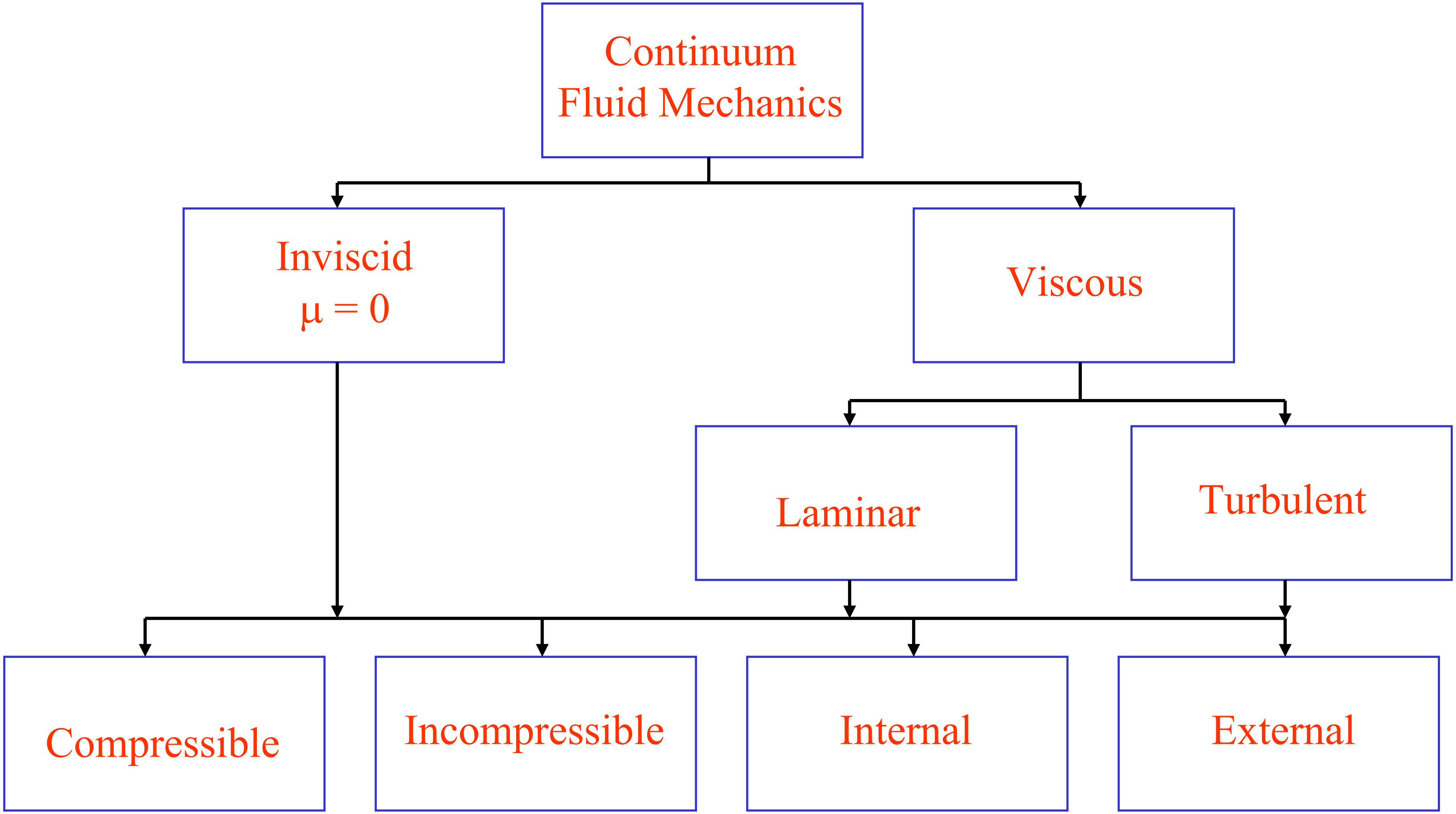
$$N_r = \frac{\rho d^2 v^2}{\mu d v} = \frac{vd}{\nu}$$

- <u>Laminar Flow</u> - smooth, streamlined flow where viscous forces dominate over inertial forces; no macroscopic mixing of adjacent fluid layers.
- <u>Turbulent Flow</u> - swirling flow where inertial forces dominate over viscous forces; macroscopic mixing of adjacent fluid layers.
- At small N_r (< 1400) flow is laminar.
- At high N_r (> 3000) flow is turbulent and has energy loss due to fluid collisions and mixing.
- For $1400 < N_r < 3000$, there is a transition from laminar to turbulent flow. Here flow depends on the local conditions and on the previous flow history.

- Laminar flow: pressure loss due to friction has a first-order relationship with the flow (analogous to electrical resistance in which voltage is linear with current).
- Turbulent flow: pressure loss becomes proportional to the square of the flow. There is no universal relationship between the stress field and the mean velocity field; one must rely heavily on semi-empirical theories and on experimental data.
- Systems with laminar flow are far simpler than those with turbulent flow; however, laminar flow is generally impractical for most systems because of the low pressures and small dimensions it requires.

Classification of Fluid Motions

Basis of observable physical characteristics of flow fields



- Incompressible Flow – density variations are negligible. Most liquid flows are essentially incompressible. However water hammer and cavitation are examples of the importance of compressibility effects in liquid flows. Gas flows may be considered incompressible provided the flow speeds are small relative to the speed of sound, e.g., < 30% of the speed of sound, which corresponds to 100 m/s for air at standard conditions.
- Compressible Flow – density variations are not negligible.

Basic Equations of Fluid Dynamics

- The basis laws governing fluid motion are:
  - Conservation of Mass
  - Newton's 2nd Law
  - Moment of Momentum
  - 1st Law of Thermodynamics
  - 2nd Law of Thermodynamics

  Not all are required to solve any one problem.
- In addition one needs the equations of state or constitutive equations that describe the behavior of physical properties of fluids under given conditions.

- Methods of Analysis
  - A system or control volume must be defined before applying the basic equations in the analysis of a problem.
  - A *system* is defined as a fixed, identifiable quantity of mass; system boundaries separate the system from the surroundings and they may be fixed or movable, however, there is no mass transfer across the system boundaries. Heat and work may cross the system boundaries.
  - A *control volume* is an arbitrary volume in space through which fluid flows. The geometric boundary of the control volume is called the control surface, which may be real or imaginary, at rest or in motion, but it must be clearly defined before beginning analysis.

- The basic laws can be formulated in terms of infinitesimal systems or control volumes, in which case the resulting equations are *differential equations*, whose solutions provide detailed, point-by-point behavior of the flow.
- When we are interested in the gross behavior of a device, we use finite systems or control volumes. The *integral formulation* of the basic laws is more appropriate.
- The method of analysis depends on the problem!
- The method of description that follows a fluid particle is called the *Lagrangian* method of description.

- The field, or *Eulerian*, method of description focuses attention on the properties of the flow at a given point in space as a function of time. The properties of the flow field are described as functions of space coordinates and time. This method of description is a logical outgrowth of the fluid-as-a-continuum assumption.
- Two Basic Reasons for using the control volume formulation rather then the system formulation:
    - Since fluid media are capable of continuous deformation, it is often extremely difficult to identify and follow the same mass of fluid at all times, as must be done in applying the system formulation.
    - We are often interested not in the motion of a given mass of fluid, but rather in the effect of the overall fluid motion on some device or structure.

- <u>Conservation of Mass</u>
  - Mass of the system is constant.

$$\left(\frac{dM}{dt}\right)_{sys} = 0$$

$$M_{sys} = \int_{sys-mass} dm = \int_{sys-vol} \rho dV$$

- <u>Newton's 2nd Law</u>
  - For a system moving relative to an inertial reference frame, the sum of all external forces acting on the system is equal to the time rate of change of the linear momentum of the system.

$$\left(\vec{F}\right)_{sys} = \left(\frac{d\vec{P}}{dt}\right)_{sys} \qquad \overrightarrow{P_{sys}} = \int_{sys-mass} \vec{V} dm = \int_{sys-vol} \vec{V} \rho dV$$

- 1st Law of Thermodynamics: Conservation of Energy
  - Rate at which heat is added to the system plus the rate at which work is done on the system by the surroundings is equal to the rate of change of the total energy of the system:

$$\dot{Q} + \dot{W} = \left(\frac{dE}{dt} \right)_{sys}$$

$$E_{sys} = \int_{sys-mass} (e)\, dm = \int_{sys-vol} (e\rho)\, dV$$

$$e = u + \frac{v^2}{2} + gz$$

- Each system equation, written on a rate basis, involves the time derivative of an extensive property of the system. Let N = any arbitrary extensive property of the system. Let η = the corresponding intensive property (extensive property per unit mass).

$$N_{sys} = \int_{sys-mass} \eta \, dm = \int_{sys-vol} \eta \rho \, dV$$

- The equation relating the rate of change of the arbitrary extensive property, N, for a system to the time variations of this property associated with a control volume at the instant when the system and control volume coincide is:

$$\left(\frac{dN}{dt}\right)_{sys} = \frac{\partial}{\partial t}\int_{CV} \eta \rho \, dV + \int_{CS} \eta \rho \vec{v} \bullet \overrightarrow{dA}$$

- This equation says: The time rate of change of the arbitrary extensive property of the system equals the time rate of change of the arbitrary extensive property within the control volume plus the net rate of efflux (rate of outflow minus the rate of inflow) of the arbitrary extensive property through the control surface.
- Note:
  - Velocity is measured relative to the control volume.
  - The time rate of change of the arbitrary extensive property within the control volume must be evaluated by an observer fixed in the control volume.

- Let's apply this equation to the three basic laws:
  - Conservation of Mass
  - Newton's 2nd Law
  - Conservation of Energy

- ## Conservation of Mass
  - The net rate of mass efflux through the control surface plus the rate of change of mass inside the control volume equals zero.

$$0 = \frac{\partial}{\partial t}\int_{CV} \rho \, dV + \int_{CS} \rho \vec{v} \bullet \overrightarrow{dA}$$

  - Velocity is measured relative to the control volume.
  - $\rho \vec{v} \bullet \overrightarrow{dA}$ is a scalar: it is positive where flow is out through the control surface, negative where flow is in through the control surface, and zero where flow is tangent to the control surface.

- <u>Newton's 2nd Law for an Inertial Control Volume</u>
  - The sum of all forces acting on a non-accelerating control volume equals the rate of change of momentum inside the control volume plus the net rate of efflux of momentum through the control surface.

$$\vec{F} = \vec{F}_S + \vec{F}_B = \frac{\partial}{\partial t}\int_{CV} \vec{v}\rho dV + \int_{CS} \vec{v}\rho\vec{v}\bullet\overrightarrow{dA}$$

  - All velocities are measured relative to the control volume.
  - All time derivatives are measured relative to the control volume.

- 1st Law of Thermodynamics: Conservation of Energy

$$\dot{Q} + \dot{W} = \frac{\partial}{\partial t}\int_{CV} e\rho dV + \int_{CS} e\rho \vec{v} \bullet \overrightarrow{dA}$$

$$e = u + \frac{v^2}{2} + gz$$

  - $\dot{W}$ is positive when work is done on the control volume by the surroundings and this can take place only at the control surface.

$$\dot{W} = \dot{W}_{shaft} + \dot{W}_{normal} + \dot{W}_{shear} + \dot{W}_{other}$$

Lumped-Parameter Approach

- Lumped-Parameter Approach
  - Some areas have used this approach more than others, e.g., electrical systems are originally conceived as a combination of R, L, C, op-amps, and other integrated-circuit modules, while mechanical systems are rarely initially conceived in terms of a combination of K, B, and M, but rather as combination of mechanisms, actuators, sensors, and controllers.
  - Fluid and Thermal systems follow a similar pattern in that system dynamics may receive relatively light conscious emphasis during the early conceptual phases.

- In addition, due to the generally less-well-defined shapes of bodies of fluid and the fact that heat flow rarely is confined to such simple and obvious paths as current, these type systems may appear less well suited to the lumped-parameter viewpoint.
- System dynamics can preserve the identity of individual components while comprehending the entire system, and thus often gives insights into needed design changes at both the component and system levels.
- So, we consider system dynamics methods for fluid and thermal machines and processes, even though they initially seem less well suited to these more amorphous systems.
- There are many practical examples of actual hardware which have been successfully studied with this approach.

Passive Fluid Elements

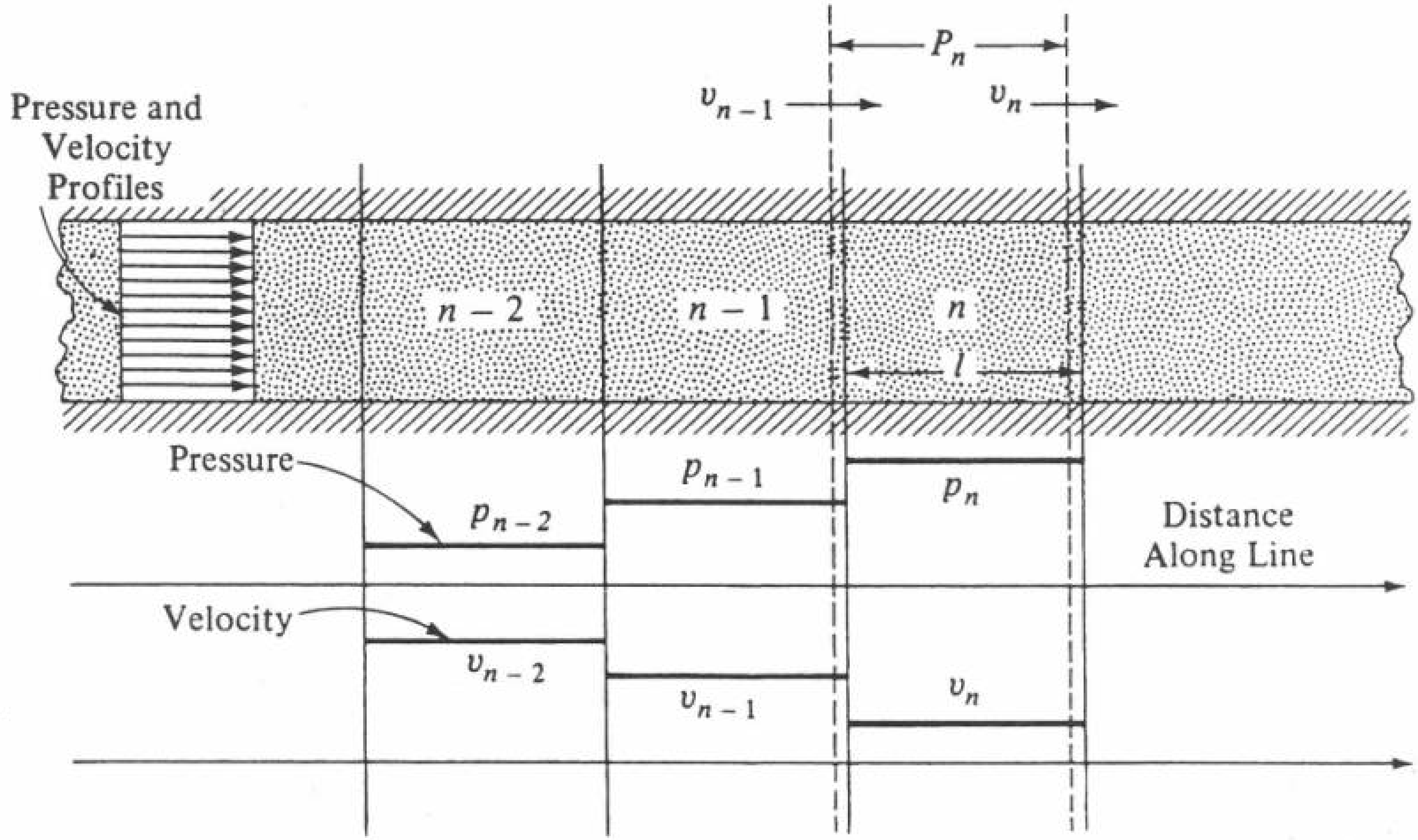
- Fluid Flow Resistance
- Fluid Compliance (Capacitance)
- Fluid Inertance (Inductance)

Fluid Flow Resistance

- Like mechanical friction and electrical resistance, this element performs the energy-dissipation function.
- The dissipation of fluid energy into heat occurs in all fluid devices to some extent.
- The simplest example is that of a fluid pipe.
  - Consider the flow of a fluid in a constant-area, rigid-walled conduit (as shown).
  - Variables of primary interest are: average fluid pressure p and the volume flow rate q.

- The average flow velocity *v* is defined as *q*/*A*, where *A* is the conduit cross-sectional area.
- The product of *p* and *q* is fluid power, just as mechanical power is *fv* and electrical power is *ei*.
- While the actual fluid pressure and velocity vary from point to point over the flow cross section in a real fluid flow, we assume a so-called one-dimensional flow model in which the velocity and pressure are uniform over the area.
- Thus, the average velocity and average pressure correspond numerically with the values at any point in the cross section. These averages are spacewise averages, not timewise averages.

Lumped Model of a Fluid Pipeline



- In a lumped-parameter dynamic analysis, the pipeline is broken into segments. Within each segment or lump, the pressure and velocity may vary arbitrarily with time, but are assumed uniform over the volume of the lump. By considering the behavior of one typical lump (the n^{th} lump) we are led to definitions of all three basic fluid elements.
- Experiments show that when a fluid is forced through a pipe at a constant flow rate, a pressure difference related to that flow rate must be exerted to maintain the flow. It is observed that it takes a larger pressure difference to cause a larger flow rate.
- In general, the relation between pressure drop and flow rate is nonlinear; however, some situations give a nearly linear effect.

- The pure and ideal fluid friction or fluid resistance element is defined as:

$$R_f \triangleq \frac{p_1 - p_2}{q} \left(\frac{\text{psi}}{\text{in}^3 / \text{sec}} \right)$$

- For nonlinear fluid resistances, we can define linearized values in the neighborhood of an operating point.
- Now consider Conservation of Mass applied to the nth lump over an infinitesimal time interval *dt*. During the time interval *dt*, the difference between mass flow into and out of the lump must equal the additional mass stored in the lump. Mass enters the lump from the left at a rate $Av_{n-1}\rho$ and leaves at the right at a rate $Av_n\rho$, where ρ is the fluid mass density, which we treat as being constant, corresponding to a constant operating-point pressure and temperature.

– For constant density, conservation of mass is the same as conservation of volume:

$$\left(Av_{n-1} - Av_{n}\right)dt = dV = \frac{V}{\beta}dp_{n} = \frac{A\ell}{\beta}dp_{n}$$

$$\left(q_{n-1} - q_{n}\right)dt = \frac{A\ell}{\beta}dp_{n}$$

$$p_{n} = \frac{1}{C_{f}}\int\left(q_{n-1} - q_{n}\right)dt$$

$$C_{f} \triangleq \frac{A\ell}{\beta} \triangleq \text{fluid compliance (capacitance)}$$

– We can easily see the electric-fluid analogy: electric current analogous to net volume flow rate, pressure analogous to voltage drop, and electrical capacitance analogous to fluid capacitance.

– Newton's 2nd Law states that the force (pressure times area) difference between the left and right ends of a lump must equal the lump mass times its acceleration. For the nth lump the result is:

$$Ap_{n-1} - Ap_n - AR_f q_n = \rho A \ell \frac{dv_n}{dt} = \rho \ell \frac{dq_n}{dt}$$

$$\left(p_{n-1} - p_n\right) - R_f q_n = \frac{\rho \ell}{A} \frac{dq_n}{dt}$$

– Since this equation contains both resistance (friction) and inertance (inertia) effects, we consider each (in turn) negligible, to separate them. With zero fluid density (no mass) we have:

$$\left(p_{n-1} - p_n\right) = R_f q_n$$

– If the resistance (friction) were zero, we have:

$$\left(p_n - p_{n-1}\right) = \frac{\rho\ell}{A}\frac{dq_n}{dt} = I_f \frac{dq_n}{dt}$$

$$I_f \triangleq \frac{\rho\ell}{A} \triangleq \text{fluid inertance}$$

– Again we see the electric-fluid analogy: pressure drop analogous to voltage drop, volume flow rate analogous to current, and fluid inertance analogous to electric inductance.

– We will return to the fluid compliance and inertance elements in more detail, since they occur in more general contexts, not just pipelines. They were introduced here to illustrate that the three elements are always present whenever a body of fluid exists.

– Whether all three will be included in a specific system model depends on the application and the judgment of the modeler.

– Let's consider fluid resistance in a more general way. When a one-dimensional fluid flow is steady (velocity and pressure at any point not changing with time), the inertance and compliance cannot manifest themselves (even though they exist), and only the resistance effect remains. We can thus experimentally determine fluid resistance by steady-flow measurements of volume flow rate and pressure drop, or if we attempt to calculate fluid resistance from theory, we must analyze a steady-flow situation and find the relation between pressure drop and volume flow rate. If a nonlinear resistance operates near a steady flow q_0, we can define a linearized resistance, good for small flow and pressure excursions from that operating point.

- Just as in electrical resistors, a fluid resistor dissipates into heat all the fluid power supplied to it. Fluid power at a flow cross section is the rate at which work is done by the pressure force at that cross section:

$$\text{Power} = pAv = pq$$

- For our assumed one-dimensional incompressible flow (volume flow rate same at both sections), the power dissipated into heat is:

$$q\left(p_1 - p_2\right) = q\Delta p = q^2 R_f = \frac{\Delta p^2}{R_f}$$

- While we can often determine flow resistance values by experimental steady-flow calibration, it is desirable to be able to calculate from theory, before a device has been built, what its resistance will be. For certain simple configurations and flow conditions, this can be done with fairly good accuracy.

- Refer to the previous discussion on Laminar and Turbulent flow and the Reynolds Number.
- Laminar flow conditions produce the most nearly linear flow resistances and these can be calculated from theory, for passages of simple geometrical shape.
- For example, a long, thin flow passage called a capillary tube of circular cross section has a fluid resistance:

$$q = \frac{\pi D^4}{128 \mu L} \Delta p$$

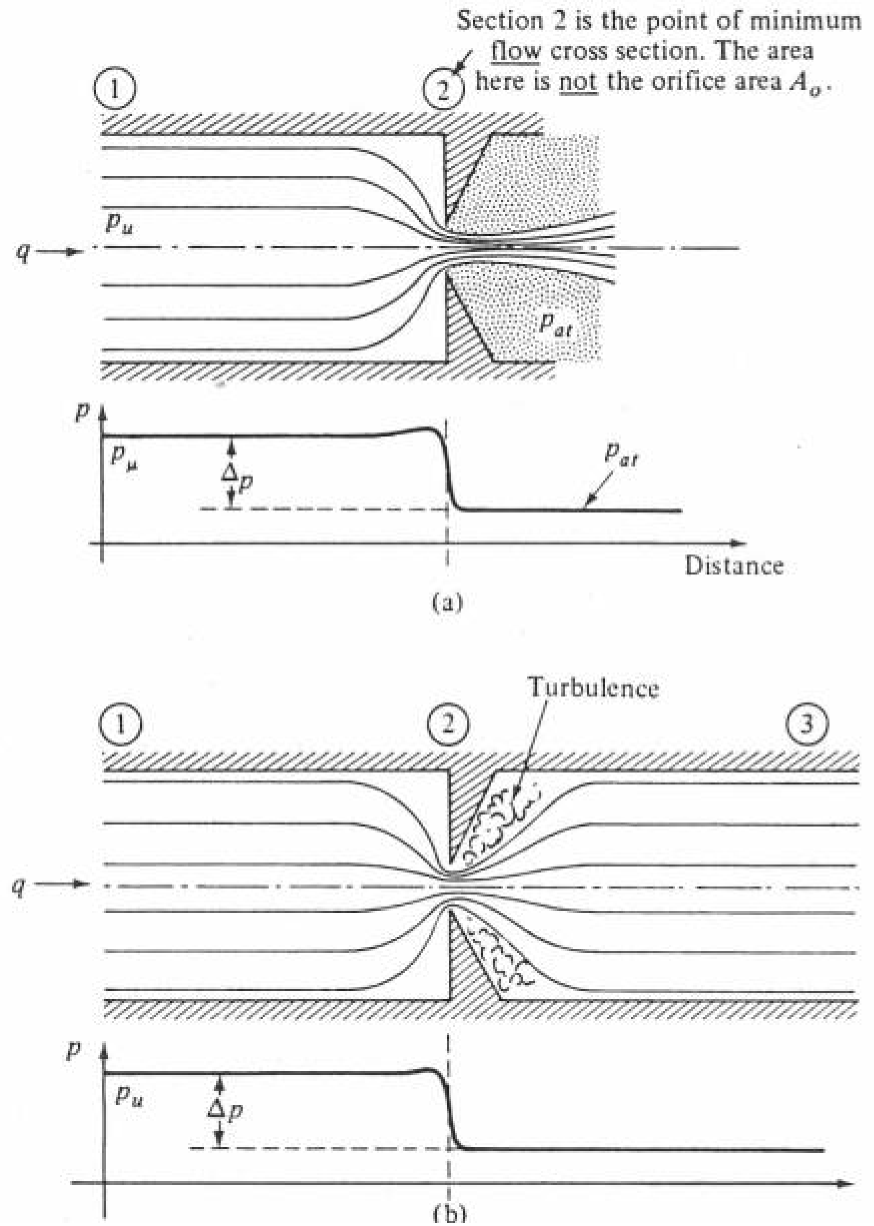
$$R_f \triangleq \frac{\Delta p}{q} = \frac{128 \mu L}{\pi D^4}$$

- This is valid for laminar flow with end effects neglected.

- Note that in the above calculations and experimental measurements of flow resistances, the approach has been to use formulas relating flow rate and pressure for steady flows as if they held for general (unsteady) flows. This approach is widely used, and usually of sufficient accuracy; however, it should be recognized as an approximation.
- Now consider orifices, where resistance is concentrated in a short distance. The pressure drop across an orifice is basically due to a conversion from the form of fluid power, pressure times volume flow rate, to the power of kinetic energy. From conservation of energy, for a level flow of a frictionless incompressible fluid:

pressure/flow power + kinetic energy power = constant

Characteristics of Orifice Flow



– Considering any two locations 1 and 2, we may write:

$$p_1 q + (\text{KE per unit time})_1 = p_2 q + (\text{KE per unit time})_2$$

$$p_1 q + \frac{\frac{\rho A L v_1^2}{2}}{\frac{AL}{q}} = p_2 q + \frac{\frac{\rho A L v_2^2}{2}}{\frac{AL}{q}} \Rightarrow p_1 - p_2 = \Delta p = \frac{\rho}{2}\left(v_2^2 - v_1^2\right)$$

– Since $q = A_1 v_1 = A_2 v_2$,

$$\frac{2\Delta p}{\rho} = \left(\frac{q}{A_2}\right)^2 - \left(\frac{q}{A_1}\right)^2 = \frac{1-\left(A_2/A_1\right)^2}{A_2^2} q^2$$

$$q = \frac{A_2}{\sqrt{1-\left(A_2/A_1\right)^2}} \sqrt{\frac{2\Delta p}{\rho}}$$

Basic pressure/flow relation (nonlinear) for an orifice

- Because of turbulence in the flow, viscosity of real fluids, and difficulties in measuring the areas of the fluid stream, the practical formula used to predict the pressure/flow relation for orifices in pipelines uses experimental data, defines the pressure drop as that between stations 1 and 3, and uses only the cross-sectional area of the hole in the metal orifice:

$$q = C_d A_o \sqrt{\frac{2\Delta p}{\rho}}$$

- The value of the orifice discharge coefficient C_d depends mainly on the Reynolds number and the area ratio $A_{pipe} / A_{orifice}$.

– The fluid resistances discussed so far are all intended to be essentially constant in numerical form. Many fluid systems require adjustable resistances, and often these take the form of some kind of valve used to smoothly modulate the flow rate. The vast majority of valves have a square-root type of pressure/flow relation and usually require experimental calibration if we want a reasonably accurate flow model. The flow modulation is achieved by somehow varying the "orifice" area A_o. A complete calibration would give a family of flow rate versus pressure drop curves, one for each flow area.

Fluid Compliance

- We have already seen that a fluid itself, whether a liquid or gas, exhibits compliance due to its compressibility.
- Certain devices may also introduce compliance into a fluid system, even if the fluid were absolutely incompressible.
  - Rubber hoses will expand when fluid pressure increases, allowing an increase in volume of liquid stored.
  - Accumulators use spring-loaded cylinders or rubber air bags to provide intentionally large amounts of compliance.

– A simple open tank exhibits compliance, since an increase in volume of contained liquid results in a pressure increase due to gravity.

- In general, the compliance of a device is found by forcing into it a quantity of fluid and noting the corresponding rise in pressure. For liquids, the input quantity is a volume of fluid V, and the ideal compliance is:

$$\text{Fluid Compliance} \triangleq C_f \triangleq \frac{V}{p} \triangleq \frac{\int q\,dt}{p}$$

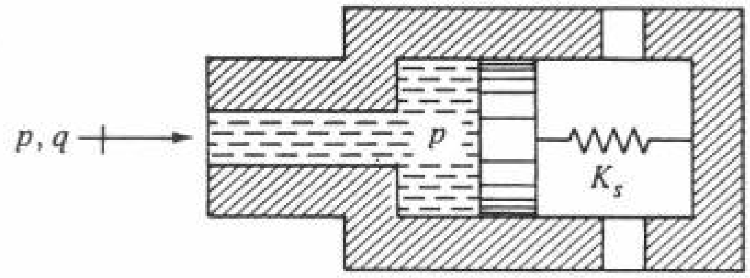
– For nonlinear compliances, the actual p-V curve can be implemented in a computer lookup table, or the local slope can be used to define an incremental compliance.

– We previously calculated the compliance of a section of hydraulic line, due to bulk modulus of the liquid itself, as: $\dfrac{A\ell}{\beta}$

– Additional contributions to compliance which may be significant here are due to entrained air bubbles and the flexibility of the tubing.

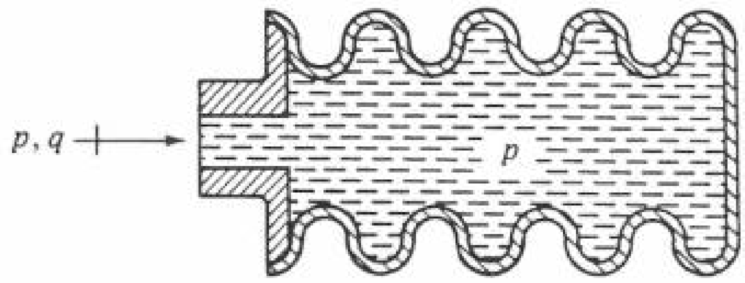
– Accumulators are devices intentionally designed to exhibit fluid compliance, e.g., spring-loaded piston and cylinder; flexible metal bellows; nitrogen-filled rubber bag. Some devices can store large amounts of fluid energy and are widely used for short-term power supplies, pulsation smoothing, and to reduce pump size in systems with intermittent flow requirements.

Accumulators

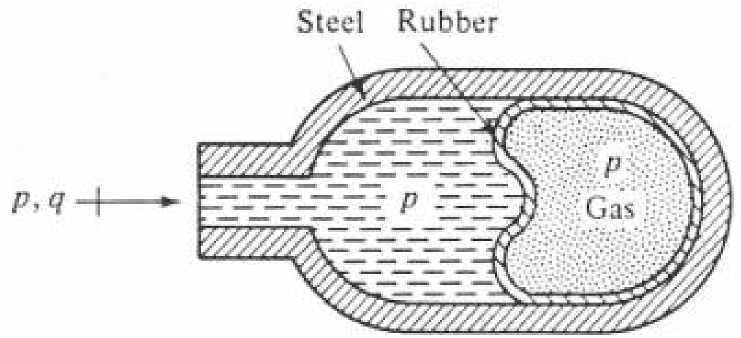


(a)

- (a) spring-loaded piston and cylinder
- (b) flexible metal bellows
- (c) nitrogen-filled rubber bag



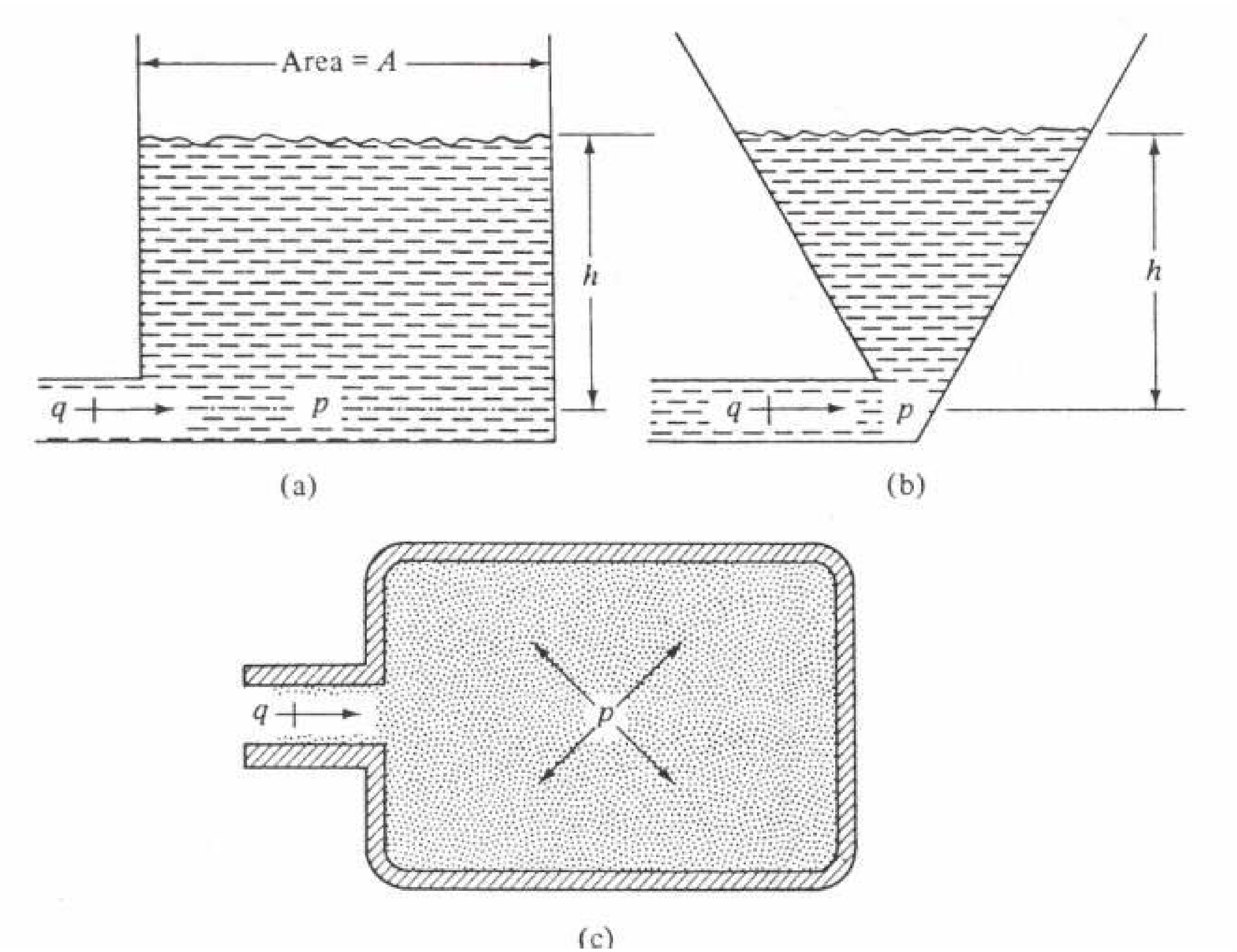
(b)



(c)

- Consider a vertical cylindrical tank of cross-sectional area A supplied with a volume flow rate q; the pressure at the tank inlet is p, liquid height is h. The vertical motion of the liquid in such tanks is usually slow enough that the velocity and acceleration have negligible effects on the pressure p and it is simply given by $p = \gamma h$, where γ is the specific weight of the liquid.
- Think of the tank as a large-diameter vertical pipe and R_f (velocity effect) and I_f (acceleration effect) are negligible relative to the height effect.
- If we add a volume V of liquid to the tank, the level h goes up an amount V/A and the pressure rises an amount $V\gamma/A$.

Liquid and Gas Tanks as Fluid Compliances



– The tank compliance is thus:

$$C_f = \frac{\text{volume change}}{\text{pressure change}} = \frac{V}{V\gamma / A} = \frac{A}{\gamma}\left(\frac{\text{in}^3}{\text{psi}}\right)$$

– For non-cylindrical tanks the compliance effect is nonlinear, but can be linearized in the usual way if desired.

– Consider a rigid tank of volume V containing a gas at pressure p. For slow (isothermal) pressure changes in which fluid density is nearly constant, we may write $pV = MRT$. If we force a mass $dM = \rho dV$ of gas into the tank we cause a pressure change dp given by:

$$dp = \frac{RT}{V} dM = \frac{RT}{V} \rho dV = \frac{RT}{V} dV \frac{p}{RT}$$

$$C_f \triangleq \left[\frac{dV}{dp} \right]_{p=p_0} = \frac{V}{p_0} \left(\frac{in^3}{psi} \right)$$

– This is a linearized compliance useful for small changes near an operating point p_0. For rapid (adiabatic) but still small pressure changes, analysis shows the compliance is:

$$C_f = \frac{V}{kp_0} \quad \text{where k is the ratio of specific heats}$$

- Fluid Capacitance relates how fluid energy can be stored by virtue of pressure.
- Law of Conservation of Mass (Continuity Equation) for a control volume:

$$\dot{m}_{net} = \frac{d}{dt}\left(M_{cv}\right) = \frac{d}{dt}\left(\rho_{cv} V_{cv}\right)$$

$$\dot{m}_{net} = \rho Q_{net} = \rho_{cv}\dot{V}_{cv} + V_{cv}\dot{\rho}_{cv}$$

- If all densities of the system are equal to ρ, then:

$$Q_{net} = \dot{V} + \frac{V}{\rho}\dot{\rho}$$

- This assumption is justified for incompressible fluids and is quite accurate for compressible fluids if pressure variations are not too large and the temperature of flow into the control volume is almost equal to the temperature of flow out of the control volume.
- Now $\rho = \rho(P, T) \Rightarrow \dot{\rho} = \frac{\partial \rho}{\partial P}\frac{\partial P}{\partial t} = \frac{\rho_0}{\beta}\dot{P}$ $\qquad \beta = \rho_0 \left.\frac{\partial P}{\partial \rho}\right|_{P_0, T_0}$

$$Q = \dot{V} + \frac{V}{\beta}\dot{P}_{cv}$$

- If the container is rigid, then $\dot{V} = 0 \Rightarrow Q = \frac{V}{\beta}\dot{P}_{cv} = C_f \dot{P}_{cv}$
- C_f is the fluid capacitance. Any large volume of a compressible fluid becomes a capacitance.

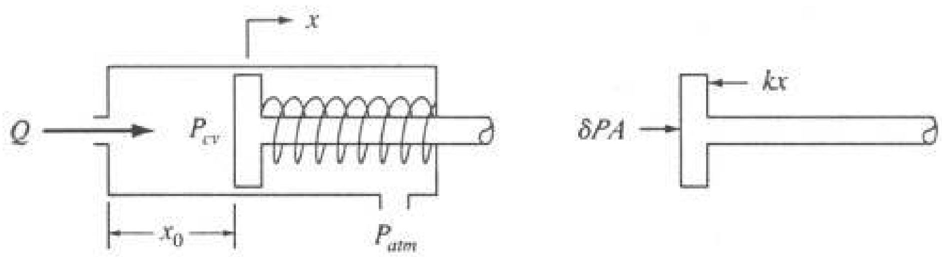
- The continuity equation gives the differential equation for the pressure inside the control volume. The following equation should be used in systems analysis (for systems with an inlet flow with the same density as the control volume), without repeating its derivation:

$$Q = \dot{V} + \frac{V}{\beta}\dot{P}_{cv}$$

$$\dot{P}_{CV} = \frac{\beta}{V}\left(Q - \dot{V}\right)$$

- The term $\dot{V}$ can take several forms depending upon the exact configuration of the system (control volume).

- For Example:



  - A spring-loaded piston of stiffness k and Area A can be used as a capacitor.
  - By applying a force balance equation, neglecting inertial and frictional effects, we obtain:

$$A\left(P_{cv}-P_{atm}\right)-kx=A\delta P-kx=0$$

$$\dot{x}=\frac{A}{k}\delta\dot{P}$$

  - The volume of the cylinder is:

$$V=A\left(x+x_0\right)$$

$$\dot{V}=A\dot{x}$$

  - x_0 is the minimum stroke representing the dead volume space in the actuator and fittings.

– Substitution yields:

$$\dot{V} = \frac{A^2}{k}\delta\dot{P}$$

– The continuity equation can now be written as:

$$Q = \dot{V} + \frac{V}{\beta}\dot{P}_{cv}$$

$$Q = \left(\frac{V}{\beta} + \frac{A^2}{k}\right)\delta\dot{P} = C_f\delta\dot{P}$$

– The total capacitance is a combination of the fluid compressibility effects and the mechanical compliance of the container.

– Thus, fluid capacitance can be either a compliant container or a volume of fluid.

– Increasing A or decreasing k increases the mechanical capacitance term, and increasing V or decreasing β increases the fluid compressibility capacitance.
– Capacitors for liquids are called accumulators. Typical accumulators are spring-loaded pistons, bellows, and gas-filled bladders for hydraulic systems.
– Since β is large for incompressible fluids, mechanical types of capacitors are used. To obtain significant compressibility capacitance with a liquid, V would have to be very large.
– Capacitance of gaseous systems can be that of mechanical capacitors or volume-type capacitance, since β is low for compressible fluids.

- The effect of fluid capacitance must be considered relative to the rest of the system. For example, the resistance connected to the capacitor and the bandwidth of interest determine how significant the capacitive effects are.

Inertance (Fluid Inductance)

- While devices for introducing resistances (orifices, capillaries) and compliance (tanks, accumulators) are often intentionally designed into fluid systems, the inertia effect is more often than not a parasitic one.
- The inertance of pipes is perhaps most commonly encountered and we will emphasize it.
  - Any flowing fluid has stored kinetic energy because of its density (mass) and velocity. The inertance of a finite-size lump of fluid represents a summing up of this kinetic energy over the volume of the lump.

– The simplest assumption possible here is that of one-dimensional flow where all the fluid particles have identical velocities at any instant of time. Since every fluid particle has the same velocity, a lump of fluid can be treated as if it were a rigid body of mass $M = \rho AL$.

– A pressure drop Δp across a pure inertance element will cause a fluid acceleration according to Newton's Law:

$$A\Delta p = \rho AL\frac{dv}{dt} = \rho AL\left(\frac{1}{A}\frac{dq}{dt}\right)$$

$$\Delta p = \frac{\rho L}{A}\frac{dq}{dt} \triangleq I_f\frac{dq}{dt}$$

$$I_f \triangleq \frac{\rho L}{A}$$

- This is analogous to $e = L(di/dt)$ for inductance in electrical systems and $f = M(dv/dt)$ for mass in mechanical systems.
- This equation is valid for liquids and gases; however, if a gas is used, the density must be evaluated at the upstream conditions.
- The significance of fluid inductance must also be evaluated relative to the rest of the system and the bandwidth or frequency of interest.

Lumped vs. Distributed Fluid System Models

- Lumped-parameter models are always approximations to the more-correct distributed models.
- The comparison of different models for a system is often best done in terms of frequency response.
- Any practical dynamic system will experience input signals whose maximum frequency will be limited to a definite value.
- In lumped modeling, the variation of pressure and velocity along the length of a pipeline was assumed a stepwise one. Within a given lump there was no variation, but pressure and velocity did change when we went to a neighboring lump.

- All distributed models allow a smooth, not stepwise, variation, which of course is more correct.
- It is clear that as a lumped model uses more and smaller lumps, the stepwise variation more nearly approximates the true smooth variation.
- How many lumps are needed to get accurate results with a lumped model?
- Experience with many kinds of systems shows that if we choose 10 lumps per wavelength of the highest frequency, we usually get good results. That is, a stepwise variation is an acceptable approximation to a sine wave if there are 10 steps per wavelength. If we decide that our lumped model needs to be good for excitations of higher frequency than we originally planned, the lumps must get smaller and there must be more of them.

- Once we have a formula for the velocity of propagation (the speed with which a disturbance propagates through the medium) and choose the highest frequency of interest, we can pick a size and number of lumps which will give good accuracy up to that frequency, using the "10 lumps per wavelength" guideline, where the wavelength λ is the velocity of propagation c divided by the frequency f. Higher operating frequencies require more and smaller lumps.
- Remember why we use lumped-parameter models. They can be solved easily for the time response to any form of input and they also allow easy simulation with standard software.

Fluid Impedance

- Most fluid systems do not really require the separation of pressure/flow relations into their resistive, compliant, and inertial components; this separation is mainly one of analytical convenience.
- For complex fluid systems where experimental measurements may be a necessity, the measurement of overall pressure/flow characteristics has become a useful tool.
- The term fluid impedance is directly analogous to mechanical and electrical impedance and is defined as the transfer function relating pressure drop (or pressure), as output, to flow rate, as input.

- **Fluid Resistance**

$$\frac{\Delta p}{q}(D) = R_f$$

$$\frac{\Delta p}{q}(i\omega) = R_f$$

- **Fluid Compliance**

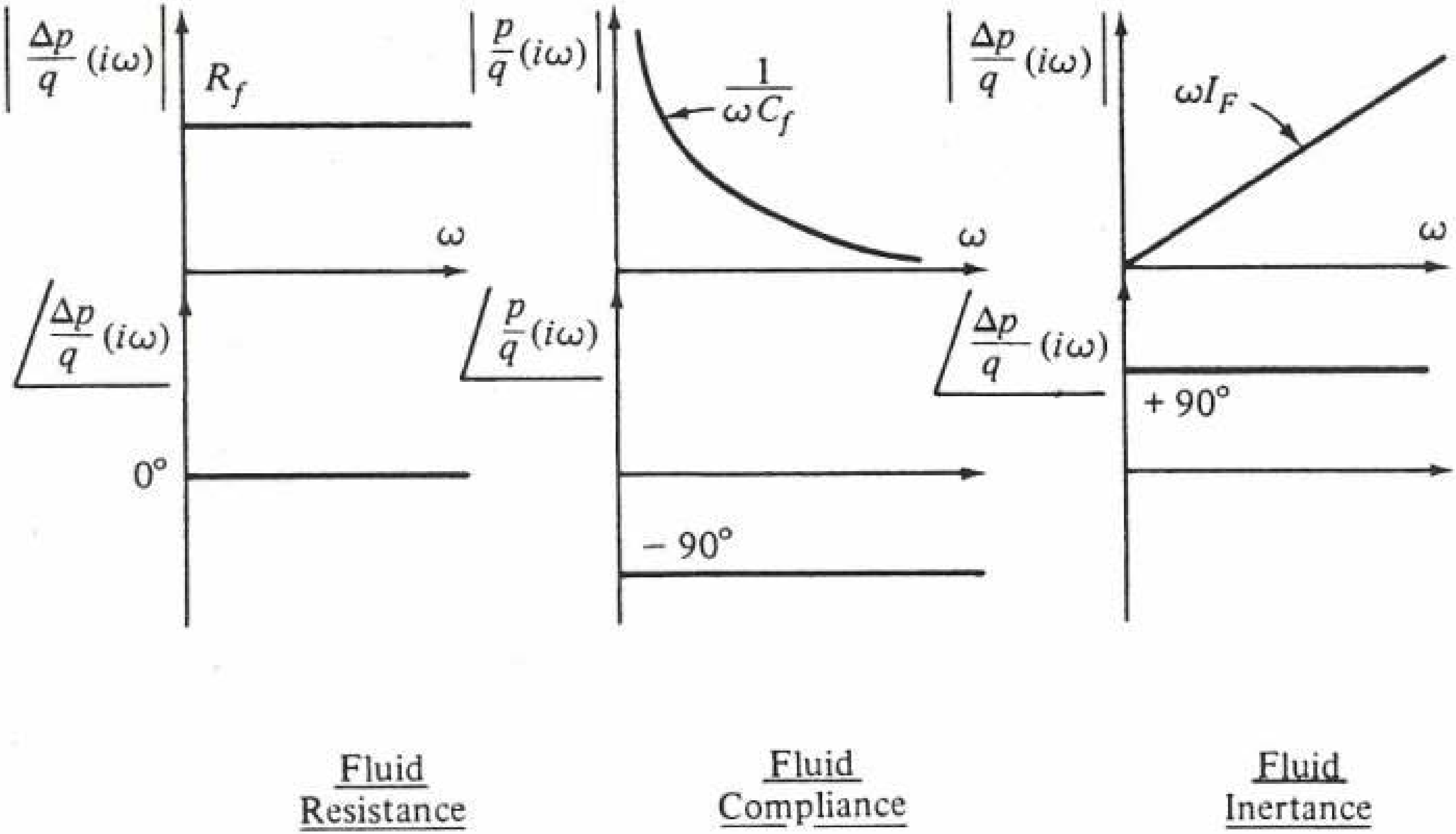
$$\frac{\Delta p}{q}(D) = \frac{1}{C_f D}$$

$$\frac{\Delta p}{q}(i\omega) = \frac{1}{\omega C_f} \angle -90^\circ$$

- **Fluid Inertance**

$$\frac{\Delta p}{q}(D) = I_f D$$

$$\frac{\Delta p}{q}(i\omega) = \omega I_f \angle +90^\circ$$



Fluid Impedance of the Three Basic Elements

- If the fluid impedance is known as an operational transfer function, one can calculate the response to any given input by solving the corresponding differential equation.
- If the fluid impedance is measured by the frequency response technique, we then do not have a transfer function in equation form, we only have curves. The response to sinusoidal inputs is of course easily calculated from such curves.
- Suppose, however, that we want to find the response to an input which is not a sine wave but rather has a time variation of arbitrary shape. Two methods are available:
  - Curve-fit the measured frequency-response curves with analytical functions, trying different forms and numerical values until an acceptable fit is achieved. Software to expedite such curve-fitting is available. Having a formula for the transfer function is the same as having the system differential equation.

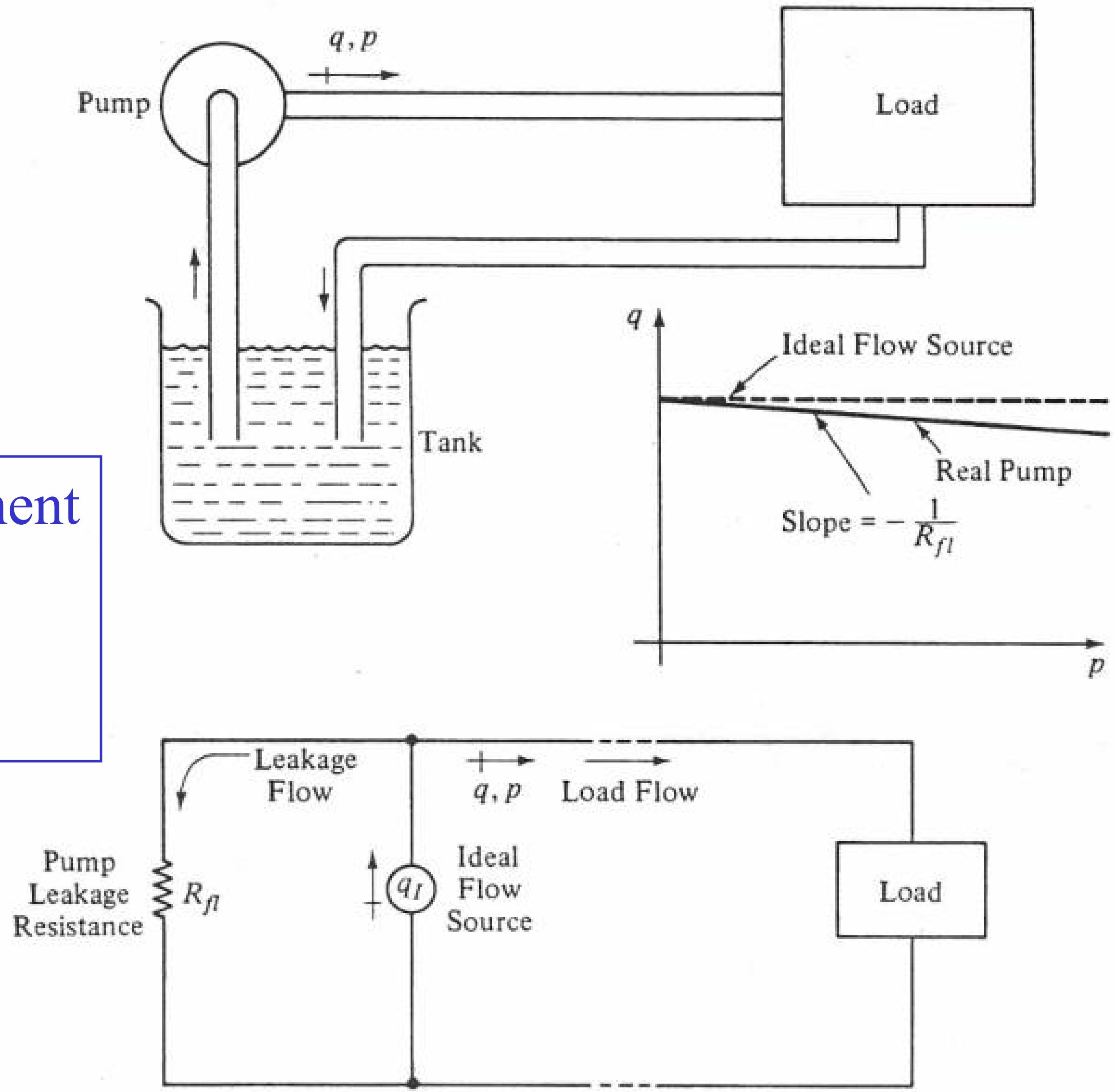
- Use the measured frequency-response curves directly, without any curve fitting. One must compute the Fourier transform of the desired time-varying input signal to get its representation in the frequency domain. This operation converts the time function into its corresponding frequency function, which will be a complex number which varies with frequency. This complex number is multiplied, one frequency at a time, with the complex number, whose magnitude and phase can be graphed versus frequency. This new complex number is the frequency representation of the output of the system. The final step is the inverse Fourier transform to convert this function back into the time domain, to give the system output as a specific, plottable, function of time.

- This discussion applies to any linear, time-invariant, dynamic system, not just fluid systems. That is, if we can measure the frequency response, we can get the response to any form of input, not just sine waves.

Fluid Sources: Pressure and Flow Rate

- An *ideal pressure source* produces a specified pressure at some point in a fluid system, no matter what flow rate might be required to maintain this pressure.
- An *ideal flow source* produces a specified flow rate, irrespective of the pressure required to produce this flow.
- In fluid systems, the most common source of fluid power is a pump or compressor of some sort.
- A positive-displacement liquid pump draws in, and then expels, a fixed amount of liquid for each revolution of the pump shaft. When driven at a constant speed, such a pump closely approximates an ideal constant-flow source over a considerable pressure range.

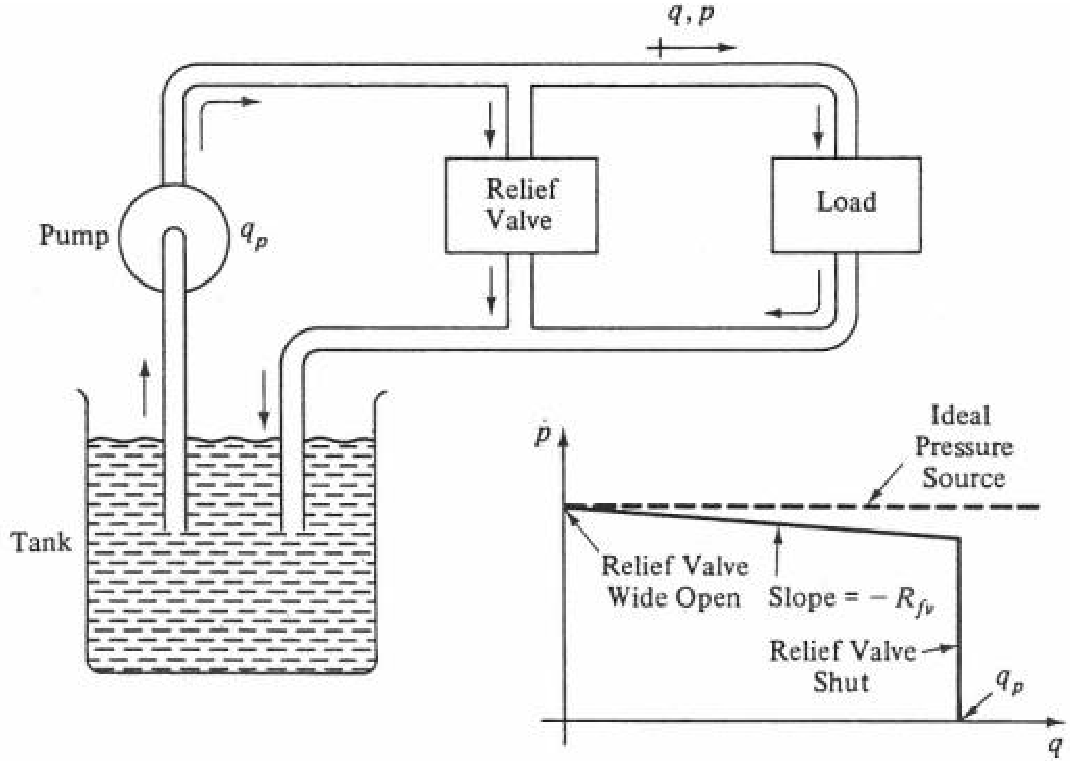
- Its main departure from ideal behavior is a decrease in flow rate, due to leakage through clearance spaces, as load pressure increases. This leakage flow is proportional to pressure; thus one can represent a real pump as a parallel combination of an ideal flow source and a linear (and large) flow resistance R_{fl}.
- If the inlet flow impedance of the load is low relative to R_{fl}, most of the flow goes into the load rather than the pump leakage path, and the pump acts nearly as an ideal flow source.

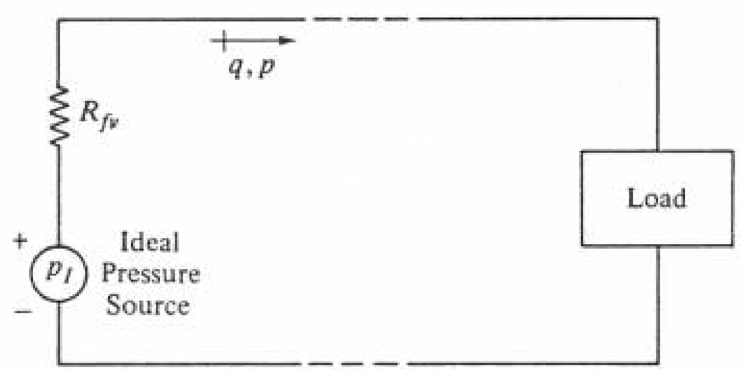


Positive-Displacement Pump as a Flow Source

- When we need to manipulate a flow rate as function of time, several possibilities exist.
  - A fixed-displacement pump may be driven at a time-varying speed. An electric motor drives the pump shaft, a flow sensor measures the flow, and a feedback controller compares the desired flow with the measured flow and commands the motor to change speed so as to keep the actual flow close to the desired at all times.
  - Instead of a fixed-displacement pump, a variable-displacement pump could be used. Here the pump shaft speed is constant, but pump output per revolution can be varied, while the pump is running. A stroking mechanism allows flow rate to be varied smoothly and quickly from full flow in one direction, through zero flow, to full flow in the reverse direction. The stroking mechanism could be driven directly or we could again use a feedback scheme.

- By combining a positive-displacement pump with a relief valve, one can achieve a practical constant-pressure source. This real source will not have the perfect characteristic of an ideal pressure source, but can be modeled as a combination of an ideal source with a flow resistance.
- A relief valve is a spring-loaded valve which remains shut until the set pressure is reached. At this point it opens partially, adjusting its opening so that the pump flow splits between the demand of the load and the necessary return flow to the tank. To achieve this partial opening against the spring, the pressure must change slightly; thus we do not get an exactly constant pressure.
- This real source can be modeled as a series combination of an ideal pressure source with a small flow resistance.


q, p
Relief
Valve
Load
Pump
q_p
Tank
p
Ideal
Pressure
Source
Relief Valve
Wide Open
Slope = $-R_{fv}$
Relief Valve
Shut
q_p
q


q, p
R_{fv}
Load
+
p_I
−
Ideal
Pressure
Source

- These examples do not exhaust the possibilities with regard to power sources in fluid systems, but they should give some idea of how real sources may be modeled in terms of ideal sources and passive elements.
- Other fluid power sources encountered in practice include centrifugal pumps, accumulators (used for short-term power supplies), elevated tanks or reservoirs (gravity is the energy source), etc.

Mechatronics Exercise:
Modeling, Analysis, & Control of an Electrohydraulic Valve-Controlled Servomechanism

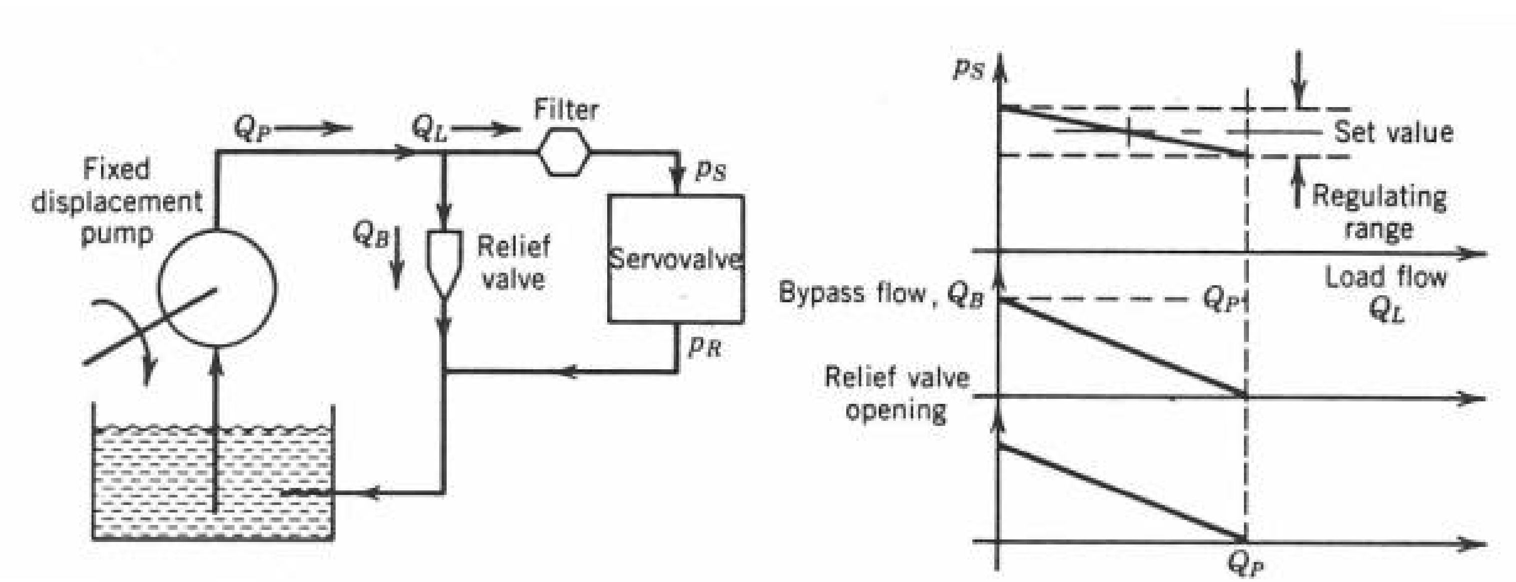
Introduction

- Although a wide variety of detailed hydraulic control schemes are in use, a useful classification is that of *pump-controlled* versus *valve-controlled* systems.
  - *Pump-controlled systems* are usually relatively high power (above 10 or 20 hp) applications, where efficiency is economically significant and response speed requirements are modest (less than 10 Hz frequency response).

– *Valve-controlled systems* are faster but are generally quite inefficient. For a low-power system, inefficiency has little economic impact. For fast high-power systems where speed specifications can only be met by valve control, the economic cost of low efficiency must be accepted.

– In *pump-controlled systems*, the fluid power supply must be included in the system model, while the analysis of *valve-controlled systems* can proceed without consideration of power supply details if one assumes the existence of a power source of constant supply pressure, p_s, irrespective of flow demand. Power supplies that approximate this behavior are available in several different forms that trade off complexity, cost, efficiency, and static/dynamic pressure regulation accuracy.

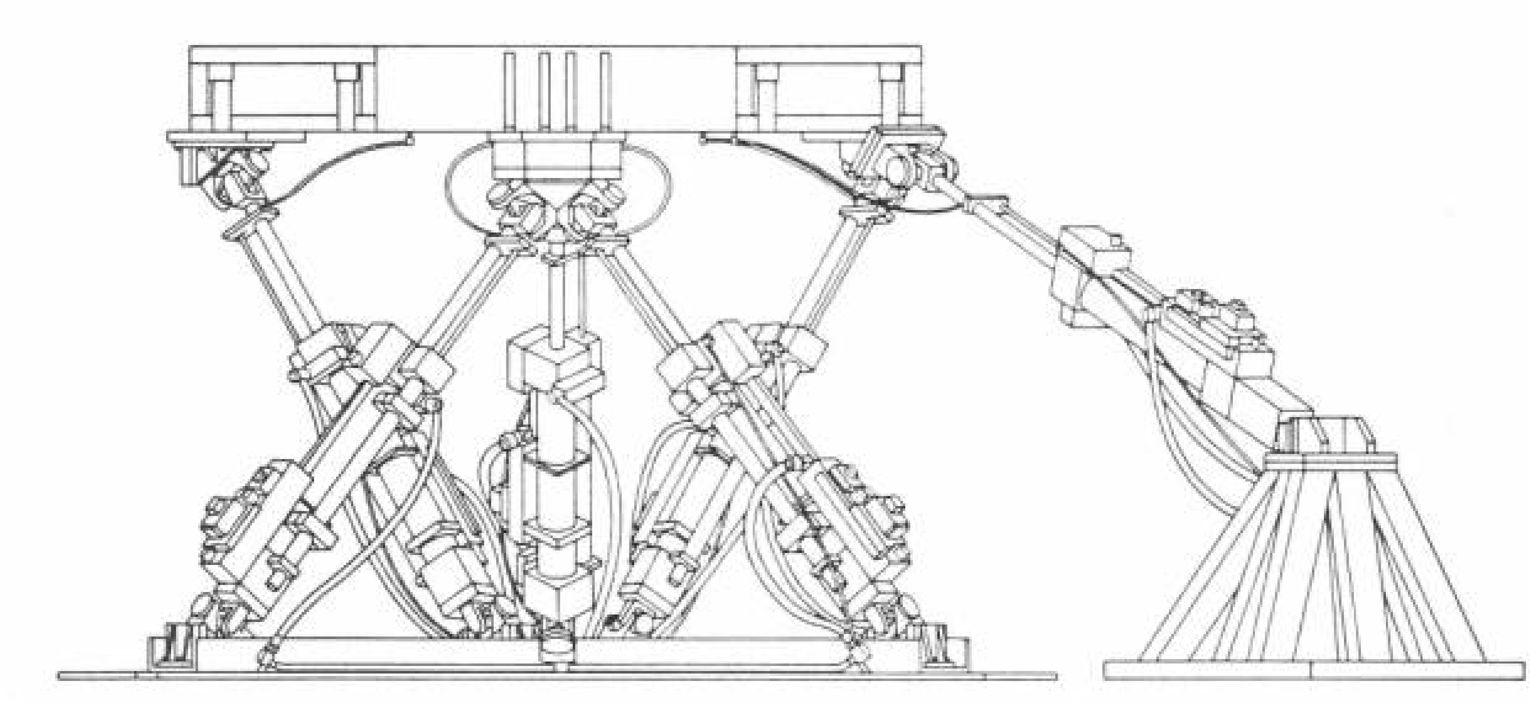
- For example, the spring-loaded relief valve is completely shut until pressure reaches the low end of the regulating range, whereupon it opens sufficiently to bypass any pump flow not required by the servovalve.
- The fluid power of the bypassed flow is completely lost and converted to heat.



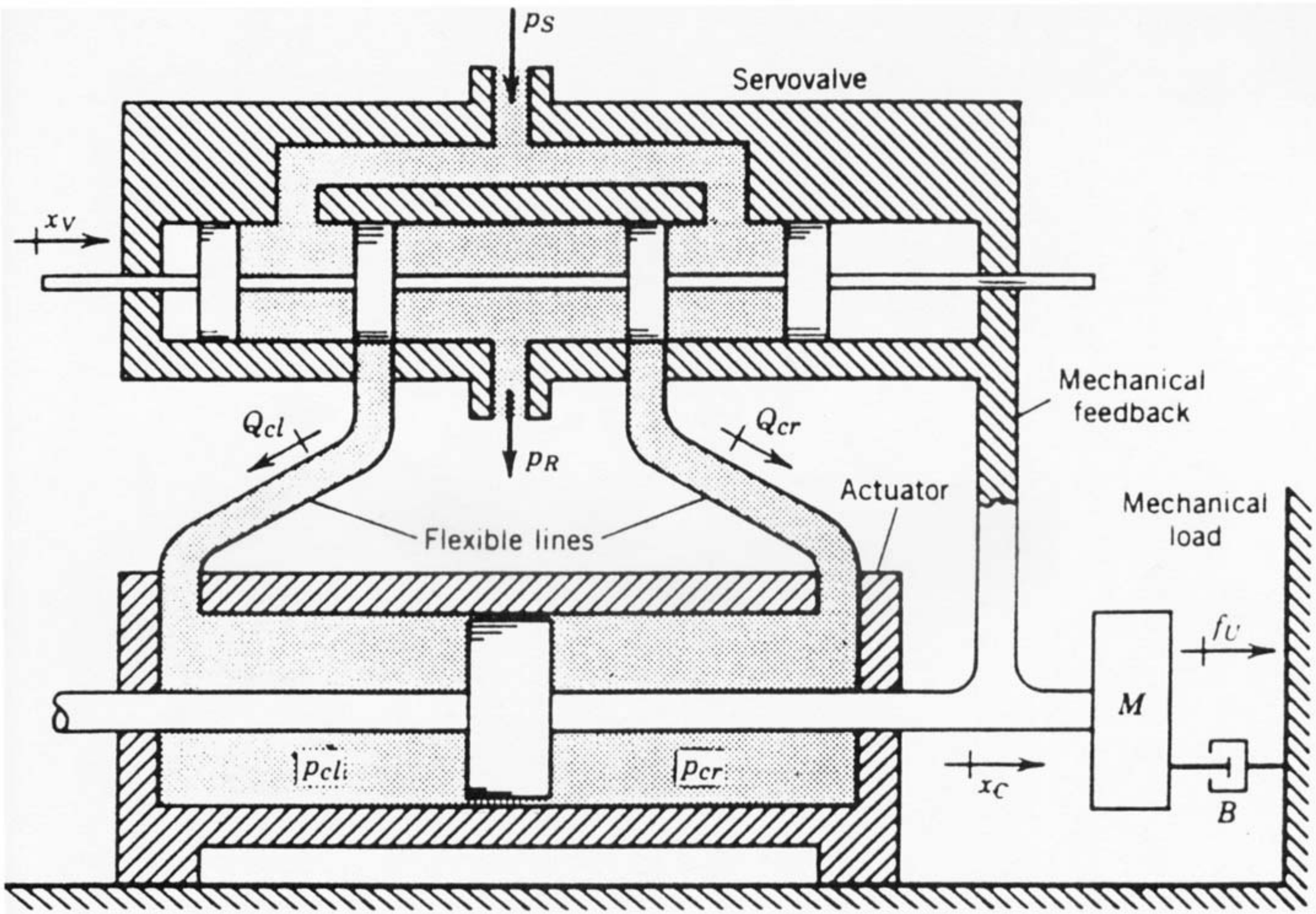
pump flow

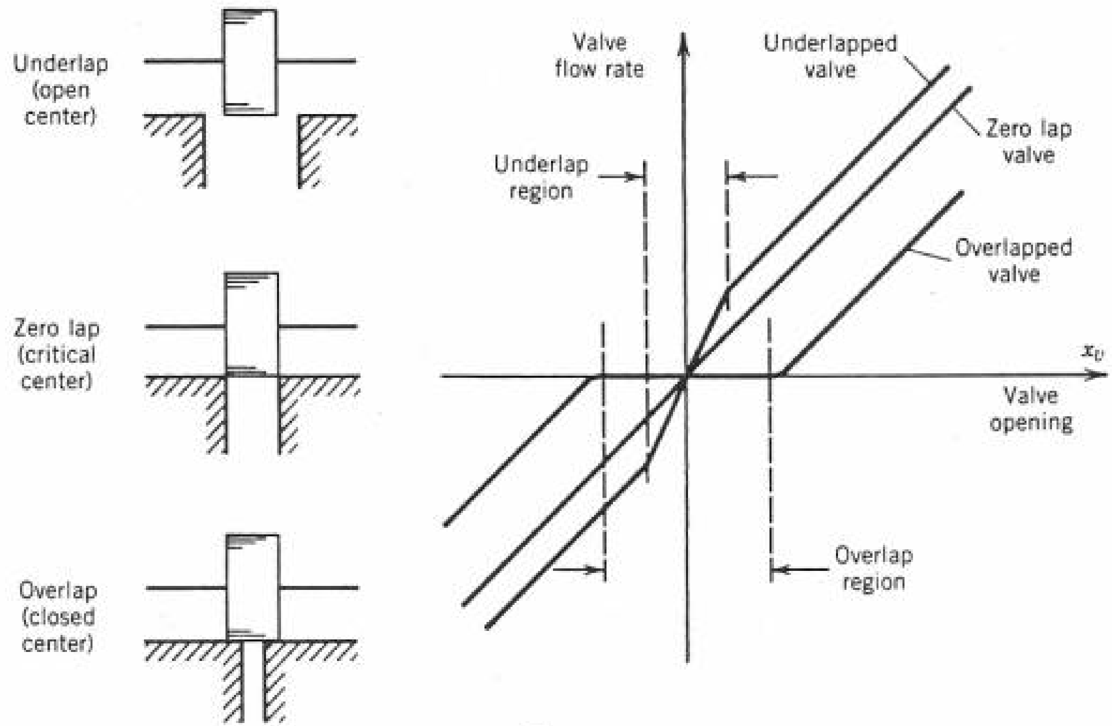
- When the servo system requires no flow, all the pump-generated power is converted to heat, giving zero efficiency.
- Supply pressure p_s varies by about ± 3% over the regulating range for steady conditions and response to transient flow requirements is quite fast.
- Pump size must be chosen to match the largest anticipated servo-system demand.
- Thus the standard assumption of constant p_s used in servo-system analysis is usually reasonable.

- The figure below shows a flight simulator, a sophisticated application of valve-controlled servos where the motions of six actuators are coordinated to provide roll, pitch, and yaw rotary motions plus x, y, z translation.



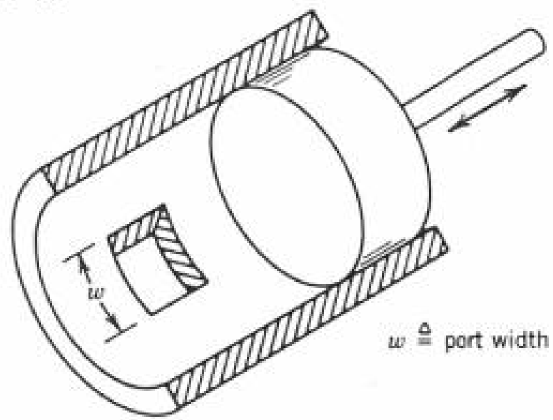
Physical System: Valve-Controlled Servo





Real valves always exhibit either underlap ($x_u > 0$) or overlap ($x_u < 0$) behavior.

Underlap / Overlap effects are usually very small.



spool valve

Physical Modeling Assumptions

- supply pressure is constant at 1000 psig
- reservoir pressure is constant at 0 psig
- valve is zero lap
- actuator pressures p_{cl} and p_{cr} each come to $p_s/2$ at the servo rest condition
- neglect inertia of the fluid
- cylinder and piston are rigid
- sensor dynamics are negligible
- parameters are constant

- compressibility effects are neglected in the orifice flow equations, but not in the cylinder equations as pressures can be high during acceleration and deceleration periods and oil compressibility can have a destabilizing effect
- both flow orifices are identical, i.e., the flow and pressure coefficients are identical for both
- disturbance to the mass is zero

Physical Model Parameters

x_u	0 inches, zero lap condition
p_s	1000 psig (constant), supply pressure
C_d	0.6, orifice discharge coefficient
w	0.5 in, valve port width
ρ	7.8E-5 lbf-s^2/in^4, fluid density
A_p	2.0 in^2, piston area
β	100,000 psi, bulk modulus of fluid
M	0.03 lbf-s^2/in, mass
K_{pl}	0.001 in^3/s-psi, piston leakage coefficient
B	100 lbf-s/in, viscous damping coefficient
V_{l0}	4.0 in^3, volume at operating point of left cylinder
V_{r0}	4.0 in^3, volume at operating point of right cylinder
p_{cl0}	500 psi, initial pressure of left cylinder
p_{cr0}	500 psi, initial pressure of right cylinder
x_{C0}	0 in, initial displacement of mass
$\dot{x}_{C0}$	0 in/sec, initial velocity of mass
f_U	0 lbf, disturbance
p_r	0 psig, return pressure

Nonlinear Mathematical Model

- Equations for the orifice volume flow rates Q_{cl} and Q_{cr} for the left and right ends of the cylinder
- Equations for conservation of mass (continuity equation) for the left and right ends of the cylinder
- Newton's 2^{nd} Law applied to the moving mass
- Kinematic relation representing the mechanical feedback

The Variable Orifice

- The variable orifice is at the heart of most fluid power systems and is the most popular device for controlling fluid flow. It is the fluid equivalent of the electrical resistor and like the resistor its use leads to energy dissipation.
- Overriding advantages include simplicity, reliability, and ease of manufacture.
- The orifice can be used in analog (infinite number of positions) or discontinuous (fully open or fully closed) modes, depending on the application.

- Knowledge of the orifice equations for incompressible and compressible flow is essential. Here we consider incompressible flow.
- The orifice equation for the volume flow rate Q of an incompressible fluid, assuming that the upstream area is much greater than the orifice area A, is:

$$Q = C_d A \sqrt{\frac{2(P_u - P_d)}{\rho}}$$

C_d = orifice discharge coefficient

- In a given system the dominating variables are usually the pressure drop and the orifice area.

- The predominant nonlinearity is the square root term, but C_d depends on the Reynolds number and cavitation.
- Cavitation refers to the formation and collapse of cavities, containing air or gas, in the liquid. If the pressure is reduced far enough hydraulic oil will vaporize and vapor cavities will form. The pressure at which vaporization commences is called the vapor pressure of the liquid and is very dependent on the temperature of the liquid. As the temperature increases, the vapor pressure increases.

- The phenomenon of cavitation damage in hydraulic machinery, turbines, pumps, and propellers is well known. It has been shown both analytically and experimentally that when cavities collapse as a result of increased hydraulic pressure, very large pressures can be developed. However, there is still controversy about the exact mechanism of the damaging process.

Orifice Flow-Rate Equations

x_u positive: valve underlap
x_u negative: valve overlap
x_u zero: valve zero lap
x_v is displacement of valve spool

$$Q_{cl} = C_d w(x_u + x_v)\sqrt{\frac{2(p_s - p_{cl})}{\rho}}$$ valid when $(x_u + x_v)$ is > 0.

This is flow into the left cylinder.

$$Q_{cl} = -C_d w(x_u - x_v)\sqrt{\frac{2(p_{cl} - p_r)}{\rho}}$$ valid when $(x_u - x_v)$ is > 0.

This is flow out of the left cylinder.

$$Q_{cr} = -C_d w(x_u + x_v)\sqrt{\frac{2(p_{cr} - p_r)}{\rho}}$$ valid when $(x_u + x_v)$ is > 0.

This is flow out of the right cylinder.

$$Q_{cr} = C_d w(x_u - x_v)\sqrt{\frac{2(p_s - p_{cr})}{\rho}}$$ valid when $(x_u - x_v)$ is > 0.

This is flow into the right cylinder.

Conservation of Mass Equations

- Conservation of Mass

$$0 = \frac{\partial}{\partial t}\int_{CV} \rho dV + \int_{CS} \rho \vec{v} \bullet \overrightarrow{dA}$$

$$0 = \rho_{CV}\dot{V}_{CV} + V_{CV}\dot{\rho}_{CV} + \rho Q_{net}$$

$$0 = \dot{V}_{CV} + \frac{V_{CV}}{\rho}\dot{\rho} + Q_{net}$$

- Here we assume that all of the densities of the system (inlet flow, outlet flow, and control volume) are the same and equal to ρ.

- This assumption is justified for incompressible fluids and is quite accurate for compressible fluids if pressure variations are not too large and the temperature of flow into the control volume is almost equal to the temperature of the flow out of the control volume.

- The isothermal bulk modulus is given by:

$$\beta = \rho_0 \left.\frac{\partial P}{\partial \rho}\right|_{P_0, T_0} = \left.\frac{\partial P}{\partial \rho / \rho_0}\right|_{P_0, T_0}$$

- Therefore:

$$\dot{\rho} = \frac{\partial \rho}{\partial P}\frac{dP}{dt} = \frac{\rho_0}{\beta}\dot{P}$$

- Conservation of Mass can be written as:

$$0 = \dot{V} + \frac{V}{\beta}\dot{P}_{CV} + Q_{net}$$

- Evaluating terms:

Left cylinder
$$\begin{cases} Q_{net} = -Q_{cl} + K_{pl}(p_{cl} - p_{cr}) \\ \dot{V} = A_p \dfrac{dx_C}{dt} \\ \dfrac{V}{\beta}\dot{P}_{CV} = \dfrac{(V_{l0} + A_p x_C)}{\beta}\dfrac{dp_{cl}}{dt} \end{cases}$$

- The resulting equations for the left and right cylinders are:

$$Q_{cr} - \frac{(V_{r0} - A_p x_C)}{\beta}\frac{dp_{cr}}{dt} + K_{pl}(p_{cl} - p_{cr}) = -A_p \frac{dx_C}{dt}$$

$$Q_{cl} - \frac{(V_{l0} + A_p x_C)}{\beta}\frac{dp_{cl}}{dt} - K_{pl}(p_{cl} - p_{cr}) = A_p \frac{dx_C}{dt}$$

- V_{l0} and V_{r0} are the volumes at the operating point of the left and right cylinders, respectively.
- β is the bulk modulus (isothermal) of the fluid defined by the expression:

$$\beta = -V \frac{dp}{dV}$$

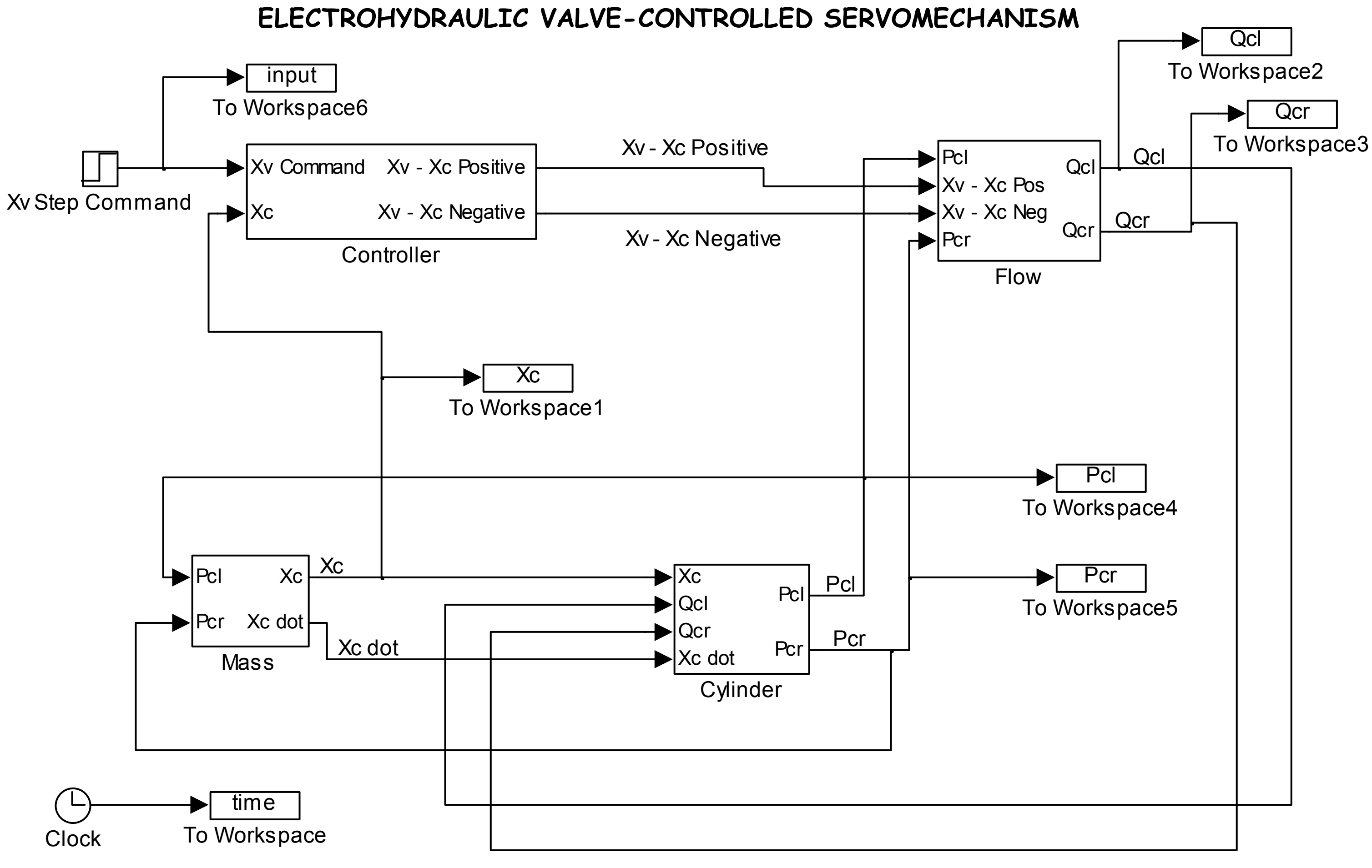
- $(V_{l0} + A_p x_C)$ and $(V_{r0} - A_p x_C)$ represent the compressed volumes of the left and right sides of the cylinder, respectively, which include the lines from the valve to the actuator plus the ends of the cylinder.

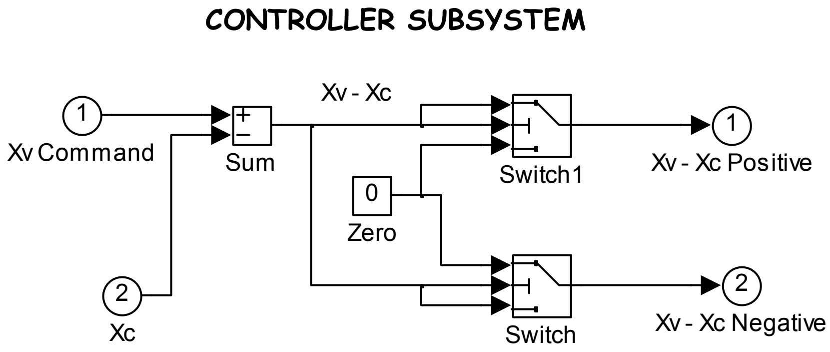
Newton's Second Law & Feedback Equations

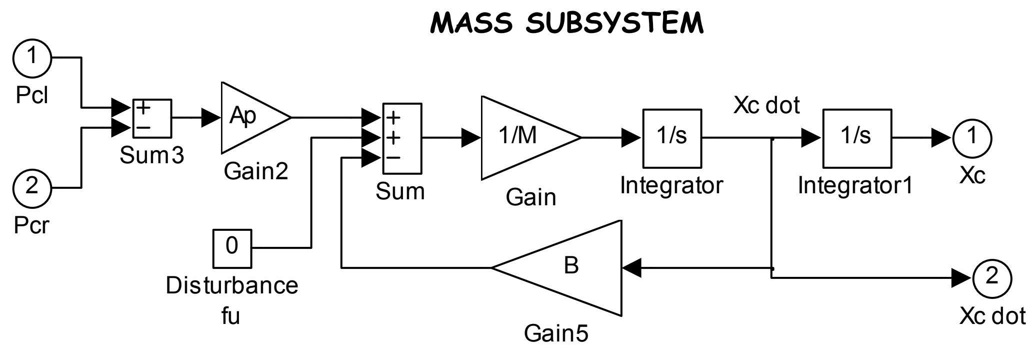
$$(p_{cl} - p_{cr})A_p - B\frac{dx_C}{dt} + f_U = M\frac{d^2x_C}{dt^2}$$

$$x_v = x_V - x_C$$

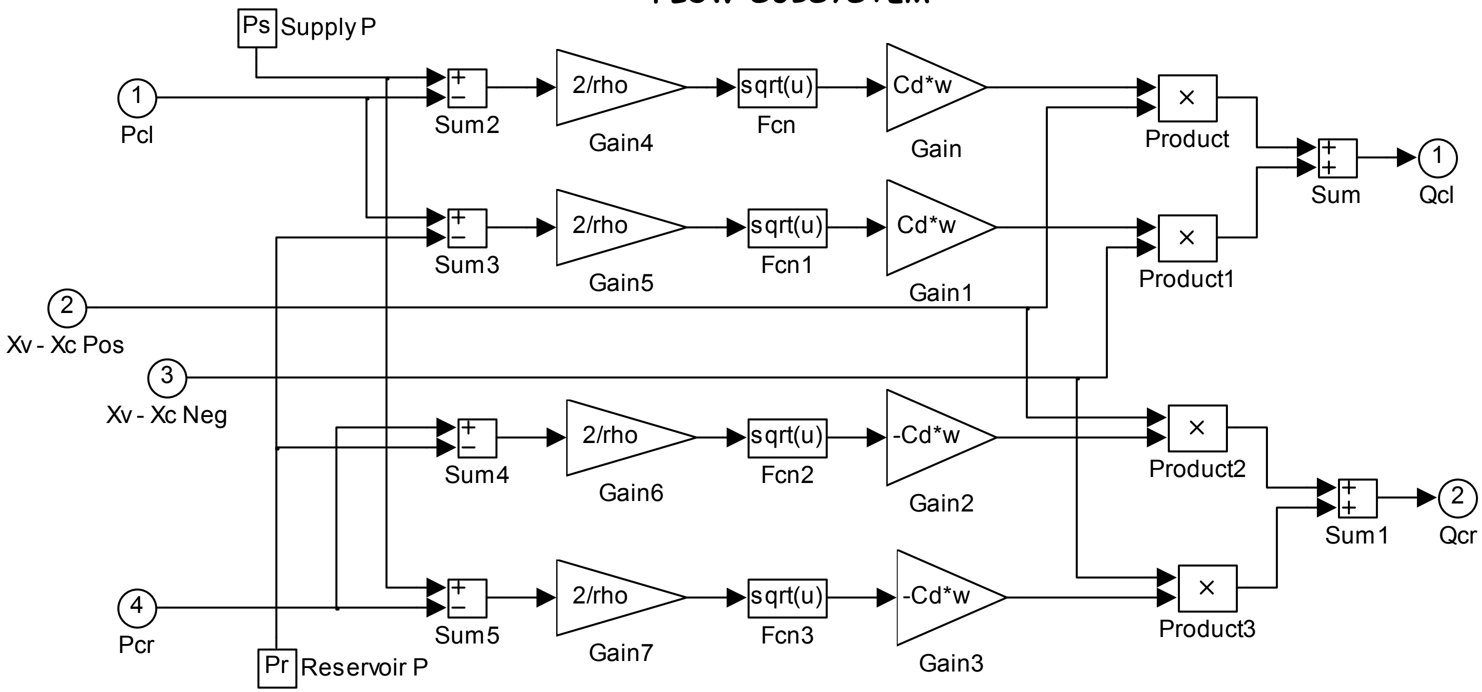
Simulink Block Diagrams

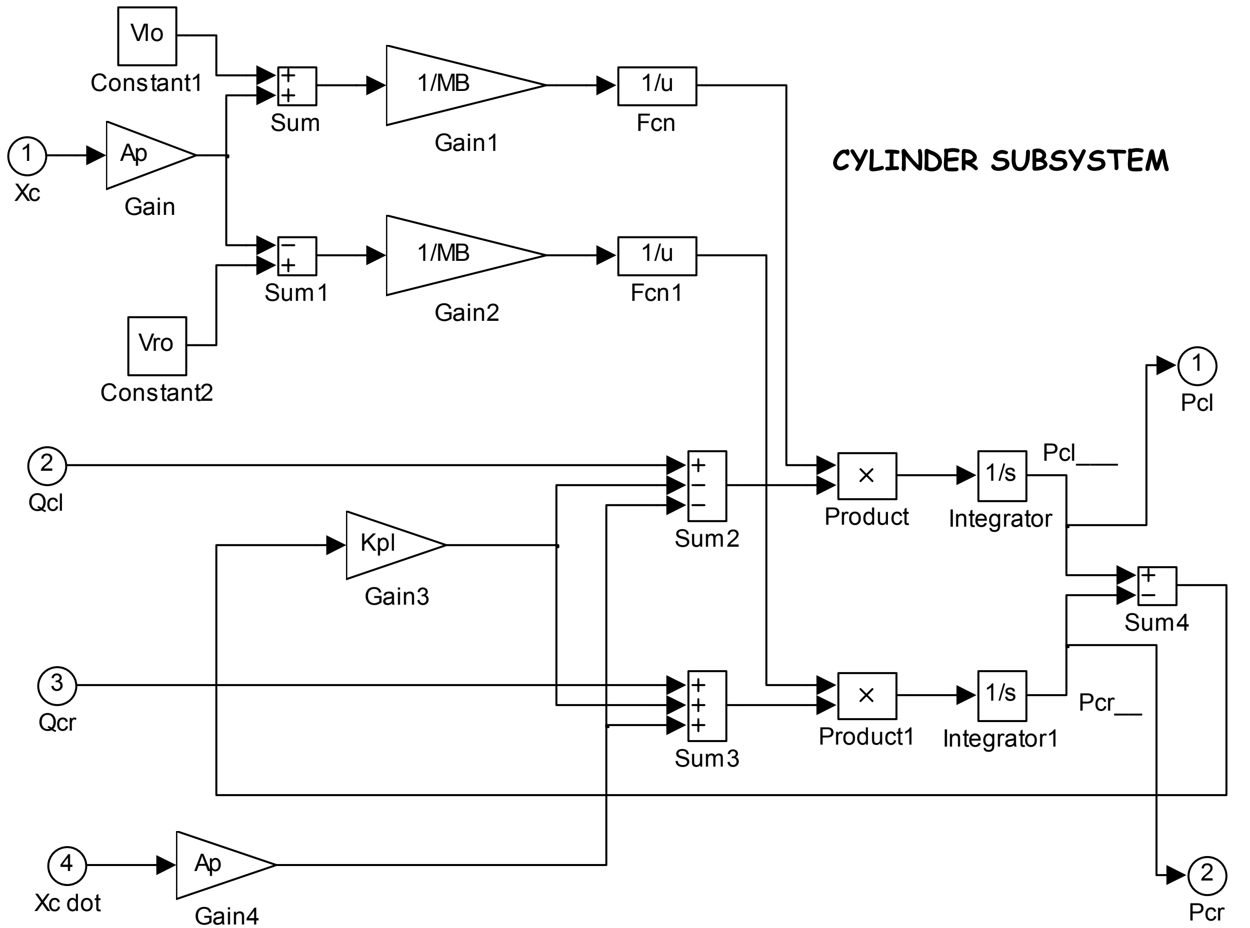



CONTROLLER SUBSYSTEM
1
Xv Command
2
Xc
Sum
Xv - Xc
0
Zero
Switch1
Switch
1
Xv - Xc Positive
2
Xv - Xc Negative


MASS SUBSYSTEM
1
Pcl
2
Pcr
Sum3
Ap
Gain2
0
Disturbance
fu
Sum
1/M
Gain
1/s
Integrator
Xc dot
1/s
Integrator1
1
Xc
B
Gain5
2
Xc dot

FLOW SUBSYSTEM




Vlo
Constant1
Sum
1/MB
Gain1
1/u
Fcn
CYLINDER SUBSYSTEM
1
Xc
Ap
Gain
1/MB
Gain2
1/u
Fcn1
Sum1
Vro
Constant2
1
Pcl
2
Qcl
Sum2
Product
1/s
Integrator
Pcl__
Kpl
Gain3
Sum4
3
Qcr
Sum3
Product1
1/s
Integrator1
Pcr__
4
Xc dot
Ap
Gain4
2
Pcr

MatLab File of Constants and Expressions

```
M=0.03;
B=100;
Ap=2.0;
Kpl=0.001;
Vlo=4.0;
Vro=4.0;
MB=100000;
Pr=0;
Ps=1000;
rho=7.8e-5;
Cd=0.6;
w=0.5;
Pclo=500;
Pcro=500;
Xcdoto=0;
Xco=0;
Vo=4.0;
Cx=1074.172;
Cp=0;
```

```
A=[-(Cp+Kpl)*MB/Vo Kpl* MB /Vo 0 -Ap* MB/Vo;
Kpl*MB/Vo -(Cp+Kpl)* MB/Vo 0 Ap* MB/Vo;
0 0 0 1;Ap/M -Ap/M 0 -B/M];
B1=[MB*Cx/Vo 0;- MB*Cx/Vo 0;0 0;0 1/M];
C=[1 0 0 0;0 1 0 0;0 0 1 0;0 0 0 1];
D=[0 0;0 0;0 0;0 0];
K=(2*Cx*Ap)/(2*Ap^2+B*(Cp+2*Kpl));
omegan=sqrt((MB*(2*Ap^2+B*(Cp+2*Kpl)))/(M*Vo));
zeta=(B+((2*MB*M)/Vo)*Kpl+((MB*M)/Vo)*Cp)/(2*sqrt(((MB*M)/Vo)
*(2*Ap^2+B*(Cp+2*Kpl))));
XcXvNum=K*omegan^2;
XcXvDen=[1 2*zeta*omegan omegan^2 0];
%[XcXvNumCL,XcXvDenCL]=cloop(XcXvNum,XcXvDen,-1);
Kc=1;
XcXvNumCL=Kc*K*omegan^2;
XcXvDenCL=[1 2*zeta*omegan omegan^2 Kc*K*omegan^2];
```

Linear System Analysis

- Restrict the analysis to small perturbations around a chosen operating point. A linearized approximate model may be obtained that provides many useful results.
- Valve flow equations can be thought of as relations between a dependent variable (flow rate) and two independent variables (spool motion and cylinder pressure) and thus can be linearized about any desired operating point.

$$Q_v \approx Q_{v,0} + \left.\frac{\partial Q_v}{\partial x_v}\right|_{\text{operating point}} x_{v,p} + \left.\frac{\partial Q_v}{\partial p_c}\right|_{\text{operating point}} p_{c,p}$$

$$\text{flow gain} = C_x = \left.\frac{\partial Q_v}{\partial x_v}\right|_{\text{operating point}}$$

$$\text{pressure coefficient} = C_p = -\left.\frac{\partial Q_v}{\partial p_c}\right|_{\text{operating point}}$$

$$Q_v \approx Q_{v,0} + C_x x_{v,p} - C_p p_{c,p}$$

- Assume that $Q_{v,0} = 0$ and that the numerical values of C_x and C_p are equal for the Q_{cl} and Q_{cr} equations (correct assumptions for commonly used operating points).

$$Q_{cl} \approx C_x x_{v,p} - C_p p_{cl,p} \qquad\qquad Q_{cr} \approx -C_x x_{v,p} - C_p p_{cr,p}$$

- Take the volumes $(V_{l0} + A_p x_C)$ and $(V_{r0} - A_p x_C)$ to be constant at $V_{l0} = V_{r0} = V_0$, a good approximation for small changes in x_C.

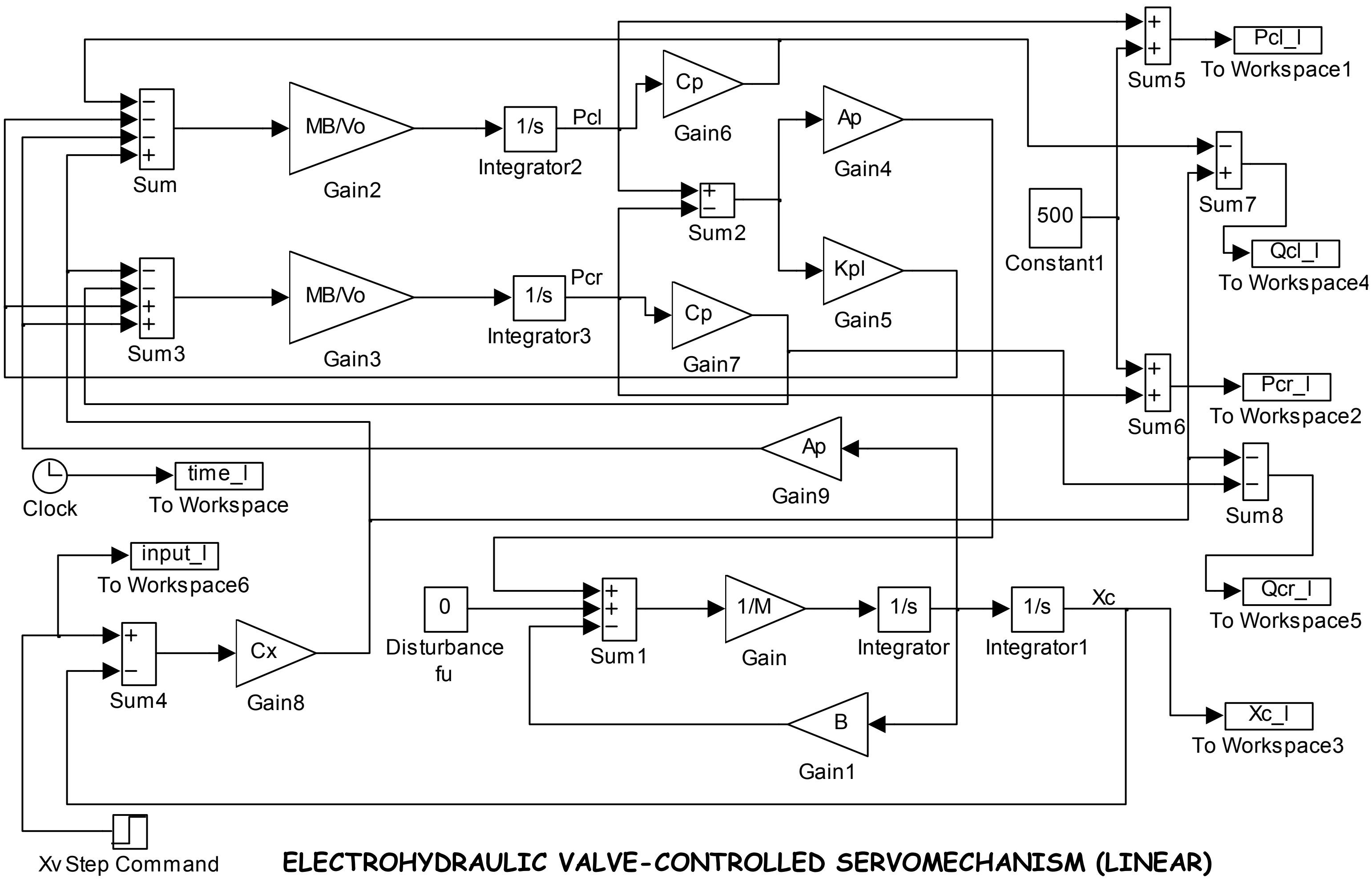
Linearized Set of Equations:

$$\left(C_x x_{v,p} - C_p p_{cl,p}\right) - \frac{V_0}{\beta}\frac{dp_{cl,p}}{dt} - K_{pl}\left(p_{cl,p} - p_{cr,p}\right) = A_p \frac{dx_{C,p}}{dt}$$

$$\left(-C_x x_{v,p} - C_p p_{cr,p}\right) - \frac{V_0}{\beta}\frac{dp_{cr,p}}{dt} + K_{pl}\left(p_{cl,p} - p_{cr,p}\right) = -A_p \frac{dx_{C,p}}{dt}$$

$$\left(p_{cl,p} - p_{cr,p}\right)A_p - B\frac{dx_{C,p}}{dt} + f_{U,p} = M\frac{d^2 x_{C,p}}{dt^2}$$

Simulink Block Diagram of the Linear System



ELECTROHYDRAULIC VALVE-CONTROLLED SERVOMECHANISM (LINEAR)

- If we take the Laplace Transform of these equations, we can derive six useful transfer functions relating the two inputs, x_v and f_U, to the three outputs, p_{cl}, p_{cr}, and x_C.

$$\begin{bmatrix} \frac{V_0 s + \beta\left(K_{pl} + C_p\right)}{C_x \beta} & \frac{-K_{pl}}{C_x} & \frac{A_p s}{C_x} \\ \frac{K_{pl}}{C_x} & \frac{-V_0 s - \beta\left(K_{pl} + C_p\right)}{C_x \beta} & \frac{A_p s}{C_x} \\ -A_p & A_p & Ms^2 + Bs \end{bmatrix} \begin{bmatrix} p_{cl} \\ p_{cr} \\ x_C \end{bmatrix} = \begin{bmatrix} x_v \\ x_v \\ f_U \end{bmatrix}$$

One of these transfer functions is:

$$\frac{x_C}{x_v}(s) = \frac{K}{s\left(\frac{s^2}{\omega_n^2} + \frac{2\zeta s}{\omega_n} + 1\right)}$$

where

$$K = \frac{2C_x A_p}{2A_p^2 + B\left(C_p + 2K_{pl}\right)}$$

$$\omega_n = \sqrt{\frac{\beta\left[2A_p^2 + B\left(C_p + 2K_{pl}\right)\right]}{MV_0}}$$

$$\zeta = \frac{B + \left(\frac{2\beta M}{V_0}\right)K_{pl} + \left(\frac{\beta M}{V_0}\right)C_p}{2\sqrt{\frac{\beta M}{V_0}\left[2A_p^2 + B\left(C_p + 2K_{pl}\right)\right]}}$$

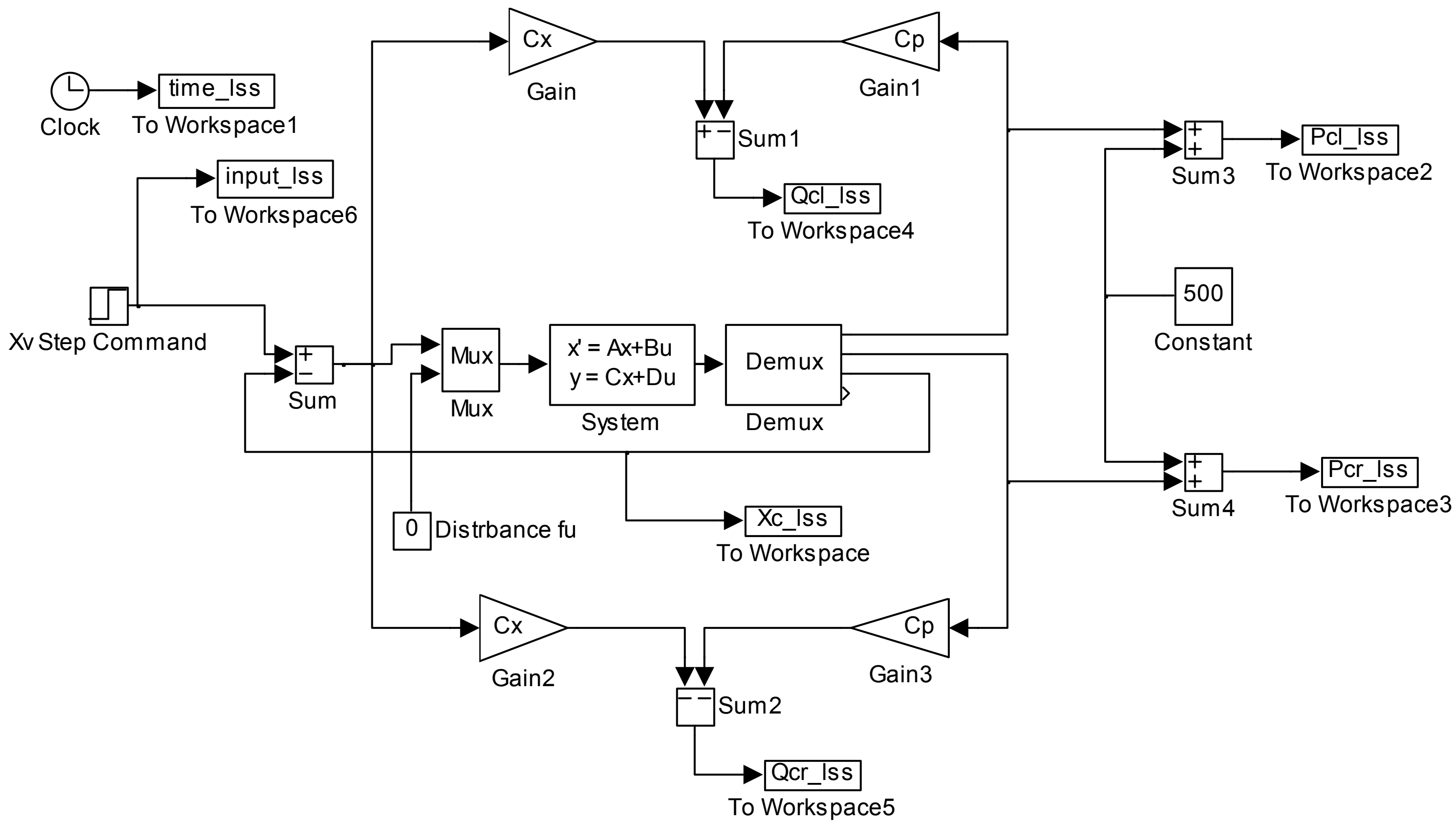
The state variables for the linearized equations are:

$$\begin{bmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \end{bmatrix} = \begin{bmatrix} p_{cl} \\ p_{cr} \\ x_C \\ \dot{x}_C \end{bmatrix}$$

The state-variable equations are:

$$\begin{bmatrix} \dot{q}_1 \\ \dot{q}_2 \\ \dot{q}_3 \\ \dot{q}_4 \end{bmatrix} = \begin{bmatrix} \frac{-\beta\left(C_p + K_{pl}\right)}{V_0} & \frac{K_{pl}\beta}{V_0} & 0 & \frac{-A_p\beta}{V_0} \\ \frac{K_{pl}\beta}{V_0} & \frac{-\beta\left(C_p + K_{pl}\right)}{V_0} & 0 & \frac{A_p\beta}{V_0} \\ 0 & 0 & 0 & 1 \\ \frac{A_p}{M} & \frac{-A_p}{M} & 0 & \frac{-B}{M} \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \end{bmatrix} + \begin{bmatrix} \frac{\beta C_x}{V_0} \\ \frac{-\beta C_x}{V_0} \\ 0 \\ 0 \end{bmatrix} \left[x_v\right]$$

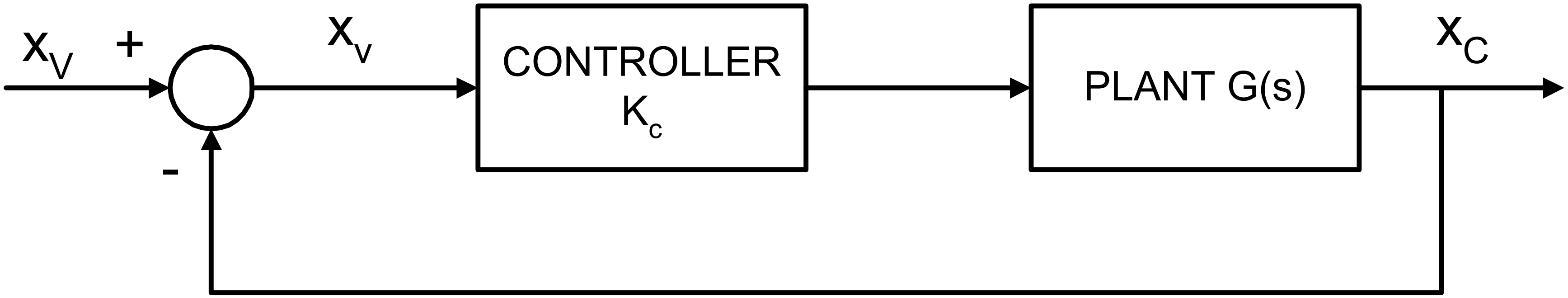
Simulink Block Diagram: Linear System in State-Variable Form



ELECTROHYDRAULIC VALVE-CONTROLLED SERVOMECHANISM (LINEAR - STATE SPACE)

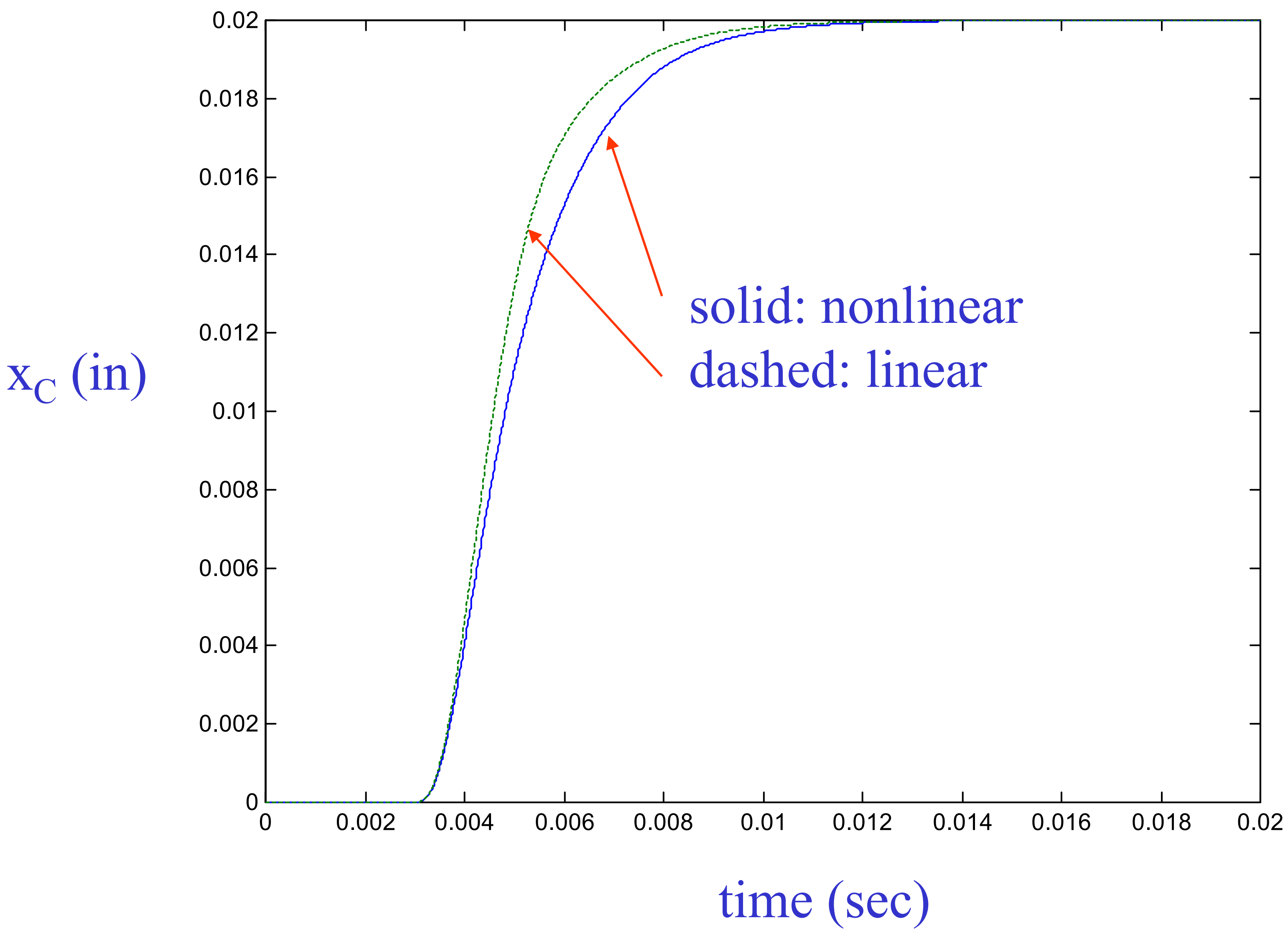
Simulation Results:

Step Command $x_V = 0.02$ in. applied at $t = 0.003$ sec

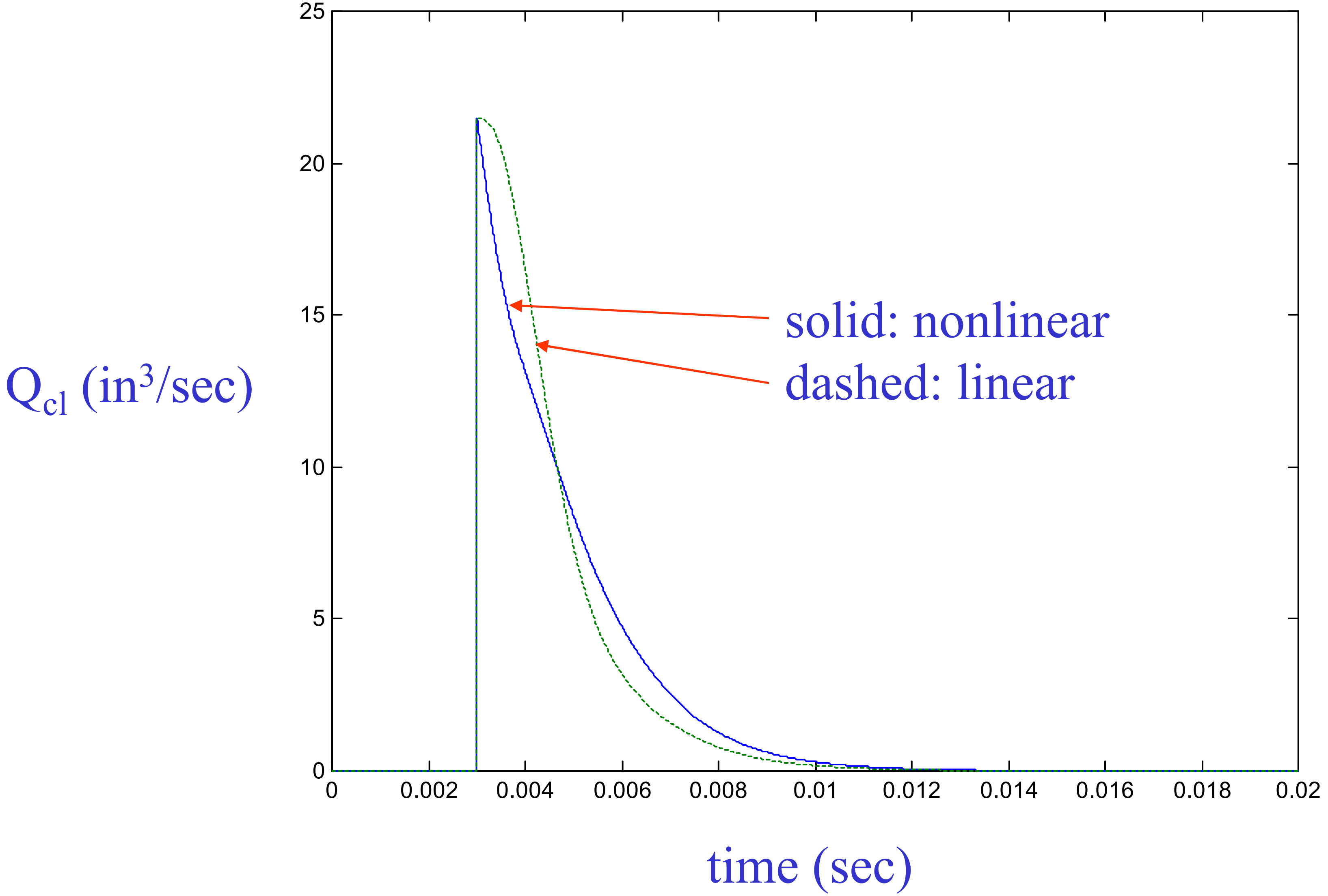


$$\frac{x_C}{x_V} = \frac{K_c G(s)}{1 + K_c G(s)} = \frac{K_c K \omega_n^2}{s^3 + 2\zeta\omega_n s^2 + \omega_n^2 s + K_c K \omega_n^2}$$

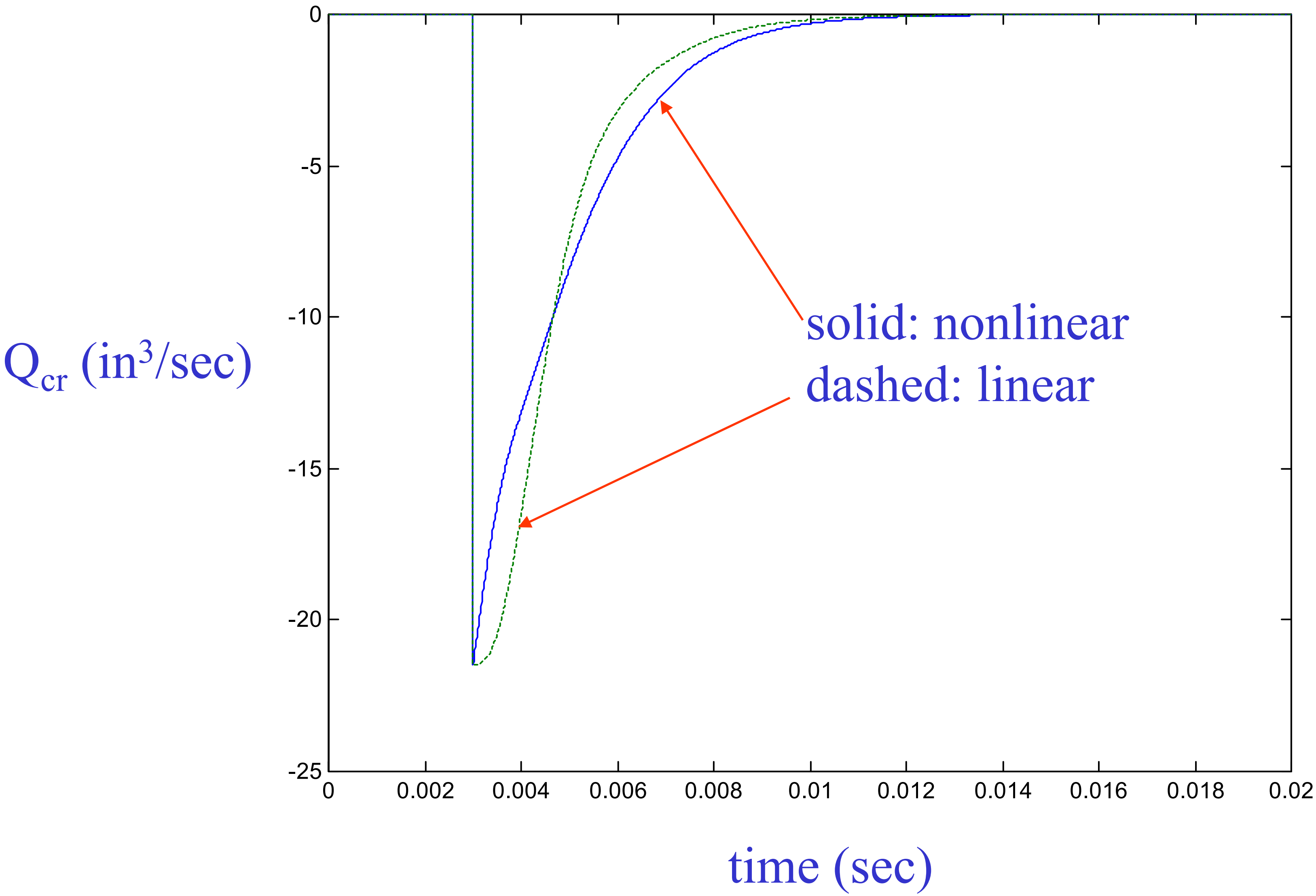
Nonlinear and Linear Simulation Results: x_C vs. time



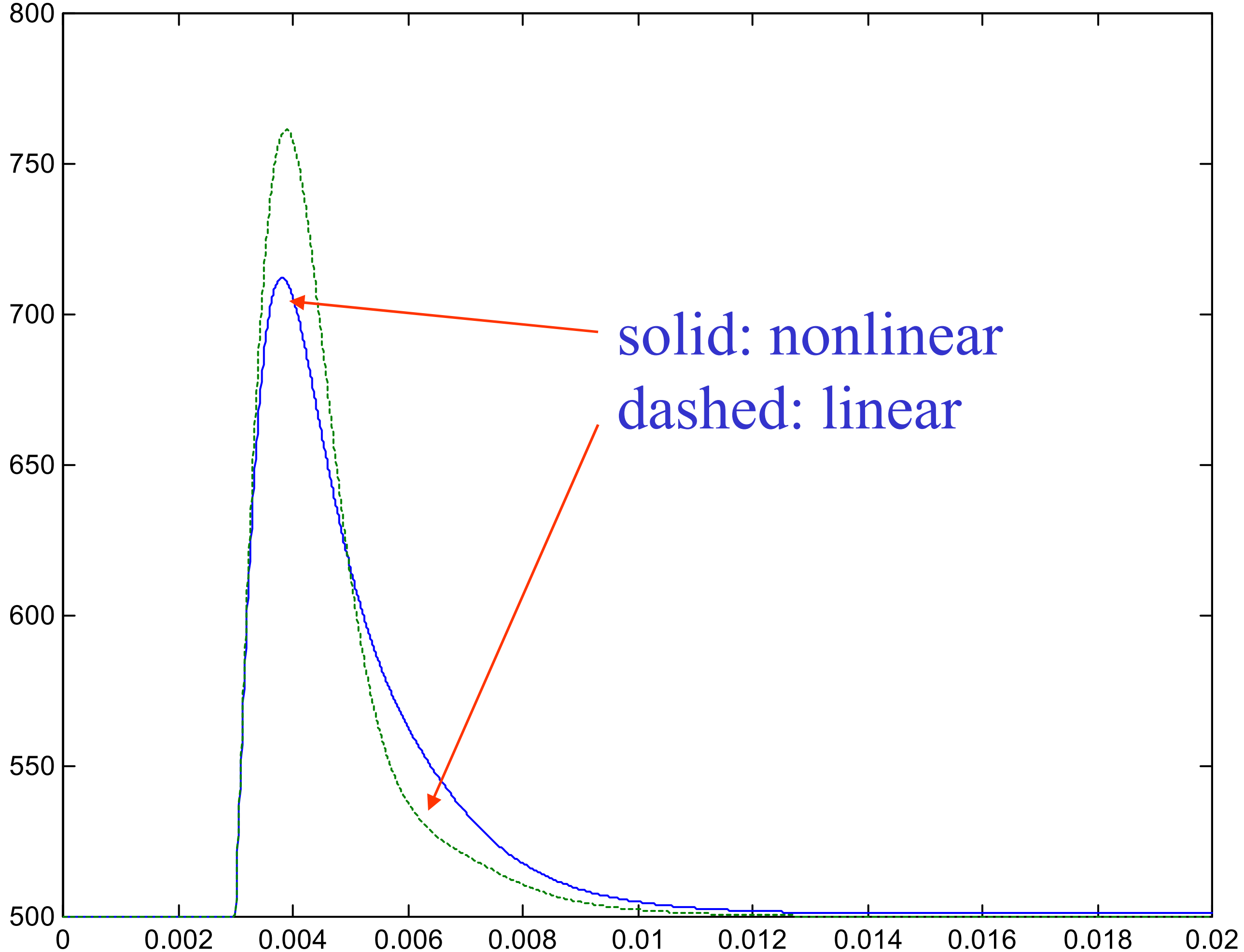
Nonlinear and Linear Simulation Results: Q_{cl} vs. time



Nonlinear and Linear Simulation Results: Q_{cr} vs. time



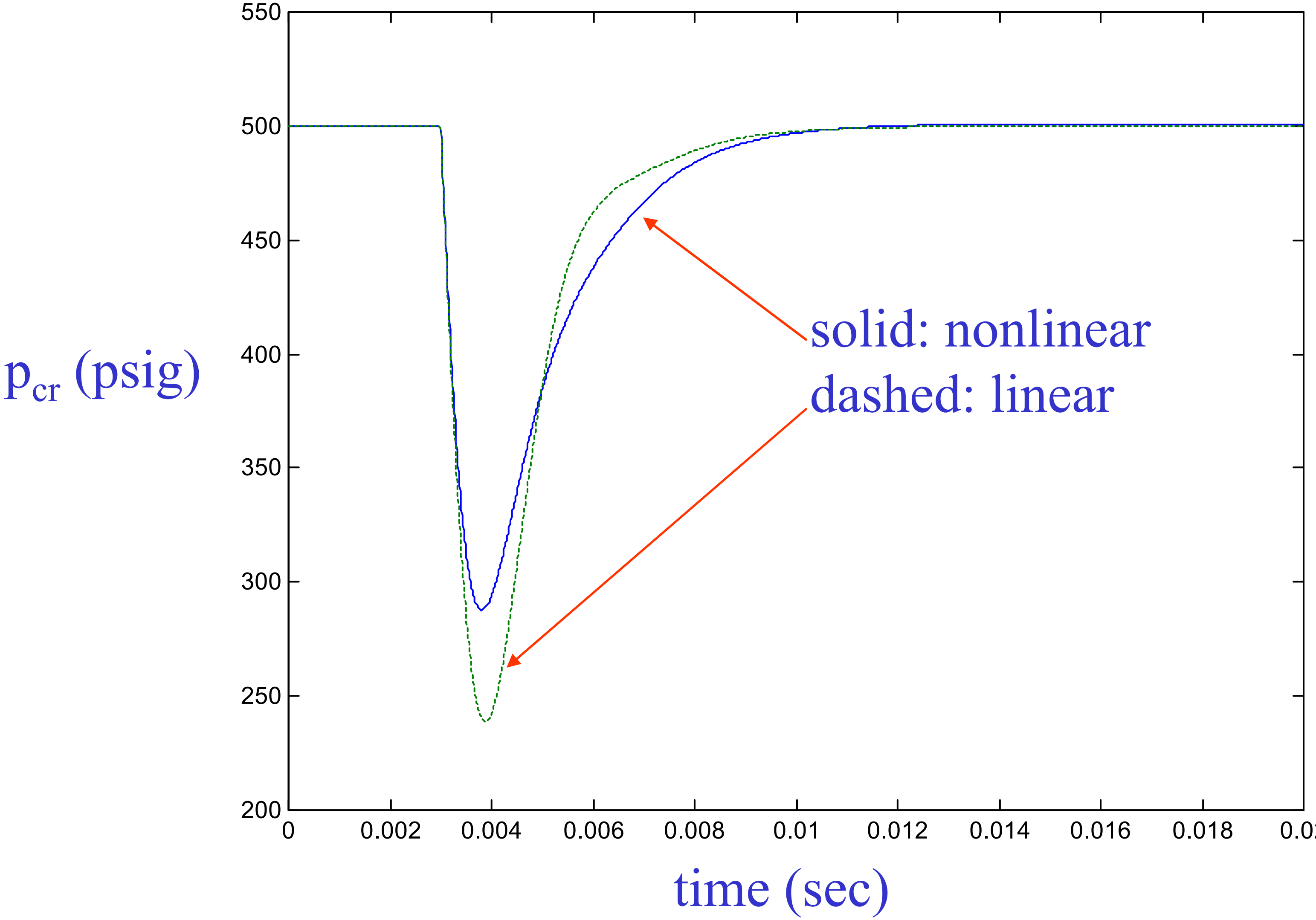
Nonlinear and Linear Simulation Results: p_{cl} time



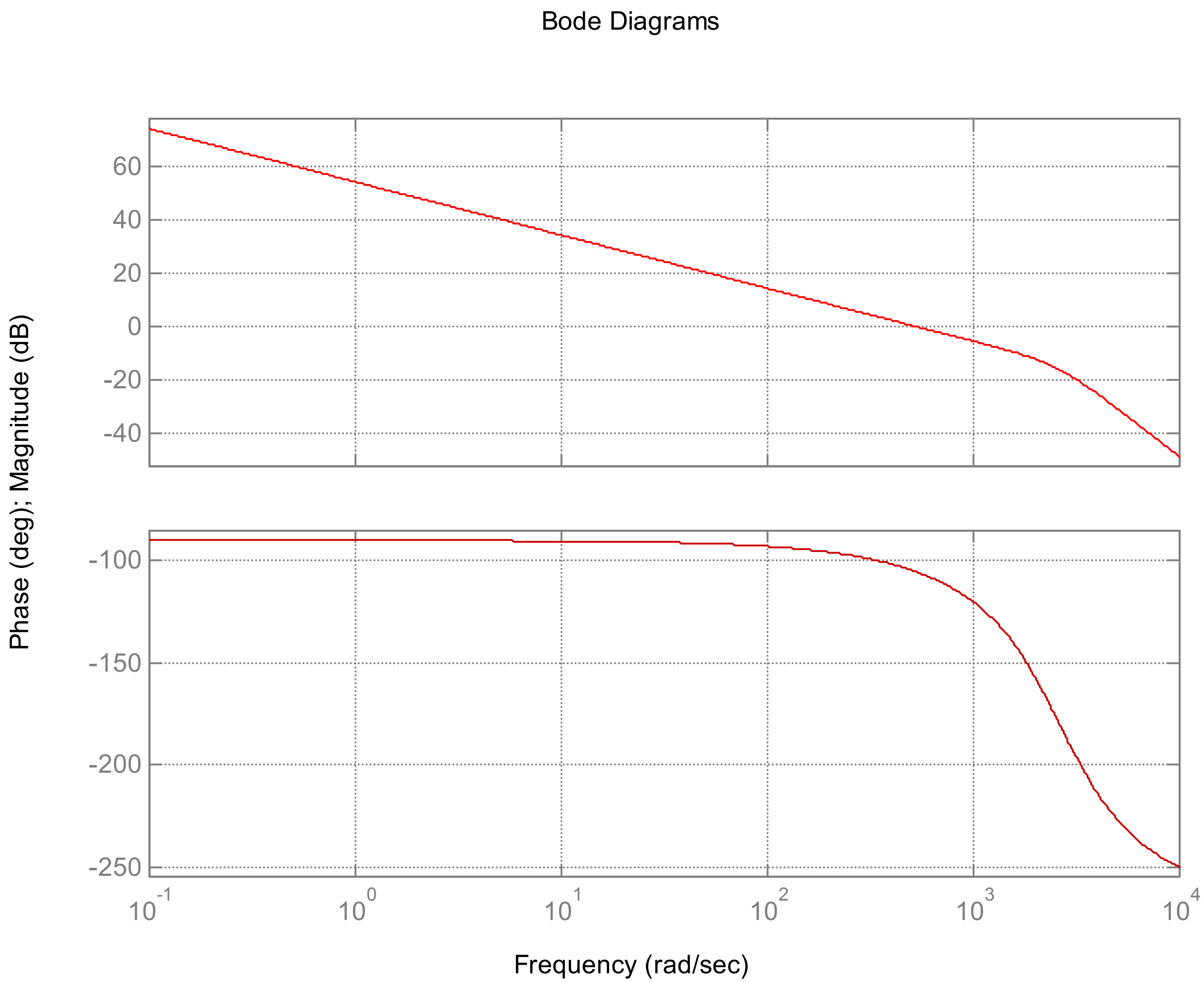
p_{cl} (psig)

time (sec)

Nonlinear and Linear Simulation Results: p_{cr} vs. time

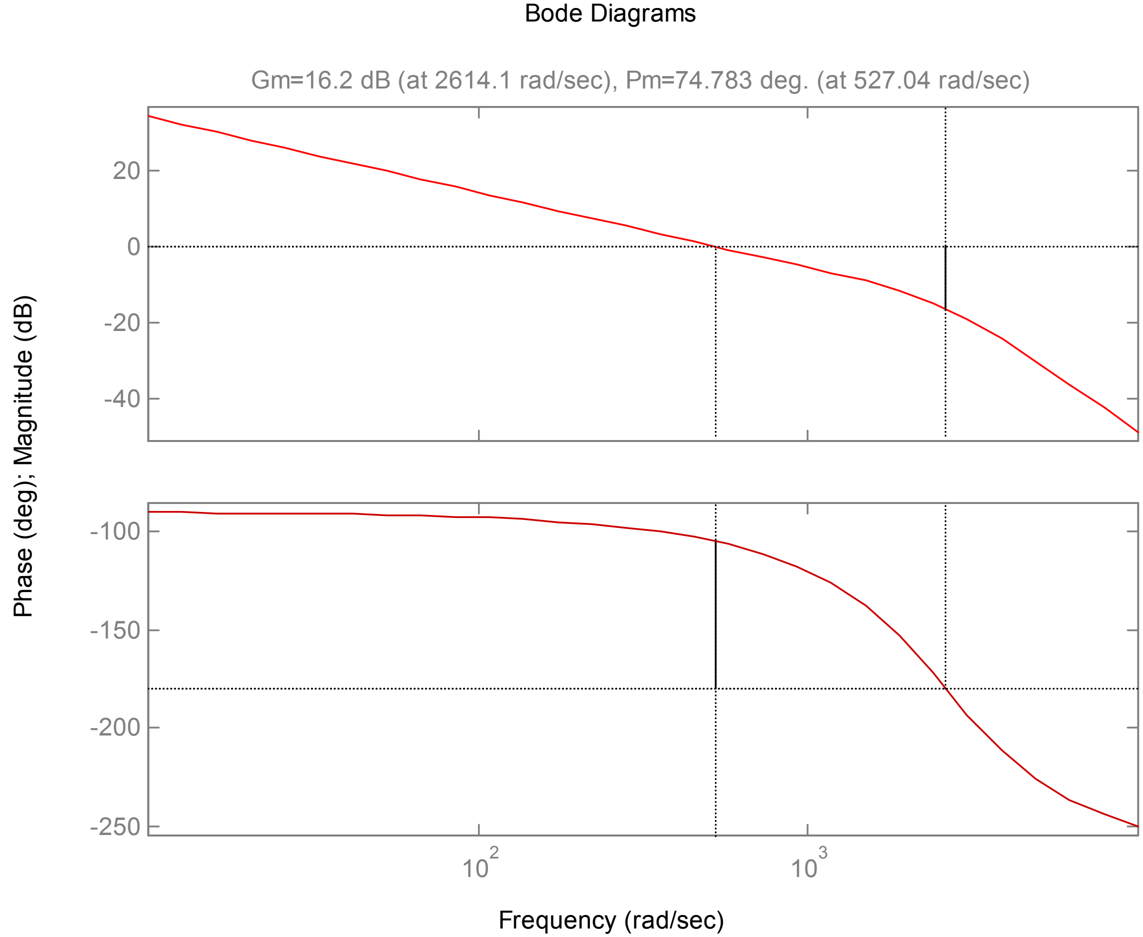


Open-Loop Frequency Response Plots with $K_c = 1$



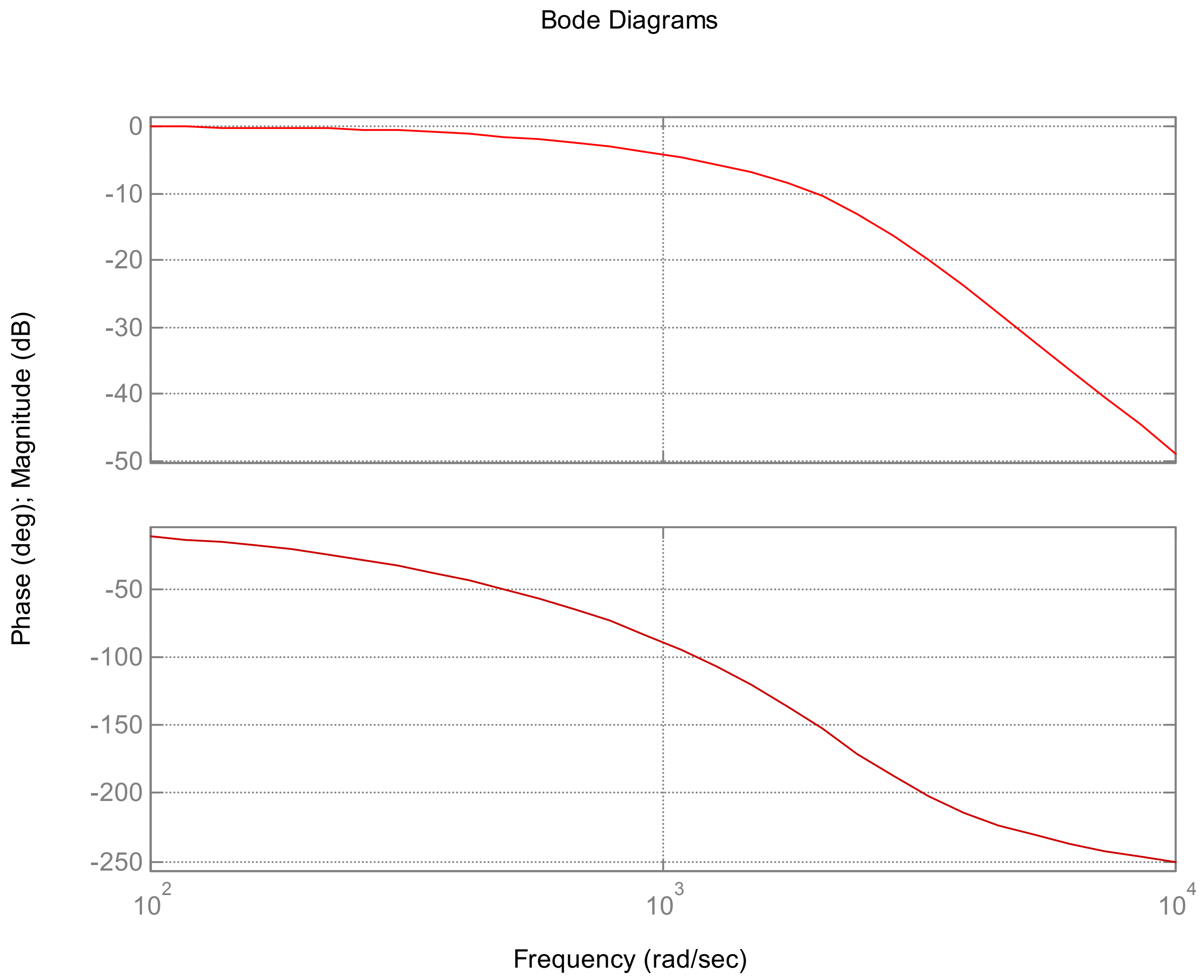
GM = 16.2 dB = 6.46

PM = 74.8°

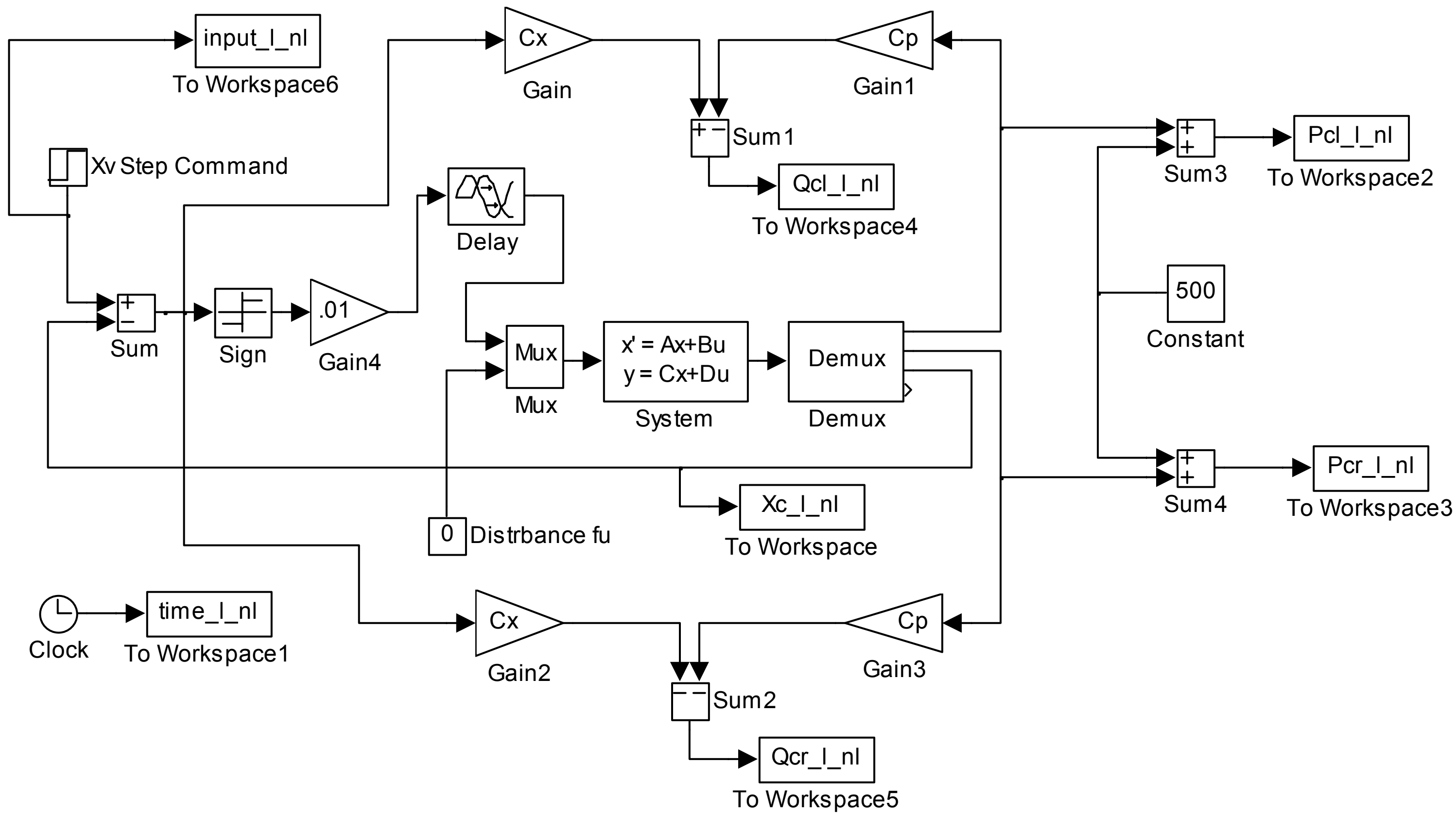


Closed-Loop Frequency Response Plots with $K_c = 1$
Closed-Loop Bandwidth = 123 Hz = 774 rad/sec

At 774 rad/sec:
Mag = 0.707
Phase = -72.2°

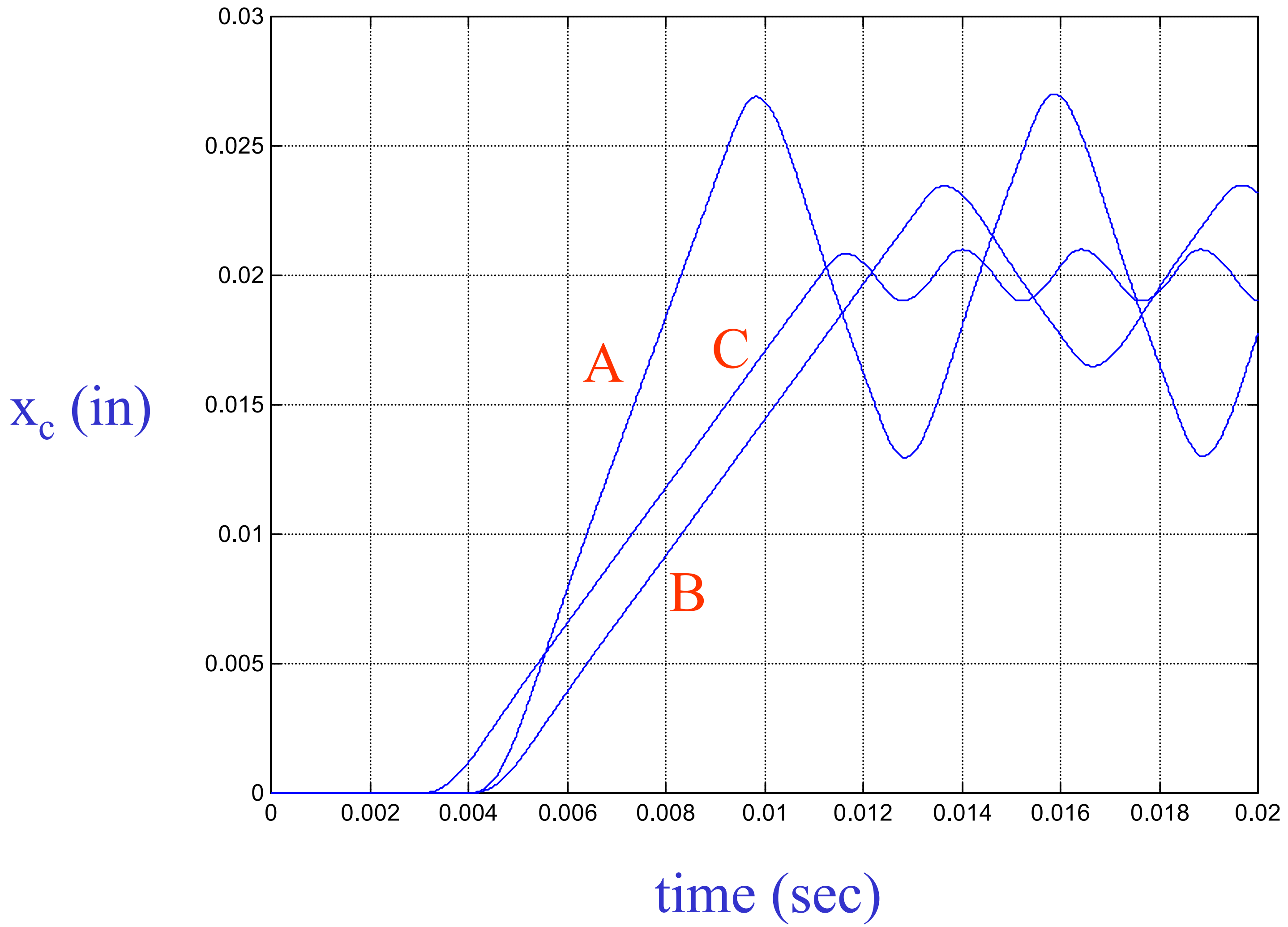


Simulink Block Diagram: Nonlinear Control with Time Delay

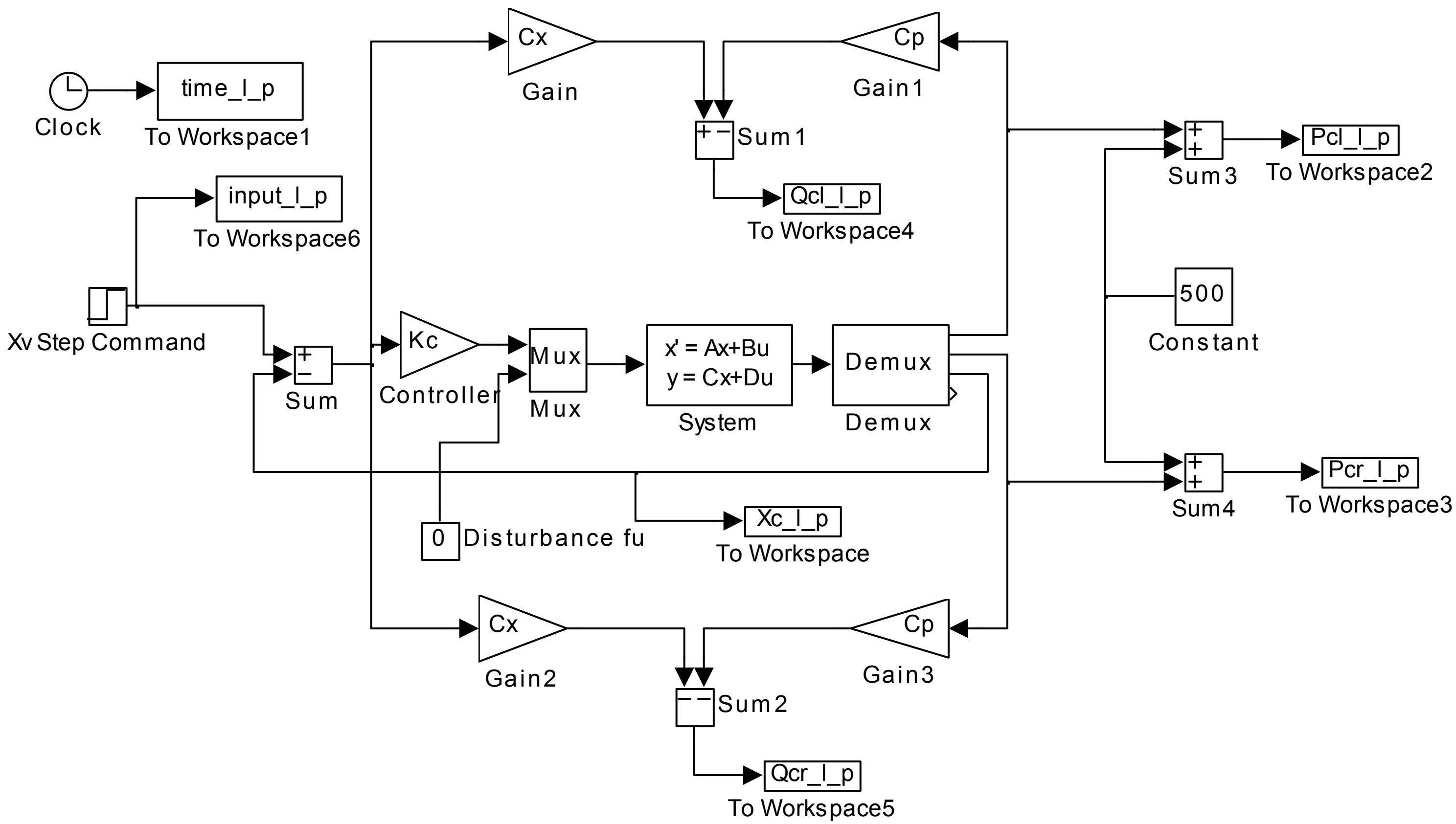


ELECTROHYDRAULIC VALVE-CONTROLLED SERVOMECHANISM (LINEAR)
with Nonlinear On-Off Controller and Time Delay

Curve A: Gain = 0.01, Delay = 0.001 sec
Curve B: Gain = 0.005, Delay = 0.001 sec
Curve C: Gain = 0.005, Delay = 0 sec



Simulink Block Diagram: Linear System with Proportional Control



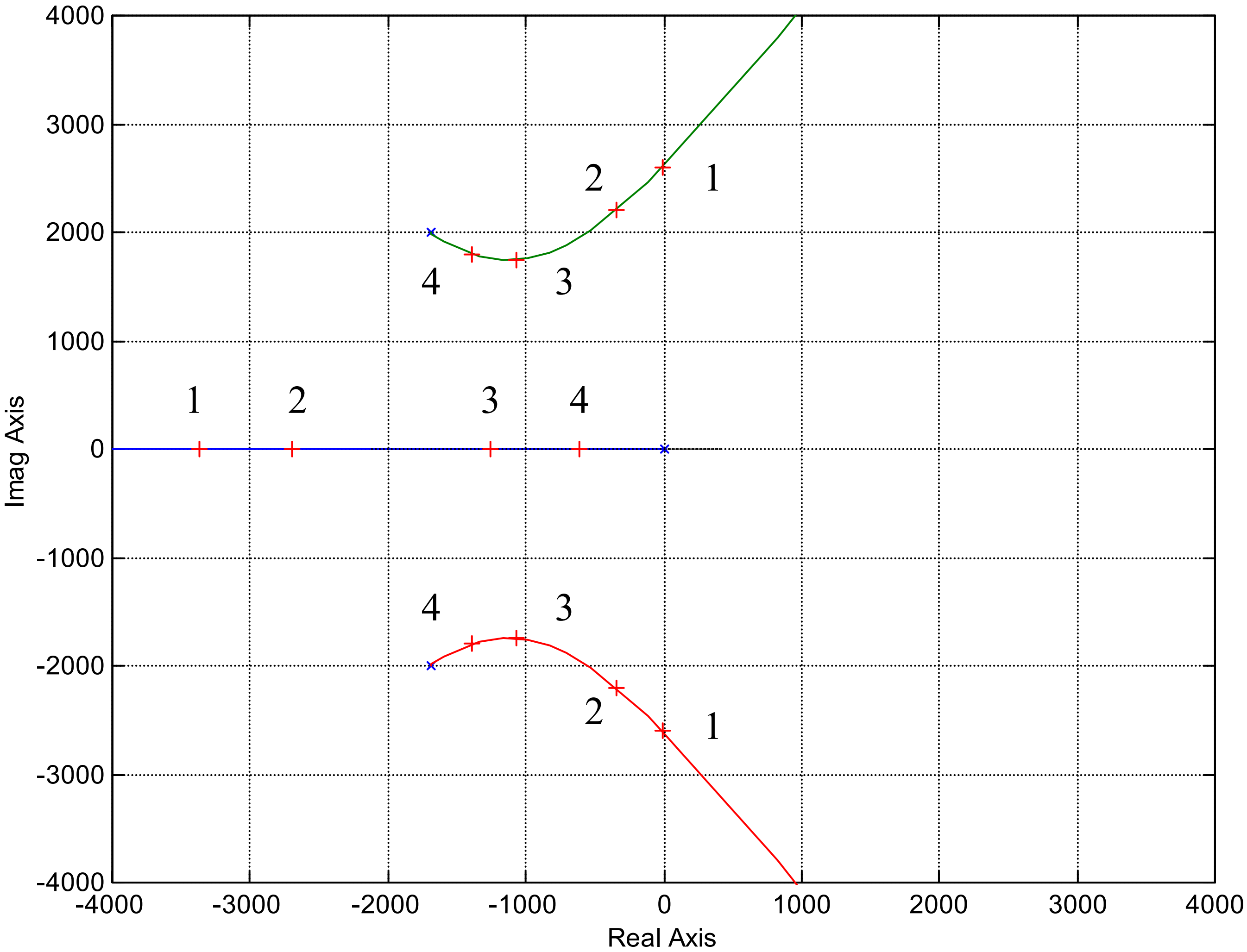
ELECTROHYDRAULIC VALVE-CONTROLLED SERVOMECHANISM (LINEAR)

With Proportional Control

Root Locus Plot

Pt. #1: $K_c = 6.46$ Pt. #2: $K_c = 3.70$

Pt. #3: $K_c = 1.36$ Pt. #4: $K_c = 1$



- Open-Loop Poles: 0, -1692 ± 1993i
- K_c =6.46

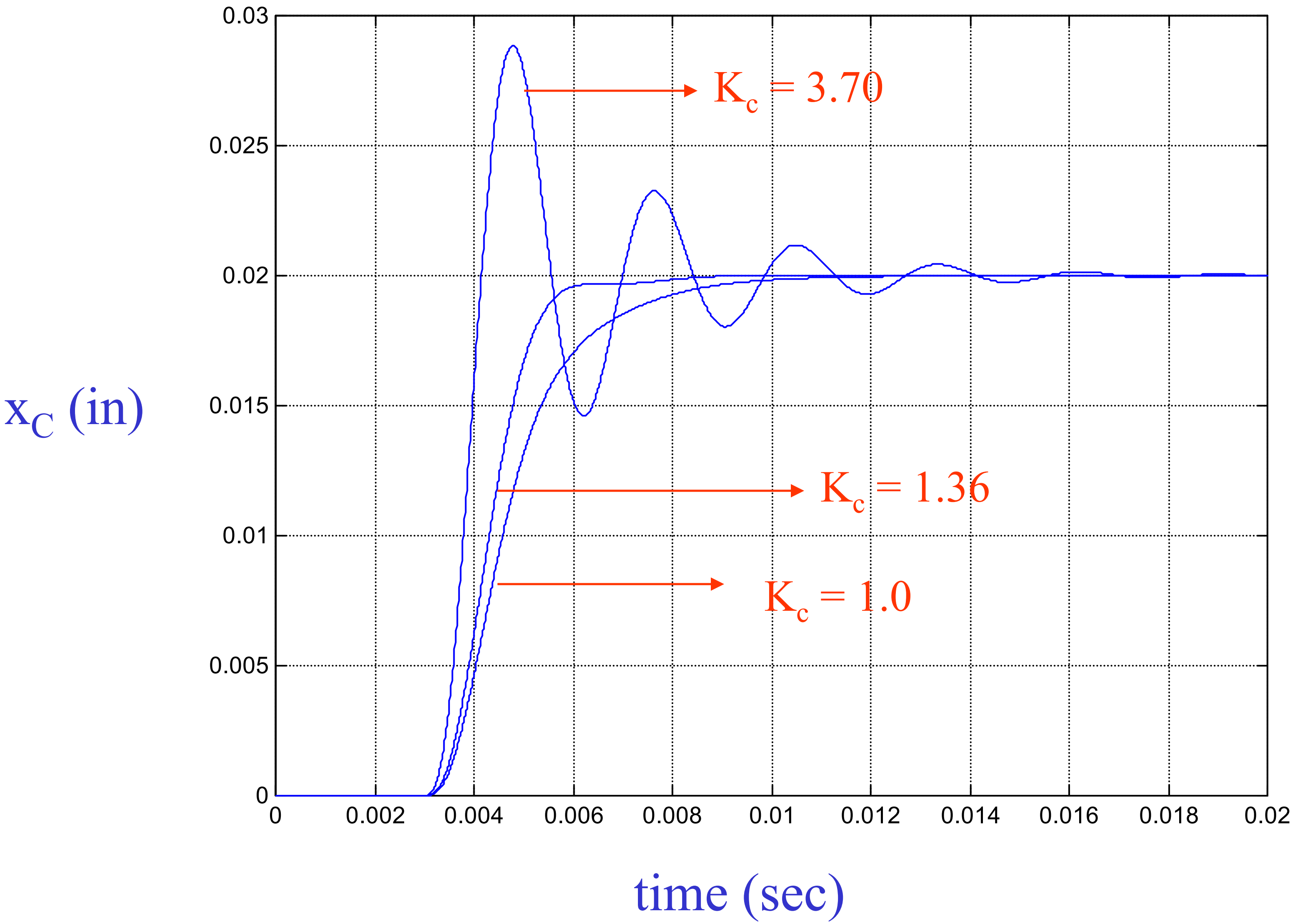
  Closed-Loop Poles: -3384, 0 ± 2614i
- K_c = 3.70

  Closed-loop Poles: -2677, -353 ± 2195i
- K_c = 1.36

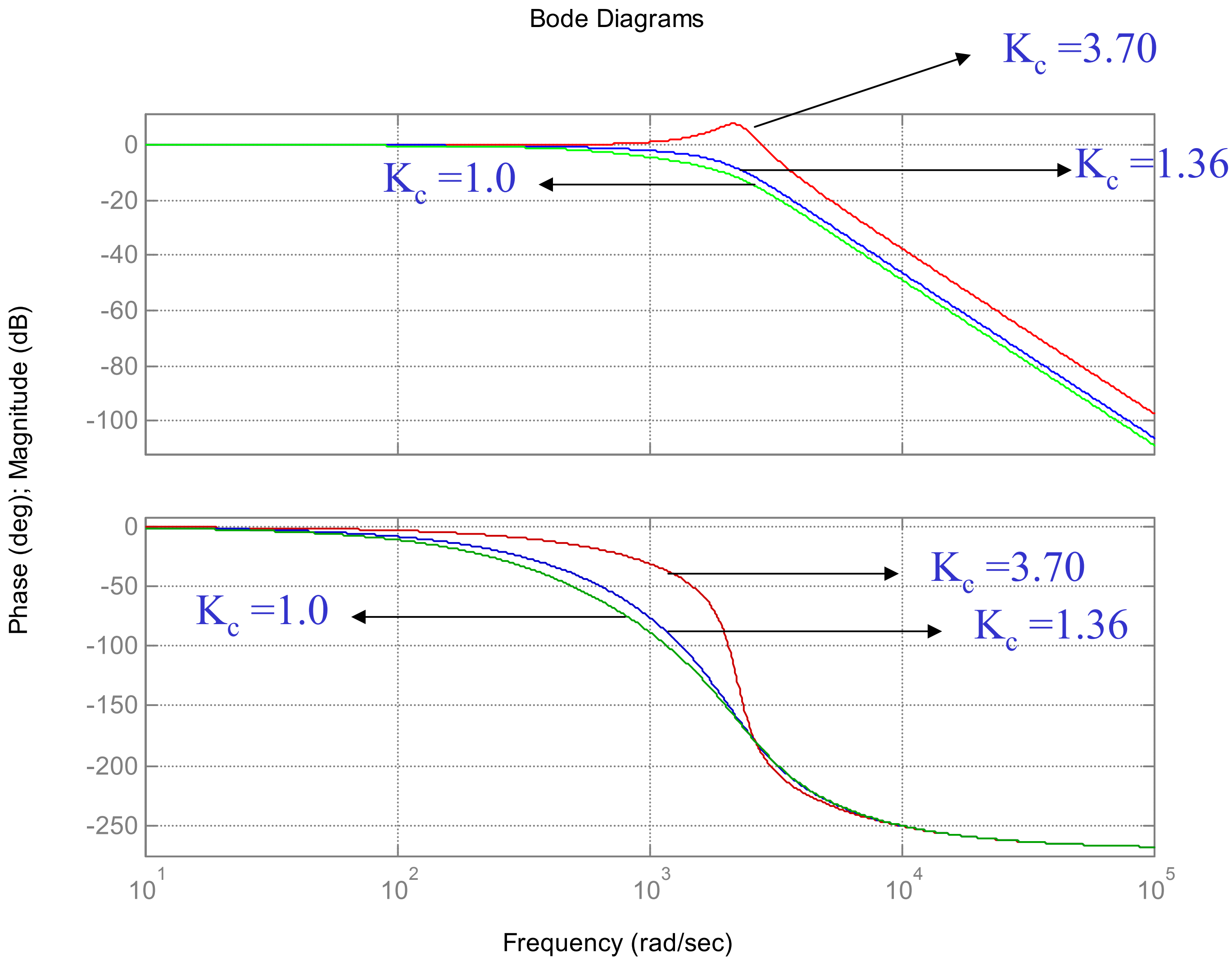
  Closed-Loop Poles: -1133, -1125 ± 1737i
- K_c = 1.0

  Closed-Loop Poles: -732, -1326 ± 1771i

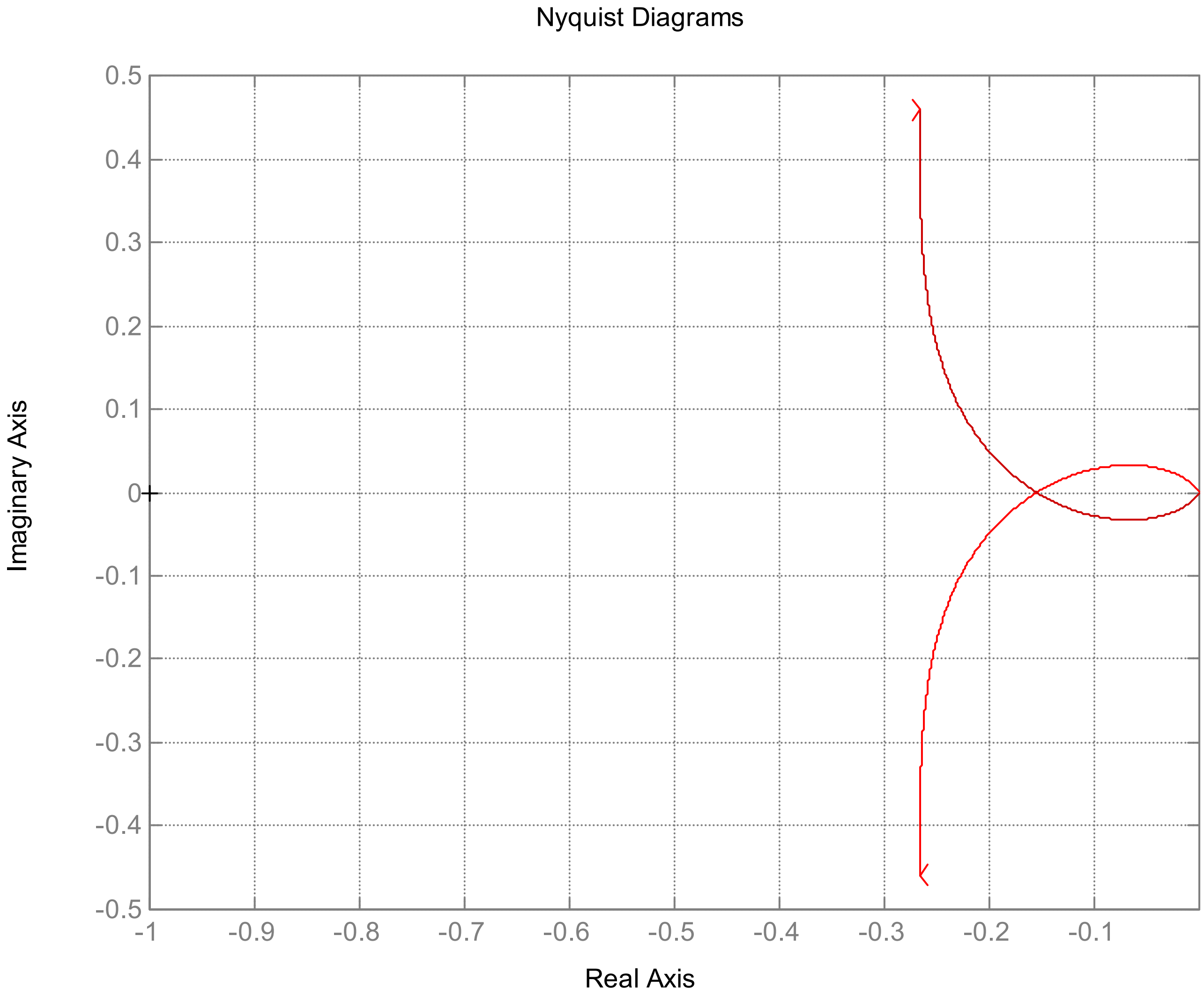
Closed-Loop Time Response (Step) Plots



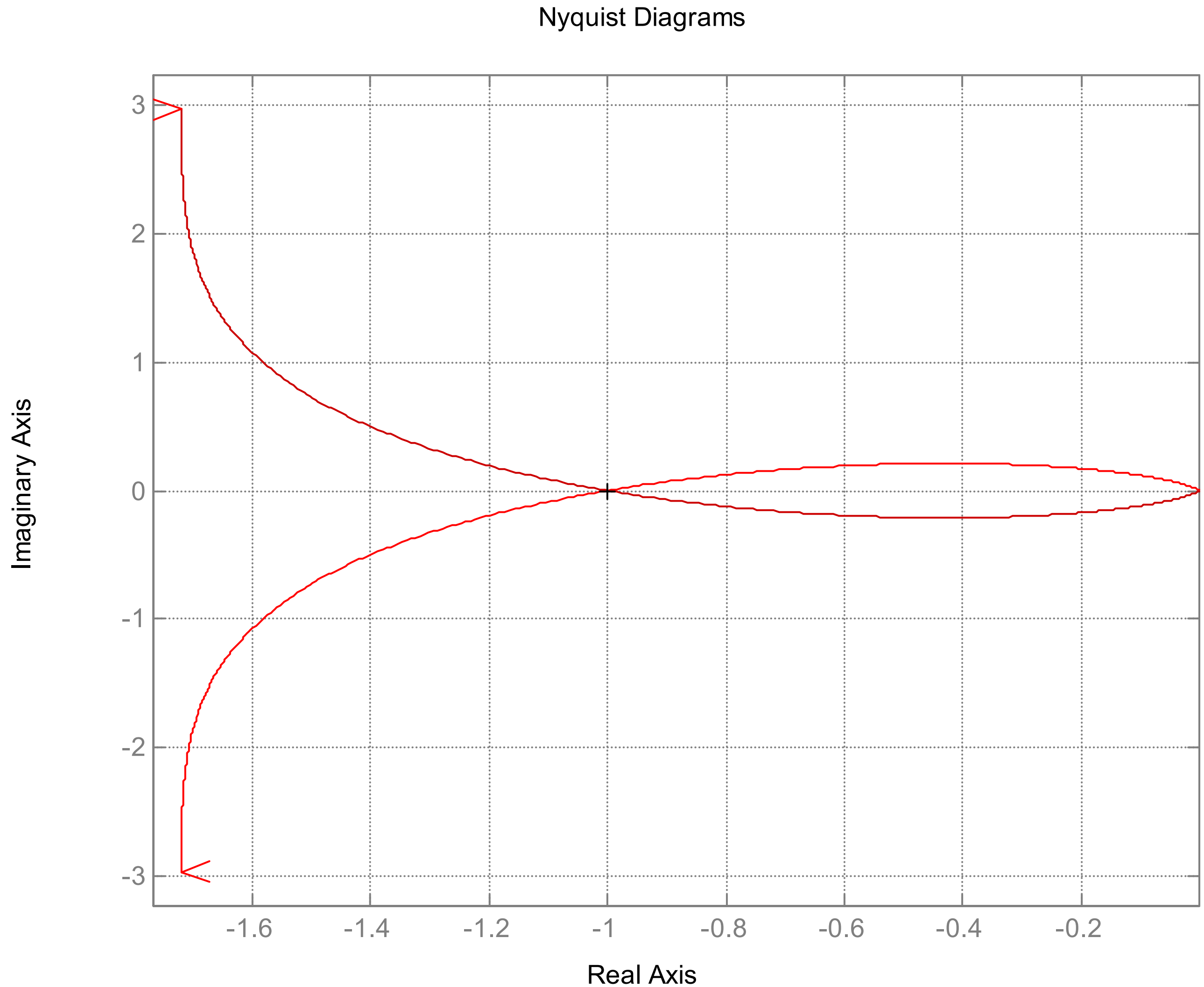
Closed-Loop Frequency Response Plots



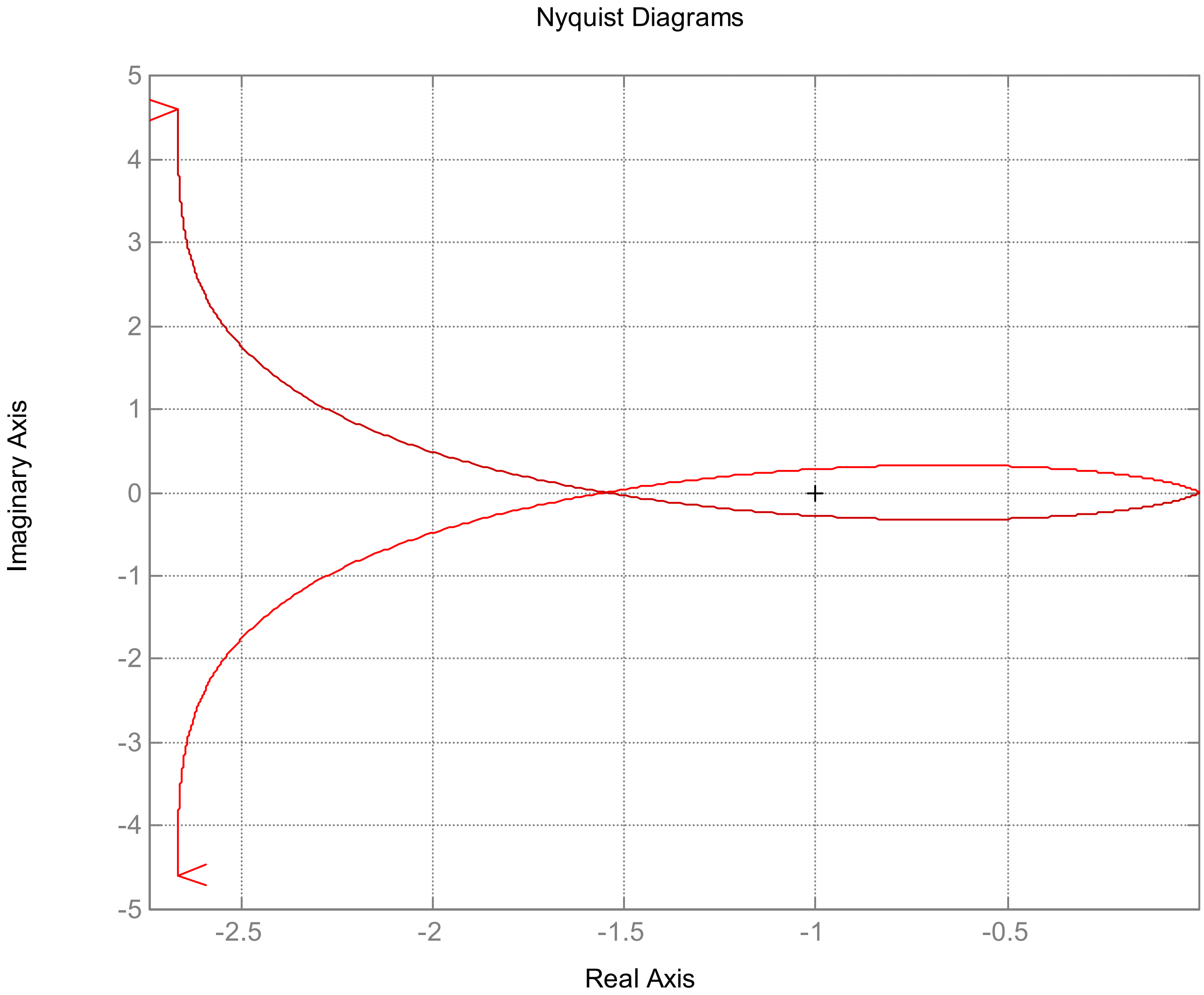
Nyquist Diagram: $K_c = 1.0$



Nyquist Diagram: $K_c = 6.46$



Nyquist Diagram: $K_c = 10.0$



Stability Considerations

<u>Closed-Loop Transfer Function</u>

$$\frac{x_C}{x_V} = \frac{K_c G(s)}{1 + K_c G(s)} = \frac{K_c K \omega_n^2}{s^3 + 2\zeta\omega_n s^2 + \omega_n^2 s + K_c K \omega_n^2}$$

$$K = \frac{2C_x A_p}{2A_p^2 + B\left(C_p + 2K_{pl}\right)}$$

$$\omega_n = \sqrt{\frac{\beta\left[2A_p^2 + B\left(C_p + 2K_{pl}\right)\right]}{MV_0}}$$

$$\zeta = \frac{B + \left(\frac{2\beta M}{V_0}\right)K_{pl} + \left(\frac{\beta M}{V_0}\right)C_p}{2\sqrt{\frac{\beta M}{V_0}\left[2A_p^2 + B\left(C_p + 2K_{pl}\right)\right]}}$$

- Neglect leakage ($K_{pl} = 0$) and consider the load as mainly inertia (B = 0, friction is ignored). The closed-loop transfer function becomes:

$$\frac{x_C}{x_V} = \frac{\dfrac{K_c C_x}{A_p}}{\dfrac{V_0 M}{2\beta A_p^2}s^3 + \dfrac{MC_p}{2A_p^2}s^2 + s + \dfrac{K_c C_x}{A_p}}$$

- Since the bulk modulus β of the fluid is defined as:

$$\beta = -\frac{\partial P}{\partial V / V_0}$$

- The combined stiffness k_0 of the two columns of fluid is:

$$k_0 = \frac{2\beta A_p^2}{V_0}$$

- The valve stiffness k_v is defined as:

$$C_x = \left.\frac{\partial Q_v}{\partial x_v}\right|_{\text{operating point}}$$

$$C_p = -\left.\frac{\partial Q_v}{\partial p_c}\right|_{\text{operating point}}$$

$$k_v = 2A_p \frac{C_x}{C_p}$$

- The closed-loop transfer function can now be written as:

$$\frac{x_C}{x_V} = \frac{\dfrac{K_c C_x}{A_p}}{\dfrac{M}{k_0}s^3 + \dfrac{C_x M}{A_p k_v}s^2 + s + \dfrac{K_c C_x}{A_p}}$$

- Applying the Routh Stability Criterion to the characteristic equation of the closed-loop transfer function gives the relationship for stability as:

$$k_0 > k_v$$

- In other words, the stiffness of the oil column must be greater than the effective valve stiffness if stability is to be satisfactory.

Sensors & Actuators in Mechatronics

MEAE 6960
Summer 2002

Assignment # 4

• Problem # 1

– For the electrohydraulic valve-controlled servomechanism discussed in class, modify the MatLab/Simulink nonlinear simulation:
  - to include the effect of valve overlap / underlap
  - to include the effect of valve-opening saturation, i.e., the valve has a finite full stroke in addition to a port width
  - to include the effect of Coulomb friction (as discussed in class)

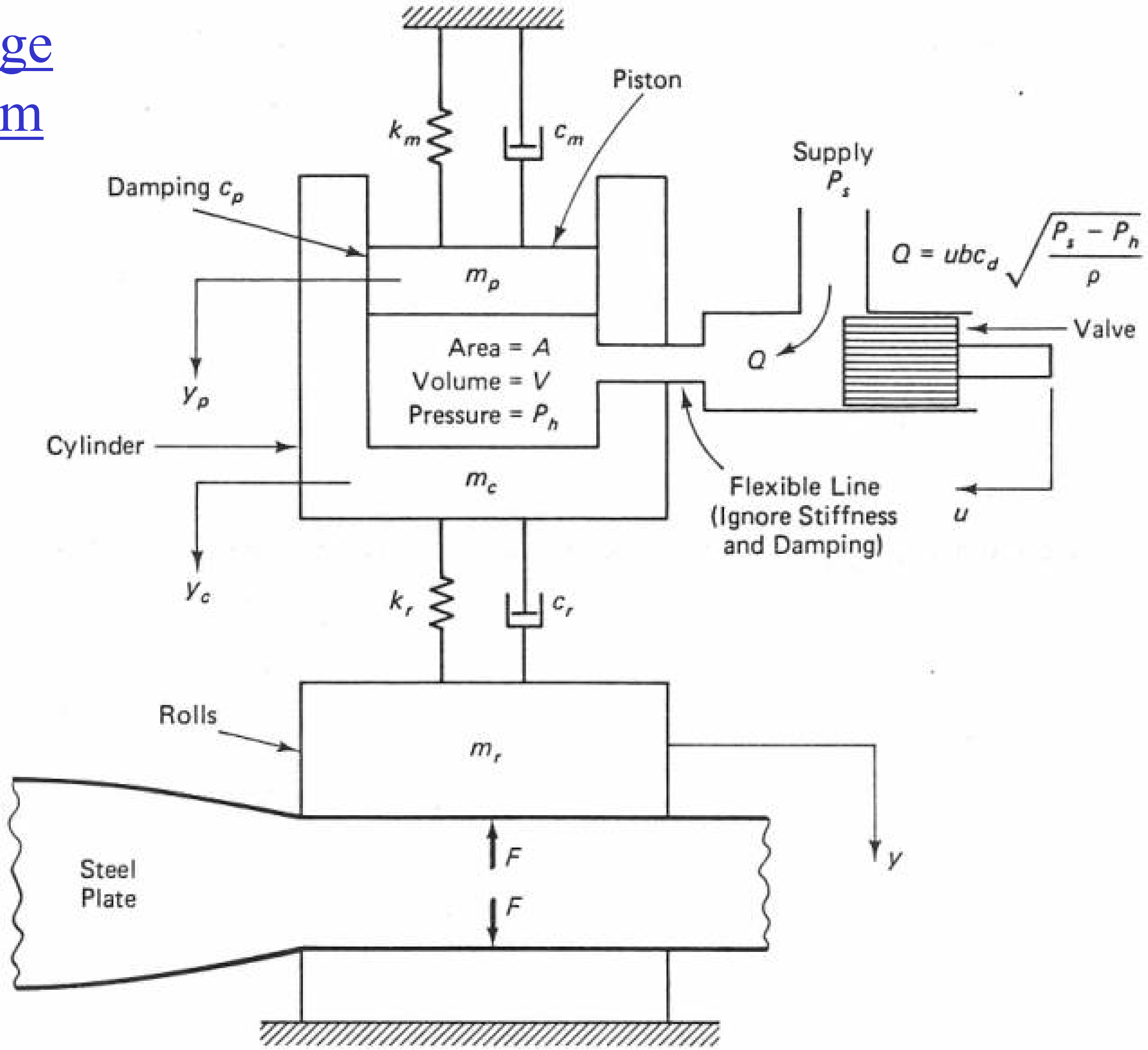
– Run simulations to demonstrate the modifications.

– Comment on the effects of valve overlap / underlap (e.g., 5% of full stroke), valve-opening saturation (e.g., 0.03 inches full stroke), and Coulomb friction on the system dynamic performance in response to a step position command.

- **Problem # 2**
  - A model for the automatic gage control system of a steel rolling mill is shown. The rolls are pressed using a single-acting hydraulic actuator with valve displacement u. The rolls are displaced through y, thereby pressing the steel that is being rolled. The rolling force F is completely known from the steel parameters for a given y.
    - Identify the inputs and the controlled variable in this control system.
    - In terms of the variables and system parameters, write dynamic nonlinear equations for the system.

- What is the order of the system? Identify the response variables.
- Draw a block diagram fro the system, clearly indicating the hydraulic actuator with valve, the mill structure, inputs, and the controlled variable.
- What variables would you measure (and feed back through suitable controllers) in order to improve the performance of the control system?

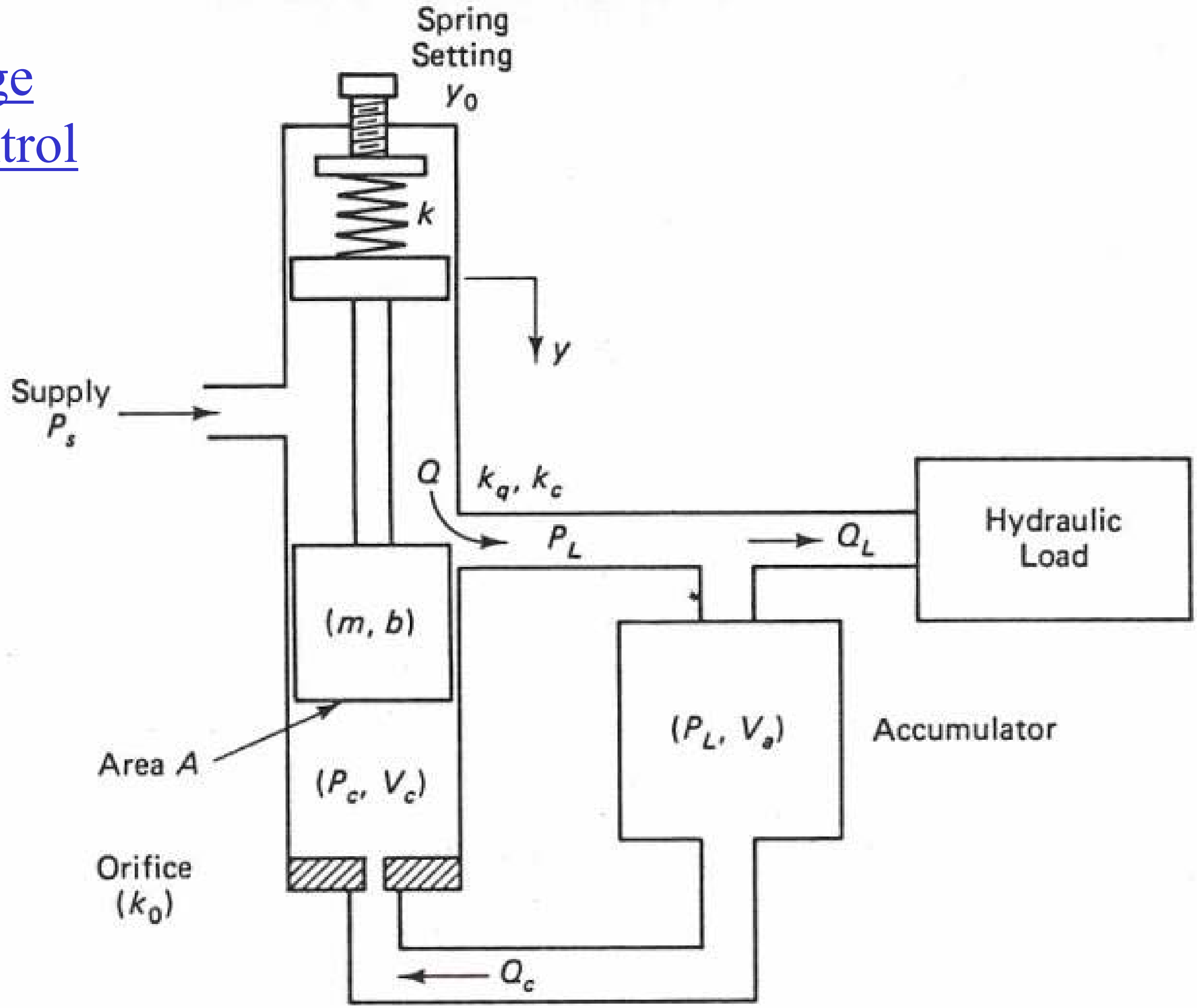
Automatic Gage Control System



- **Problem # 3**
  - A single-stage pressure control valve is shown. The purpose of the valve is to keep the load pressure P_L constant. Volume rates of flow, pressures, and the volumes of fluid subjected to those pressures are indicated. The mass of the spool and appurtenances is m, the damping constant of the damping force acting on the moving parts is b, and the effective bulk modulus of oil is β. The accumulator volume is V_a. The flow into the valve chamber (volume V_c) is through an orifice. This flow may be taken as proportional to the pressure drop across the orifice, the constant of proportionality being k_o. A compressive spring of stiffness k restricts the spool motion. The initial spring force is set by adjusting the initial compression y_0 of the spring.

- Identify the reference input, the primary output, and a disturbance input for the valve system.
- By making linearization assumptions and introducing any additional parameters that might be necessary, write equations to describe the system dynamics.
- Set up a block diagram for the system, showing various transfer functions.

Single-Stage Pressure Control Valve

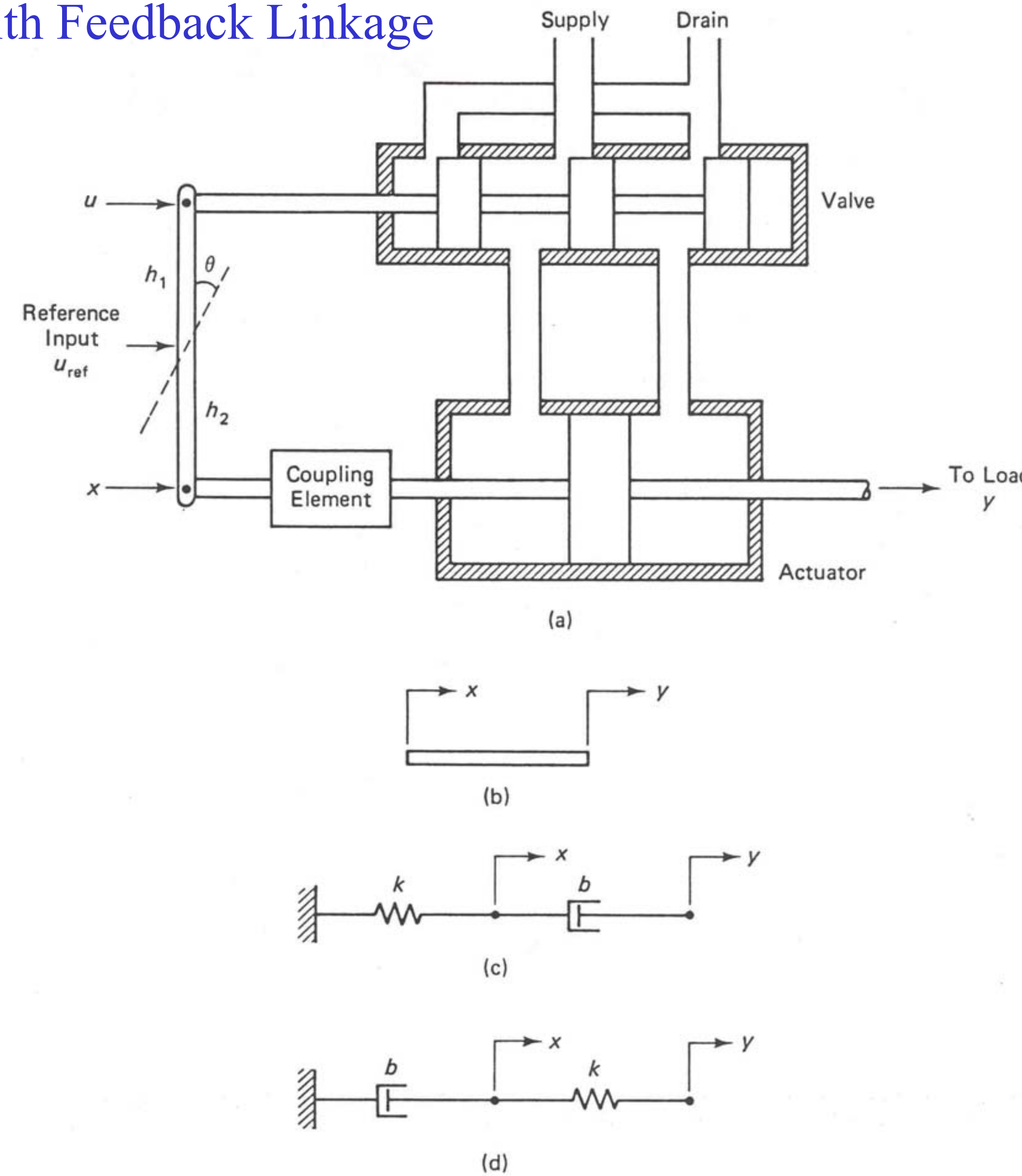


• Problem # 4

– A mechanical linkage is employed as the feedback device for a servovalve of a hydraulic actuator, as shown. The reference input is u_{ref}, the input to the servovalve is u, and the displacement (response) of the actuator piston is y. A coupling element is used to join one end of the linkage to the piston rod. The displacement at this location of the linkage is x.

  - Show that rigid coupling gives proportional feedback action.
  - If a viscous damper (damping constant b) is used as the coupling element and if a spring (stiffness k) is used to externally restrain the coupling end of the linkage, show that the resulting feedback action is lead compensation.
  - If the viscous damper (damping constant b) and the spring (stiffness k) are interchanged, what is the resulting feedback control action?

(a) Servovalve and Actuator with Feedback Linkage
(b) Rigid Coupling
(c) Damper-Spring Coupling
(d) Spring-Damper Coupling



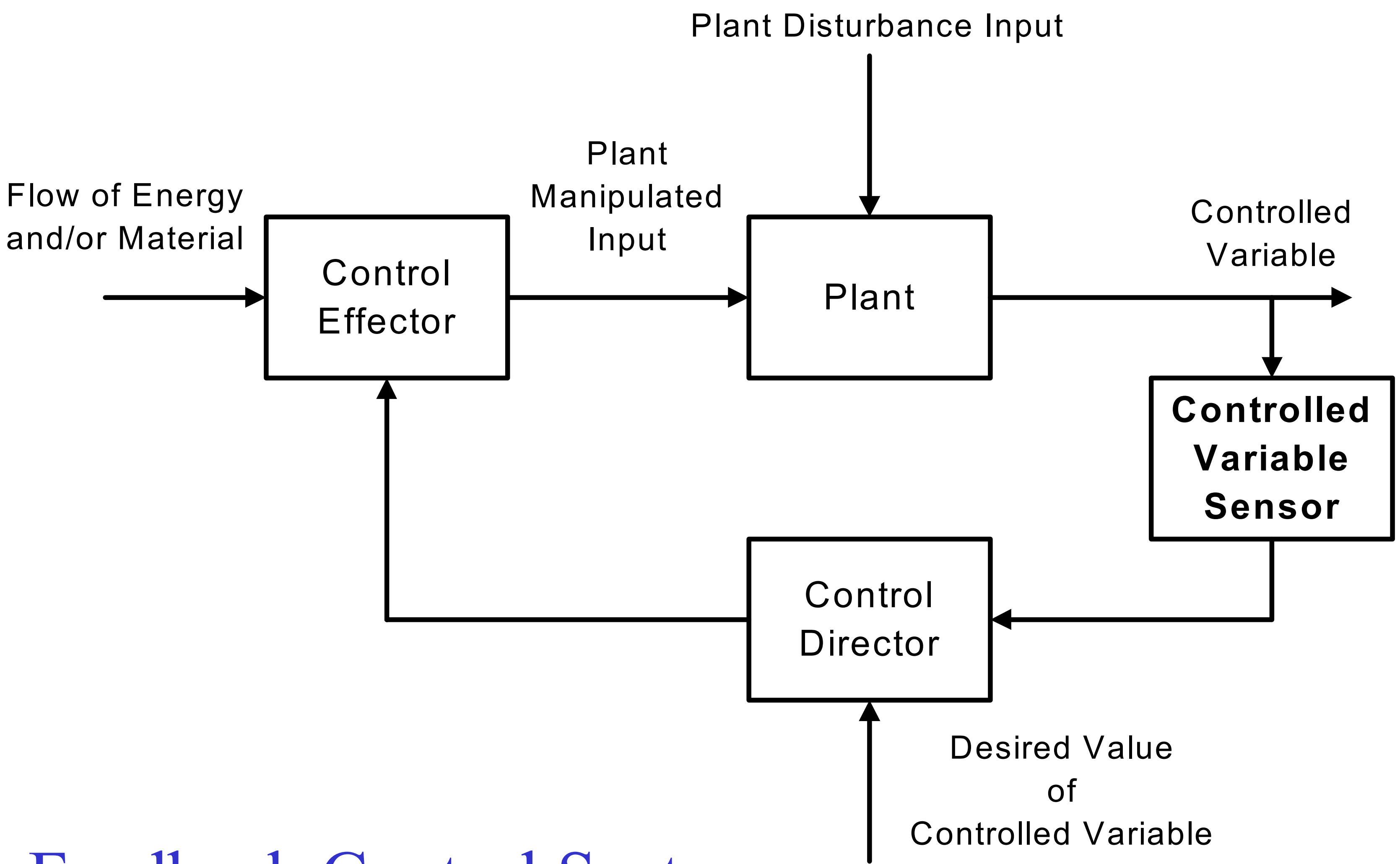
Introduction to Sensors

- Types of Applications of Measurement Instrumentation
- Generalized Configurations and Functional Descriptions of Measuring Instruments

Types of Applications of Measurement Instrumentation

- Monitoring of Processes and Operations
- Control of Processes and Operations
- Experimental Engineering Analysis

- <u>Monitoring of Processes and Operations</u>
  - Certain applications of measuring instruments may be characterized as having essentially a monitoring function, e.g., thermometers, barometers, and water, gas, and electric meters.
- <u>Control of Processes and Operations</u>
  - An instrument can serve as a component of a control system. To control any variable in a feedback control system, it is first necessary to measure it. A single control system may require information from many measuring instruments, e.g., industrial machine and process controllers, aircraft control systems.



Feedback Control System

- Experimental Engineering Analysis
  - In solving engineering problems, two general methods are available: theoretical and experimental. Many problems require the application of both methods and theory and experiment should be thought of as complimenting each other.
  - Features of Theoretical Methods
    - Often gives results that are of general use rather than for restricted application.
    - Invariably require the application of simplifying assumptions. The theoretically predicted behavior is always different from the real behavior, as a physical/mathematical model is studied rather than the actual physical system.
    - In some cases, may lead to complicated mathematical problems.

- Require only pencil, paper, computers, etc. Extensive laboratory facilities are not required.
- No time delay engendered in building models, assembling and checking instrumentation, and gathering data.

– Features of Experimental Methods

- Often gives results that apply only to the specific system being tested. However, techniques such as dimensional analysis may allow some generalization.
- No simplifying assumptions necessary if tests are run on an actual system. The true behavior of the system is revealed.
- Accurate measurements necessary to give a true picture. This may require expensive and complicated equipment. The characteristics of all the measuring and recording equipment must be thoroughly understood.
- Actual system or a scale model required. If a scale model is used, similarity of all significant features must be preserved.
- Considerable time required for design, construction, debugging.

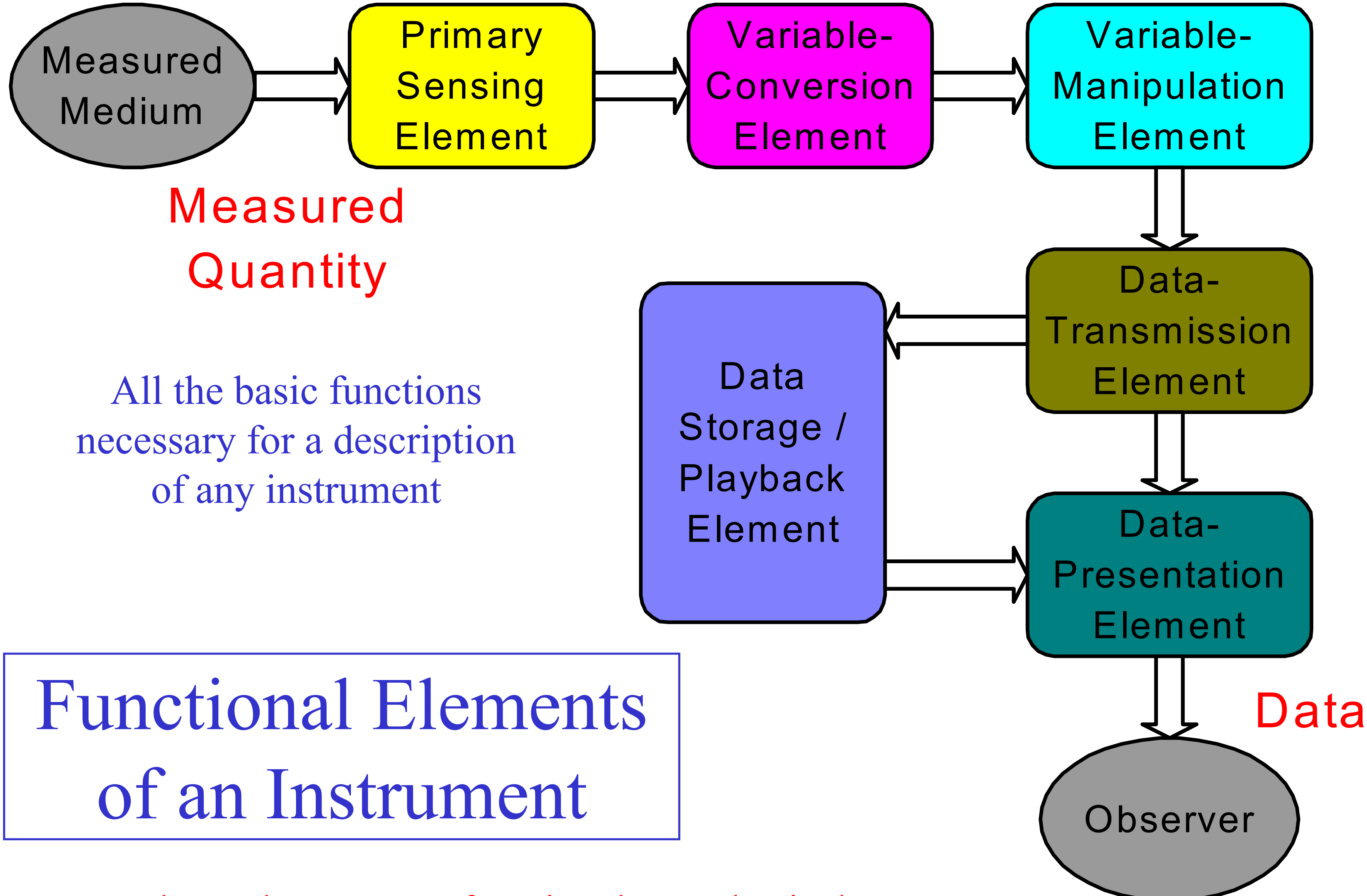
- Note
  - The distinction among monitoring, control, and analysis functions is not clear-cut; the category decided on may depend somewhat on your point of view and the apparent looseness of the classification should not cause any difficulty.
  - It is now extremely common for engineers to include in the design of a machine or process, as dedicated components, computers of various sizes. Computers are important, but the computer "component" of many machines and processes is often *not* the critical system element in terms of either technical or economic factors. Rather, components external to the computer, the actuators and sensors, are more often the limiting factors in the system design.

Generalized Configurations and Functional Descriptions of Measuring Instruments

- It is desirable to describe both the operation and performance of measuring instruments and associated equipment in a generalized way without recourse to specific physical hardware.
- Here we focus on the operation which can be described in terms of the functional elements of an instrument or instrument system.

- By concentrating on these functions and the various physical devices available for accomplishing them, we develop our ability to synthesize new combinations of elements leading to new and useful instruments.

Functional Elements of an Instrument



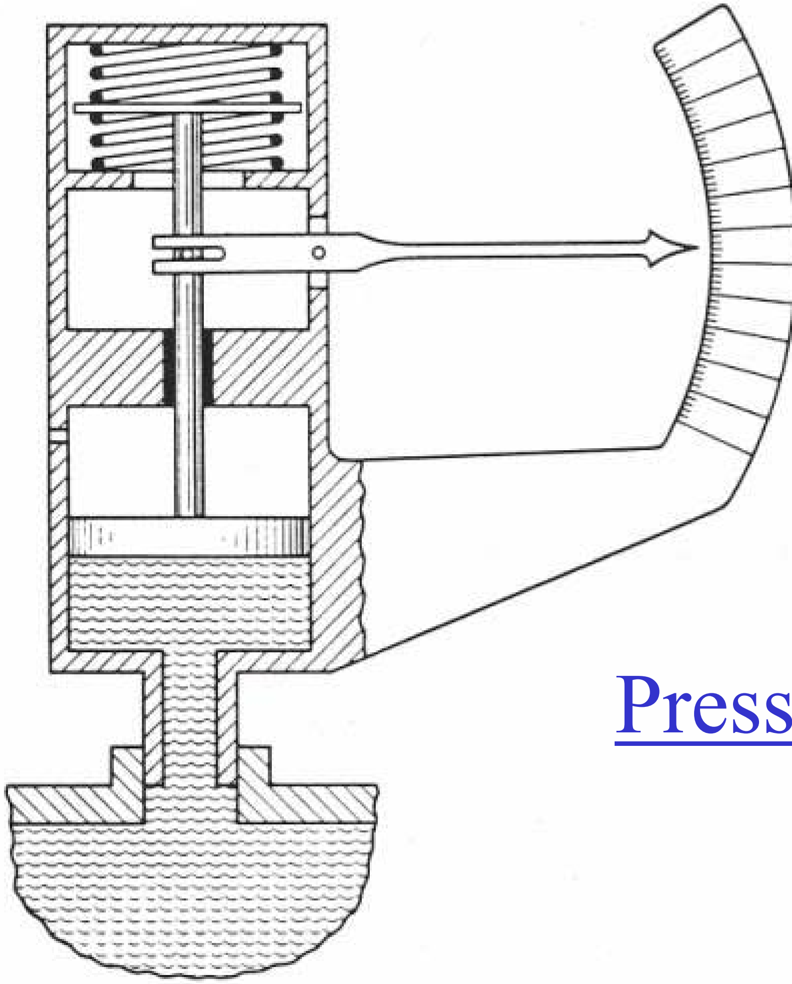
All the basic functions necessary for a description of any instrument

Note: These elements are functional, not physical.

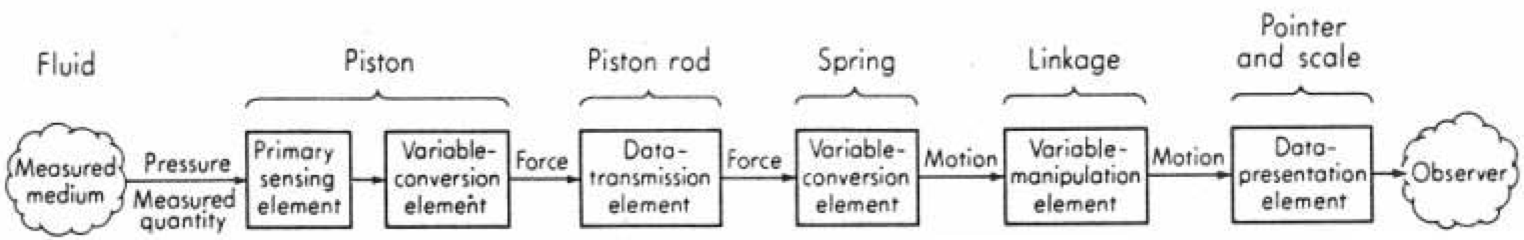
- Primary Sensing Element
  - This is the element that first receives energy from the measured medium and produces an output depending in some way on the measured quantity. The output is some physical variable, e.g., displacement or voltage. An instrument always extracts some energy from the measured medium. The measured quantity is always disturbed by the act of measurement, which makes a perfect measurement theoretically impossible.
- Variable-Conversion Element
  - It may be necessary to convert the output signal of the primary sensing element to another more suitable variable while preserving the information content of the original signal. This element performs this function.

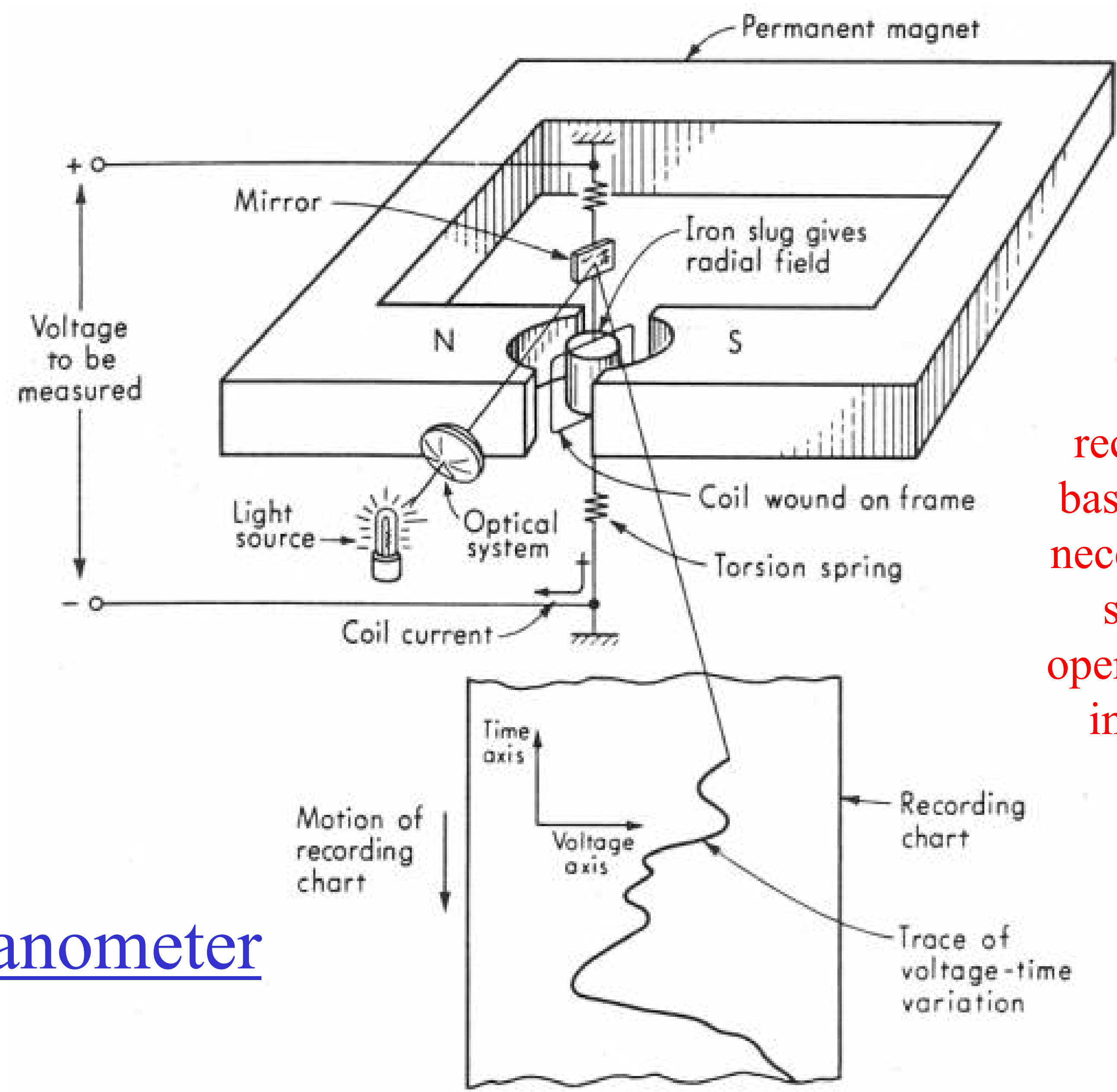
- Variable-Manipulation Element
  - An instrument may require that a signal represented by some physical variable be manipulated in some way. By manipulation we mean specifically a change in numerical value according to some definite rule but a preservation of the physical nature of the variable. This element performs such a function.
- Data-Transmission Element
  - When functional elements of an instrument are actually physically separated, it becomes necessary to transmit the data from one to another. This element performs this function.

- Data-Presentation Element
  - If the information about the measured quantity is to be communicated to a human being for monitoring, control, or analysis purposes, it must be put into a form recognizable by one of the human senses. This element performs this “translation” function.
- Data Storage/Playback Element
  - Some applications require a distinct data storage/playback which can easily recreate the stored data upon command.
- Note
  - A given instrument may involve the basic functions in any number, combination, or order. A given physical component may serve several of the basic functions.



Pressure Gage





Can you recognize the basic functions necessary to the successful operation of this instrument?

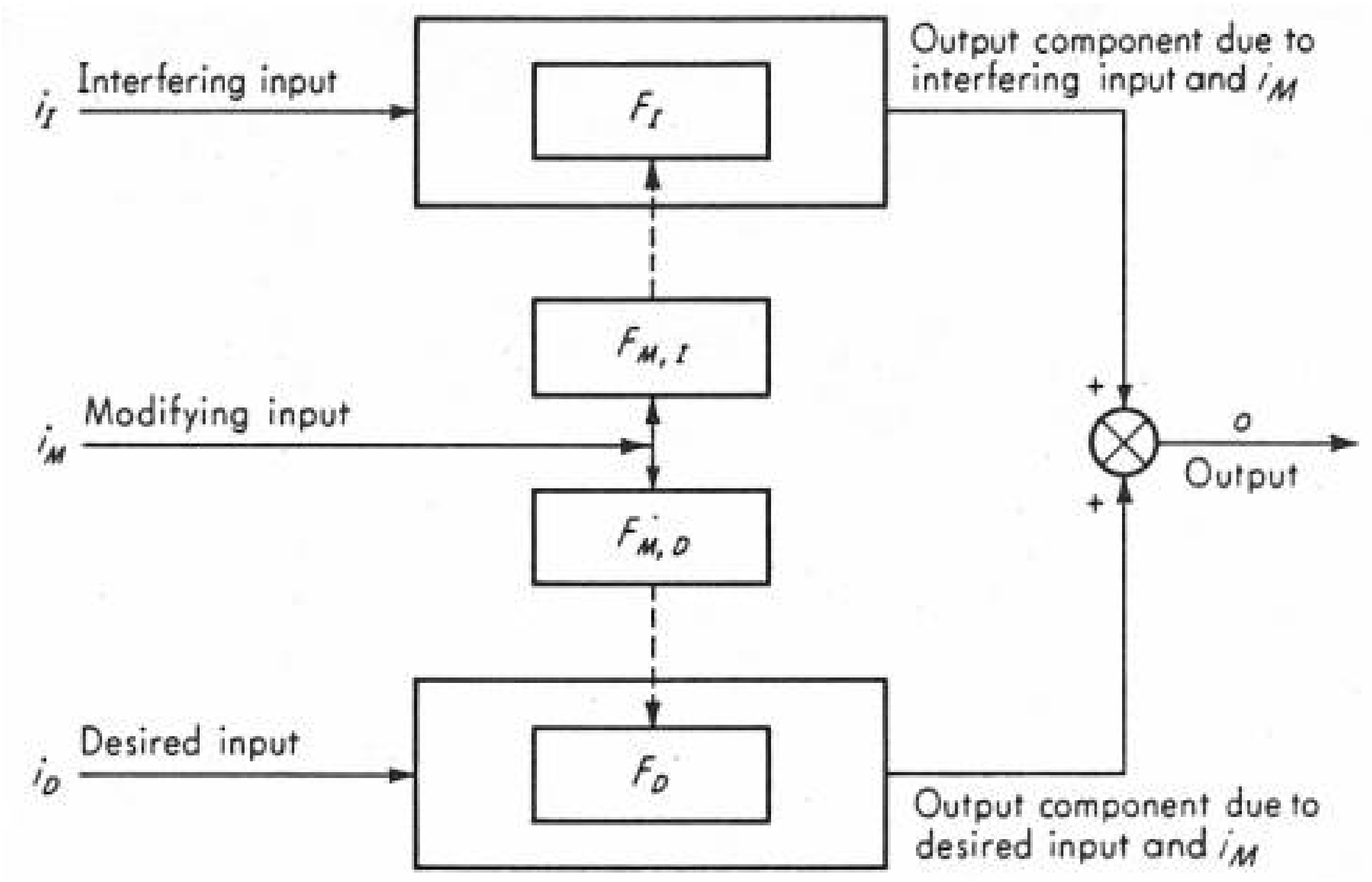
Galvanometer

- <u>Active vs. Passive Transducers</u>
  - A component whose output energy is supplied entirely or almost entirely by its input signal is commonly called a *passive transducer*. The output and input signals may involve energy of the same form or there may be an energy conversion from one form to another.
  - An *active transducer* has an auxiliary source of power which supplies a major part of the output power while the input signal supplies only an insignificant portion. Again, there may or may not be a conversion of energy from one form to another.

- Analog vs. Digital Modes of Operation
  - For analog signals, the precise value of the quantity (voltage, rotation angle, etc.) carrying the information is significant; the specific waveform of input and output signals is of vital importance.
  - Digital signals are basically of a binary (on/off) nature, and variations in numerical value are associated with changes in the logical state (true/false) of some combination of switches.
    - +2 V to +5 V represents ON state
    - 0 V to +0.8 V represents OFF state
  - In digital devices, it is simply the presence (logical 1) or absence (logical 0) of a voltage within some wide range that matters; the precise value of the signal is of no consequence.

- Digital devices are very tolerant of noise voltages and need not be individually very accurate, even though the overall system can be extremely accurate.
- When combined analog/digital systems are used (often the case in measurement systems), the digital portions need not limit system accuracy; these limitations generally are associated with analog portions and/or the analog/digital conversion devices.
- Since most measurement and control apparatus is of an analog nature, it is necessary to have both A/D converters and D/A converters, which serve as translators that enable the computer to communicate with the outside analog world.

- # Input-Output Configuration of Measuring Instruments and Instrument Systems



- Input quantities are classified into three categories:
  - Desired Inputs
    - These are quantities that the instrument is specifically intended to measure.
  - Interfering Inputs
    - These are quantities to which the instrument is unintentionally sensitive.
  - F_D and F_I are input-output relations, i.e., the mathematical operations necessary to obtain the output from the input. They represent different concepts depending on the particular input-output characteristic being described, e.g., a constant, a mathematical function, a differential equation, a statistical distribution function.

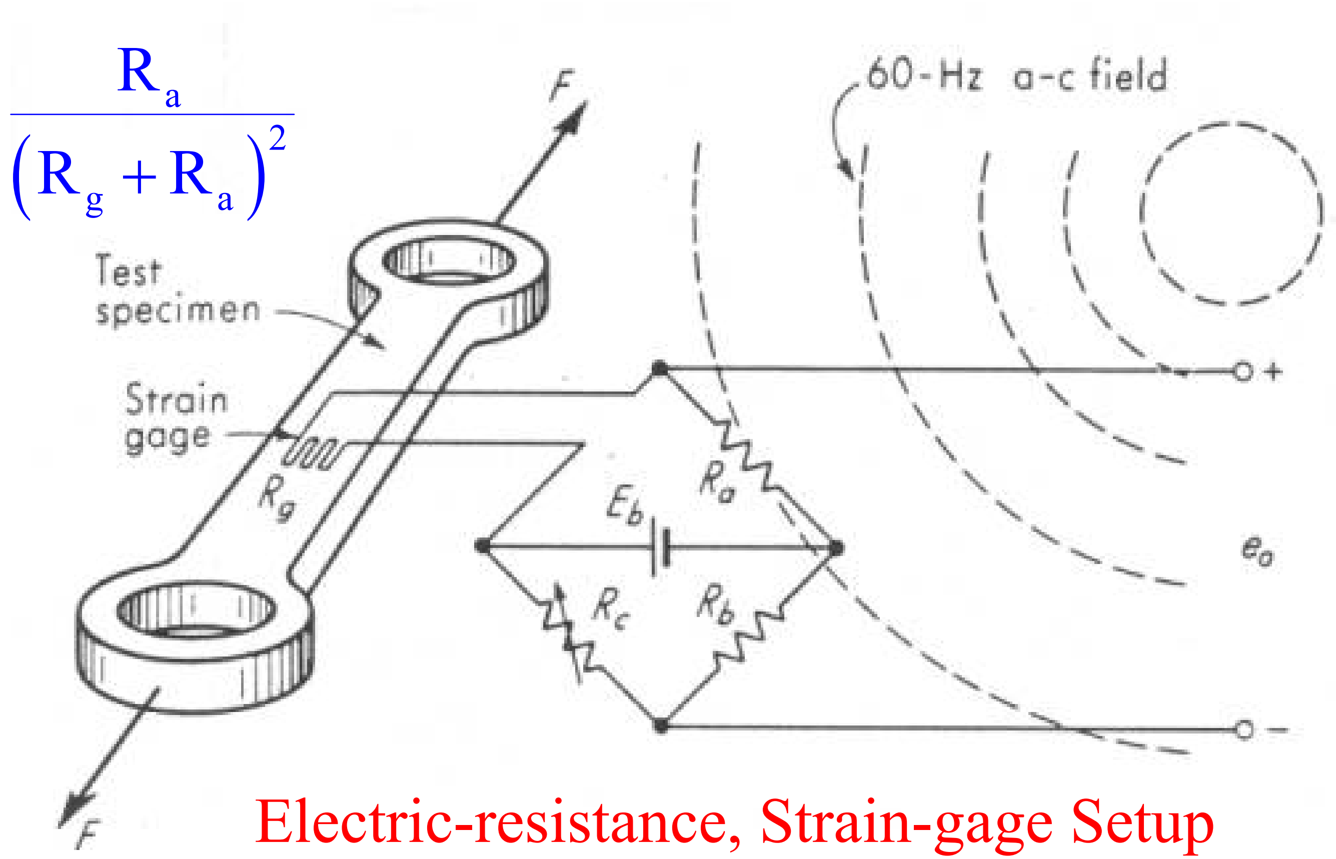
– Modifying Inputs

- These are quantities that cause a change in the input-output relations for the desired and interfering inputs, i.e., they cause a change in F_D and/or F_I. $F_{M,I}$ and $F_{M,D}$ represent the specific manner in which i_M affects F_I and F_D, respectively.

Examples of Input Quantities

$$\Delta R_g = (GF) R_g \varepsilon$$

$$e_0 = -(GF) R_g \varepsilon E_b \frac{R_a}{(R_g + R_a)^2}$$

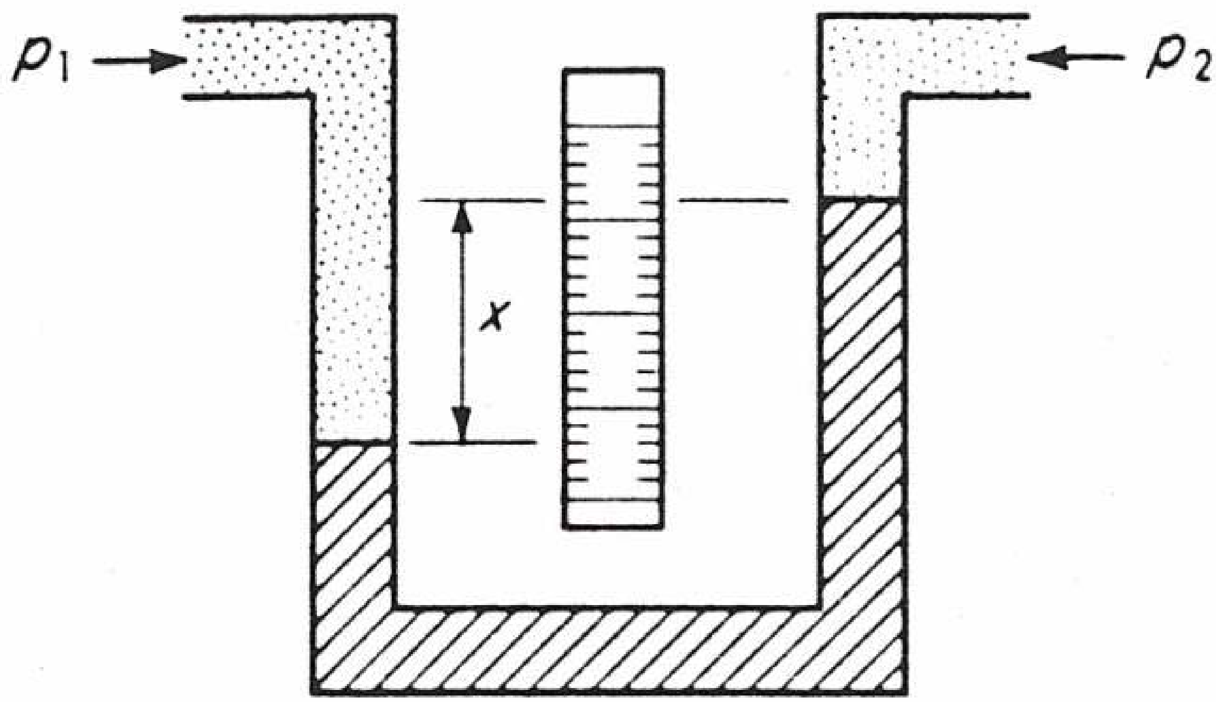


Electric-resistance, Strain-gage Setup

Desired Input: strain

Interfering Inputs: 60 Hz field, temperature

Modifying Inputs: temperature, battery voltage



U-Tube Manometer for Differential Pressure Measurement

<u>Desired Inputs</u>: p_1 and p_2

<u>Interfering Inputs</u>: acceleration, tilt angle

<u>Modifying Inputs</u>: temperature, gravitational force, tilt angle

- Methods for Nullifying or Reducing the Effects of Spurious Inputs:
  - Method of Inherent Sensitivity
  - Method of High-Gain Feedback
  - Method of Calculated Output Corrections
  - Method of Signal Filtering
  - Method of Opposing Inputs

- <u>Method of Inherent Insensitivity</u>
  - This method proposes the sound design philosophy that the elements of the instrument should inherently be sensitive to only the desired inputs. This approach requires that somehow F_I and/or $F_{M,D}$ be made as nearly equal to zero as possible. Thus, even though i_I and/or i_M may exist, they cannot affect the output.
- <u>Method of High-Gain Feedback</u>
  - Suppose we wish to measure a voltage e_i by applying it to a motor whose torque is applied to a spring, causing a displacement x_o, which may be measured on a calibrated scale. For this properly designed open-loop system:

$$x_o = \left(K_{Mo} K_{Sp}\right) e_i$$

- If modifying inputs i_{M1} and i_{M2} exist, they can cause changes in K_{Mo} and K_{Sp} that lead to errors in the relation between e_i and x_o. These errors are in direct proportion to the changes in K_{Mo} and K_{Sp}.
- Consider a closed-loop system. Here x_o is measured by a feedback device which produces a voltage e_o proportional to x_o. This voltage is subtracted from the input voltage e_i, and the difference is applied to an amplifier which drives the motor and thereby the spring to produce x_o:

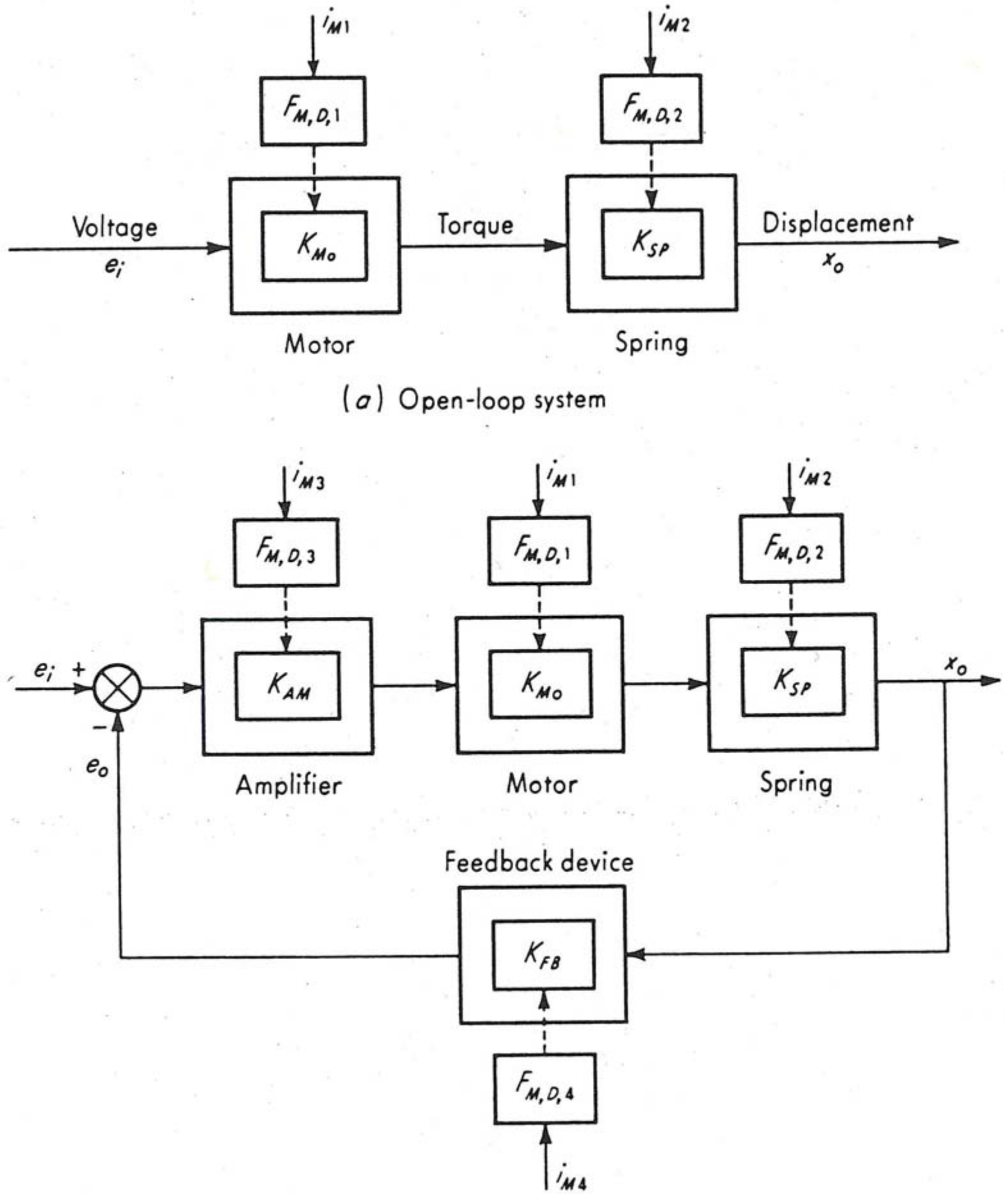
$$x_o = \frac{K_{Am} K_{Mo} K_{Sp}}{1 + K_{Am} K_{Mo} K_{Sp} K_{Fb}} e_i$$

- Suppose we design K_{Am} to be very large (a "high-gain" system) so that $K_{Am}K_{Mo}K_{Sp}K_{Fb} >> 1$. Then

$$x_o \approx \frac{1}{K_{Fb}} e_i$$

- The effects of variations in K_{Am}, K_{Sp}, and K_{Mo} (as a result of modifying inputs i_{M1}, i_{M2}, and i_{M3}) on the relation between e_i and output x_o have been made negligible. We now require that K_{Fb} stay constant (unaffected by i_{M4}) in order to maintain constant input-output calibration.

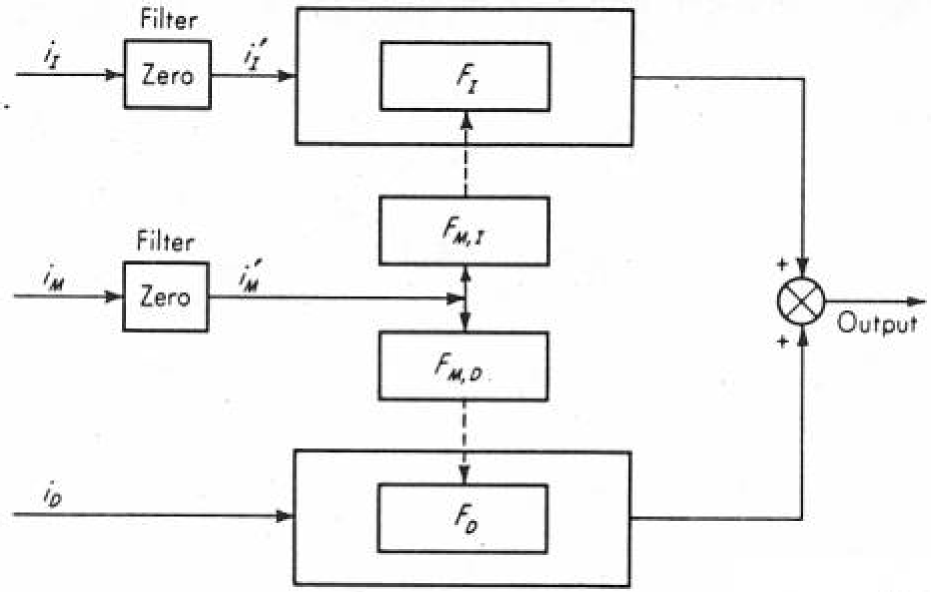
Method of High-Gain Feedback



(b) Closed-loop or feedback system

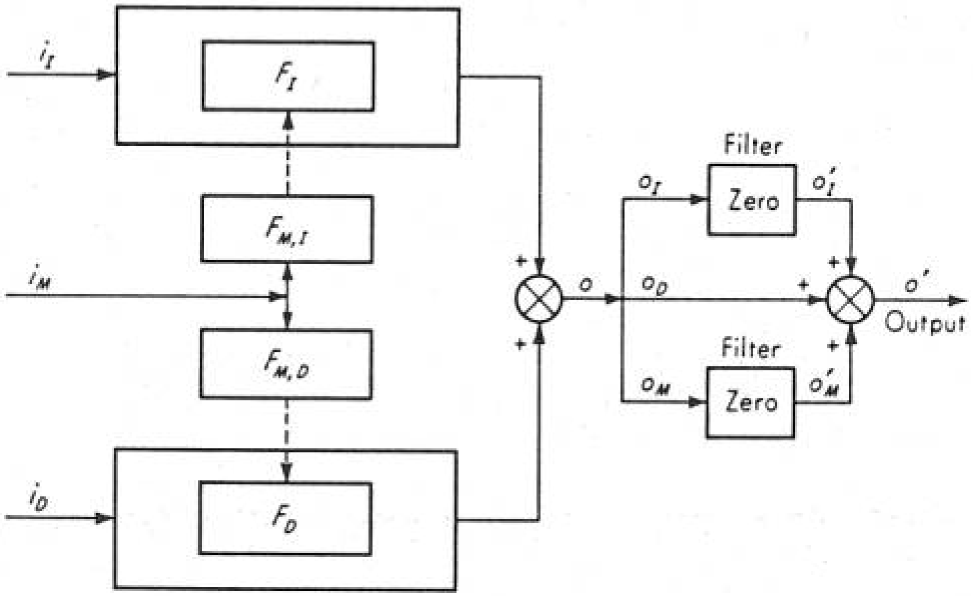
- Method of Calculated Output Corrections
  - This method requires one to measure or estimate the magnitudes of the interfering and/or modifying inputs and to know quantitatively how they affect the output. Then it is possible to calculate corrections which may be added to or subtracted from the indicated output so as to leave (ideally) only that component associated with the desired input.
  - Since many measurement systems today can afford to include a microcomputer to carry out various functions, if we also provide sensors for the spurious inputs, the microcomputer can implement the method of calculated output corrections on an automatic basis.

- Method of Signal Filtering
  - This method is based on the possibility of introducing certain elements (“filters”) into the instrument which in some fashion block the spurious signals, so that their effects on the output are removed or reduced. The filter may be applied to any suitable signal in the instrument, be it input, output, or intermediate signal.



(a) Input filtering

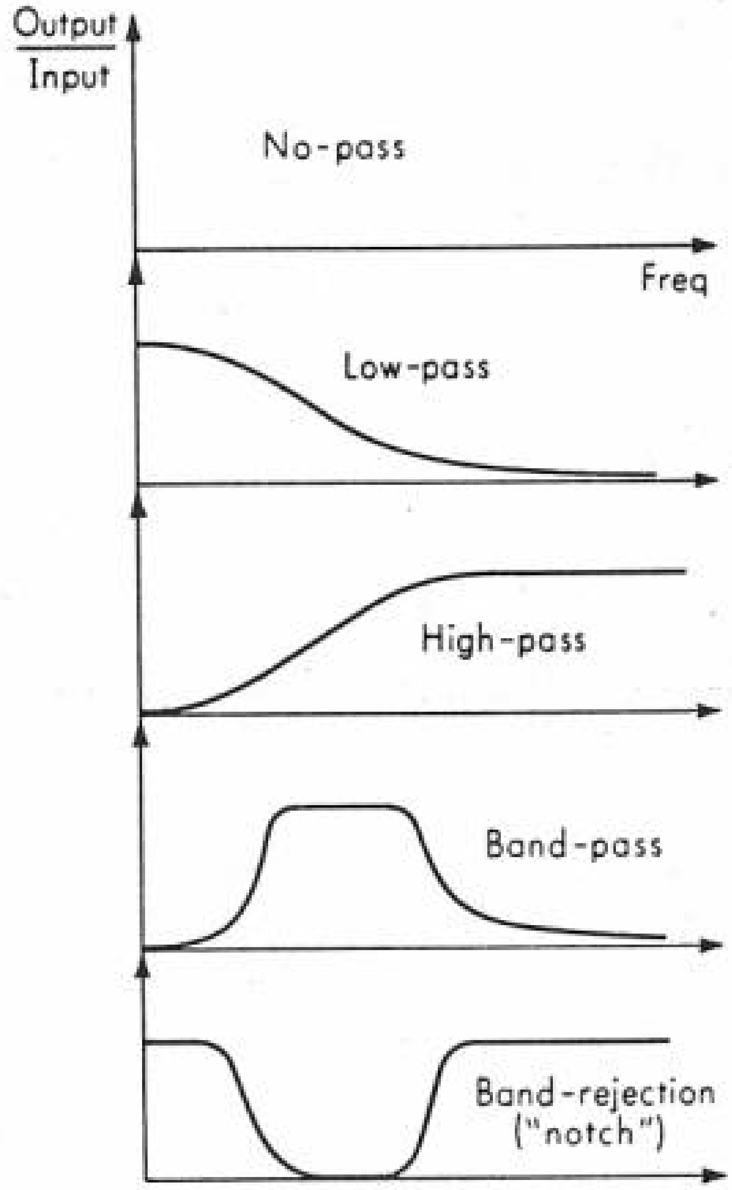
Input and Output Filtering



(b) Output filtering

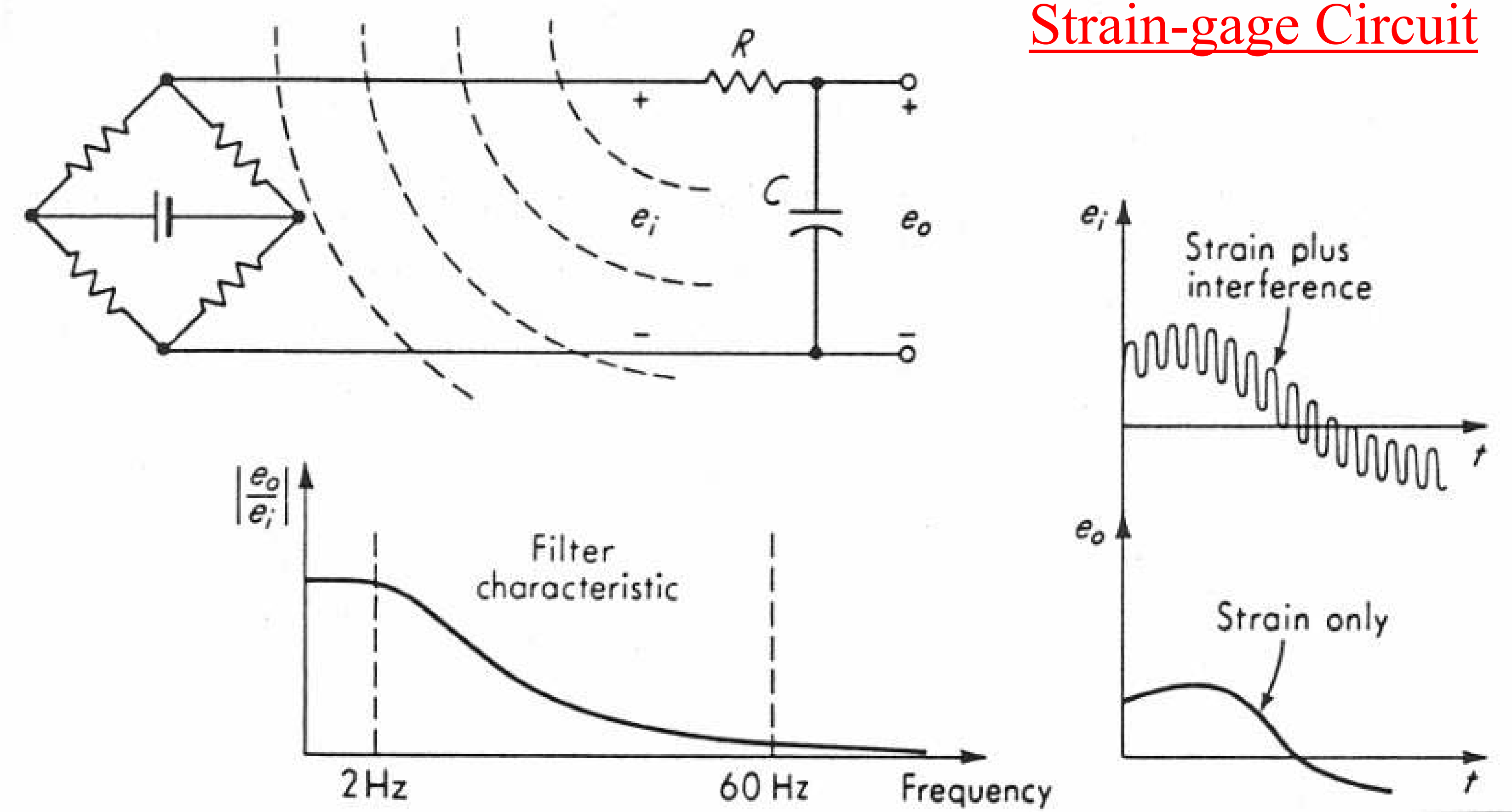
Basic Filter Types

It is usually possible to design filters of mechanical, electrical, thermal, pneumatic, etc. nature which separate signals according to their frequency content in some specific manner.

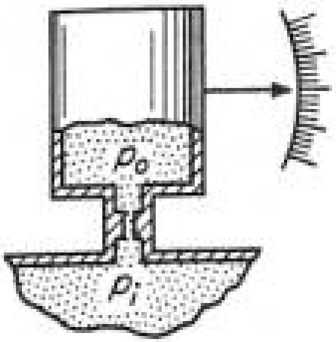


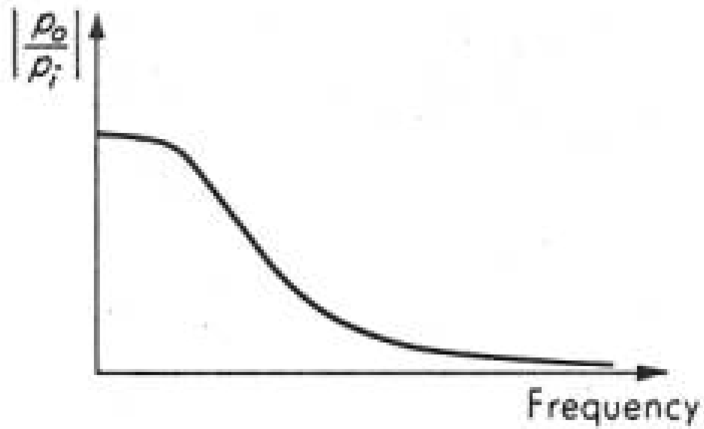
Examples of Filtering

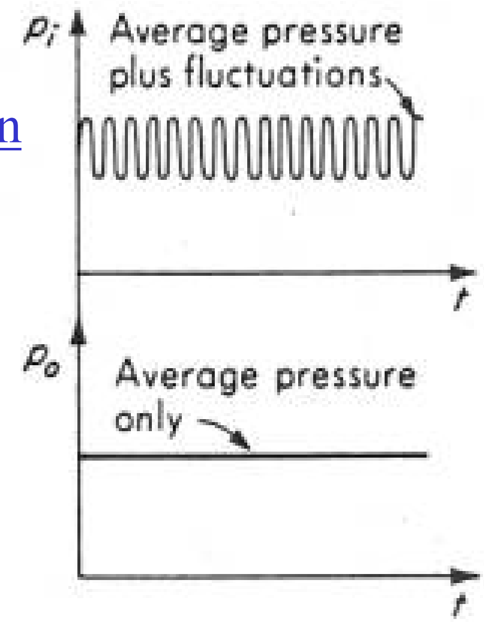
Strain-gage Circuit



Pressure Gage with Flow Restriction



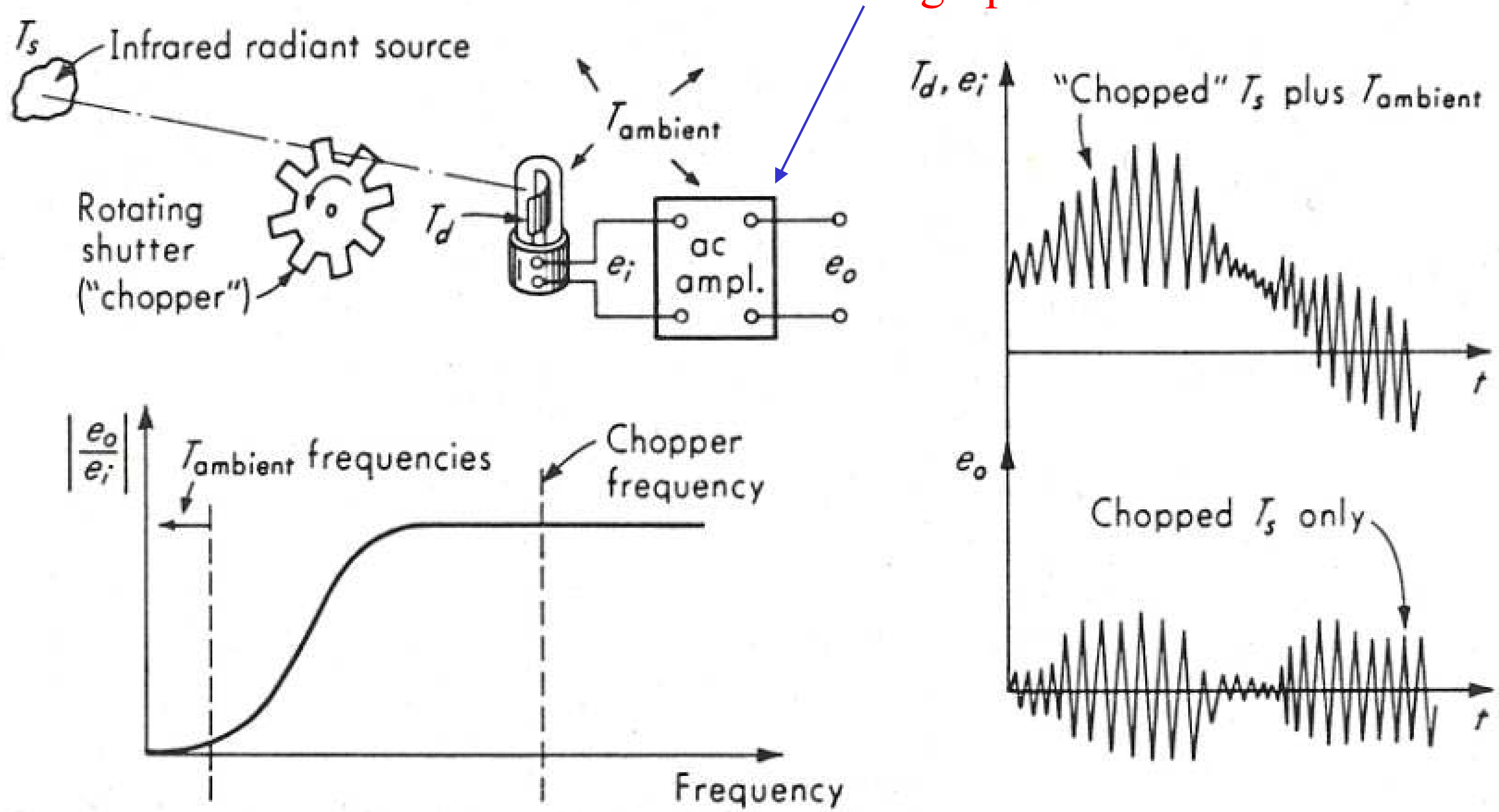




Pressure gage modified by the insertion of a flow restriction (e.g., needle valve) between the source of the pressure and the piston chamber.

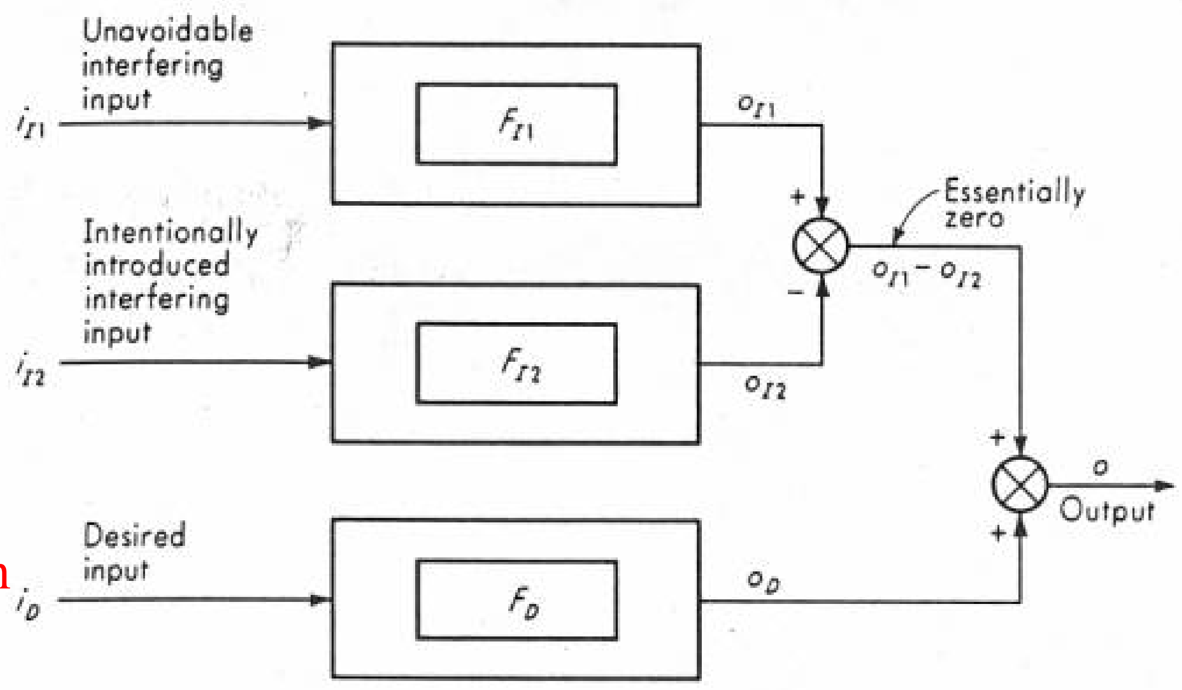
Pulsations in air pressure may be smoothed by the pneumatic filtering effect of the flow restriction and associated volume.

high-pass filter



Chopped Radiometer

- Method of Opposing Inputs
  - This method consists of intentionally introducing into the instrument interfering and/or modifying inputs that tend to cancel the bad effects of the unavoidable spurious inputs.



Method applied to interfering inputs; extension to modifying inputs is obvious.

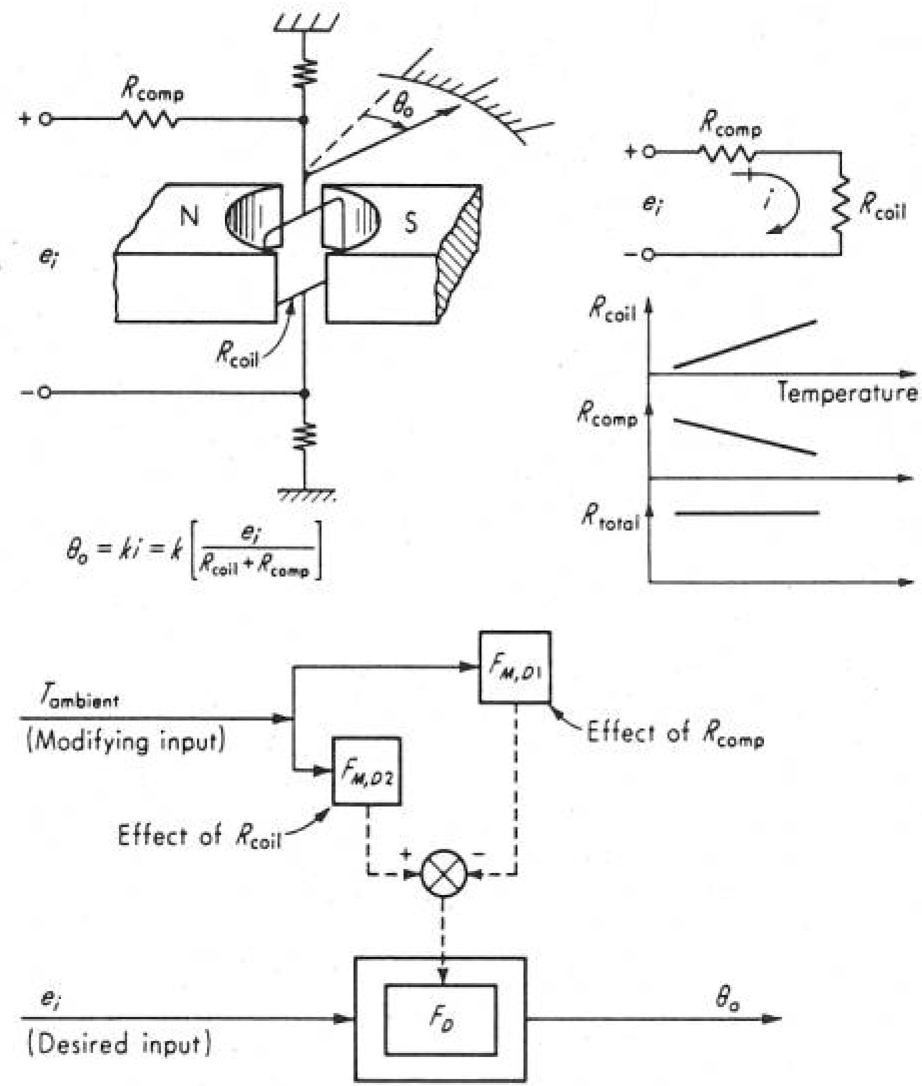
- The intentionally introduced input is designed so that the signals o_{I1} and o_{I2} are essentially equal, but act in opposite sense; thus the net contribution $o_{I1} - o_{I2}$ to the output is essentially zero.
- This method might actually be considered as a variation on the method of calculated output corrections. However, the "calculation" and application of the correction are achieved automatically owing to the structure of the system, rather than by numerical calculation by a human operator.

Examples of Method of Opposing Inputs

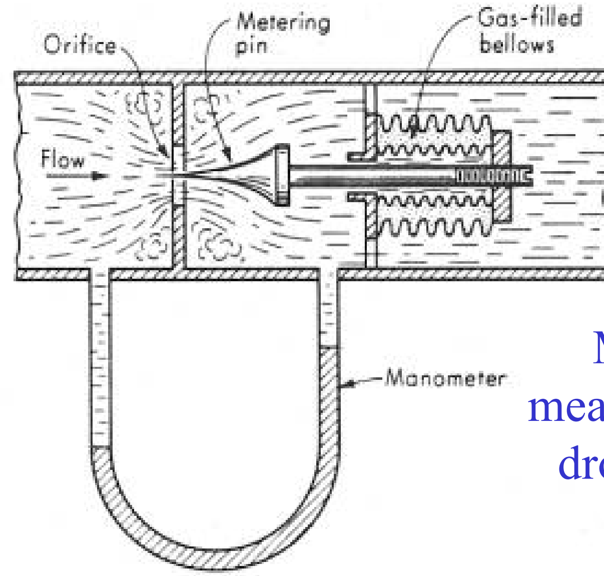
Millivoltmeter

Millivoltmeter is basically a current-sensitive device whose scale can be calibrated in voltage as long as the total circuit resistance is constant.

Ambient temperature is a modifying input here.

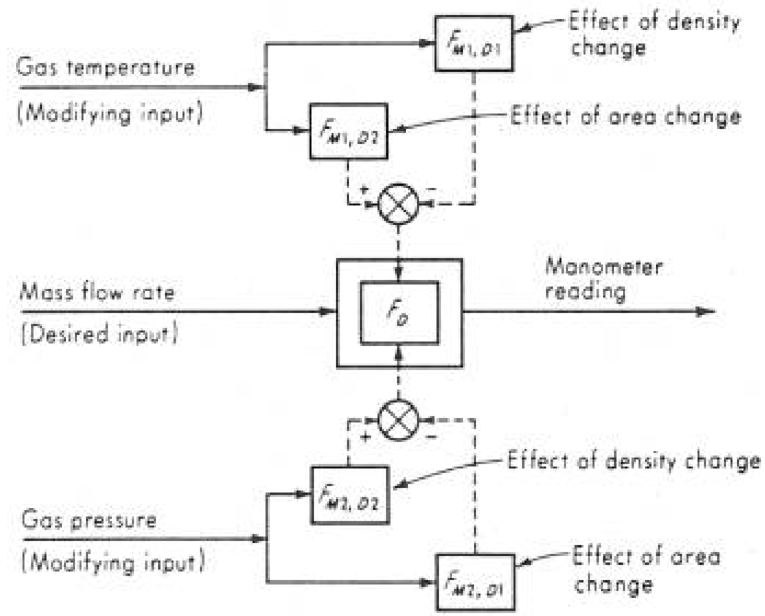


Device for The Measurement of the Mass Flow Rate of Gases



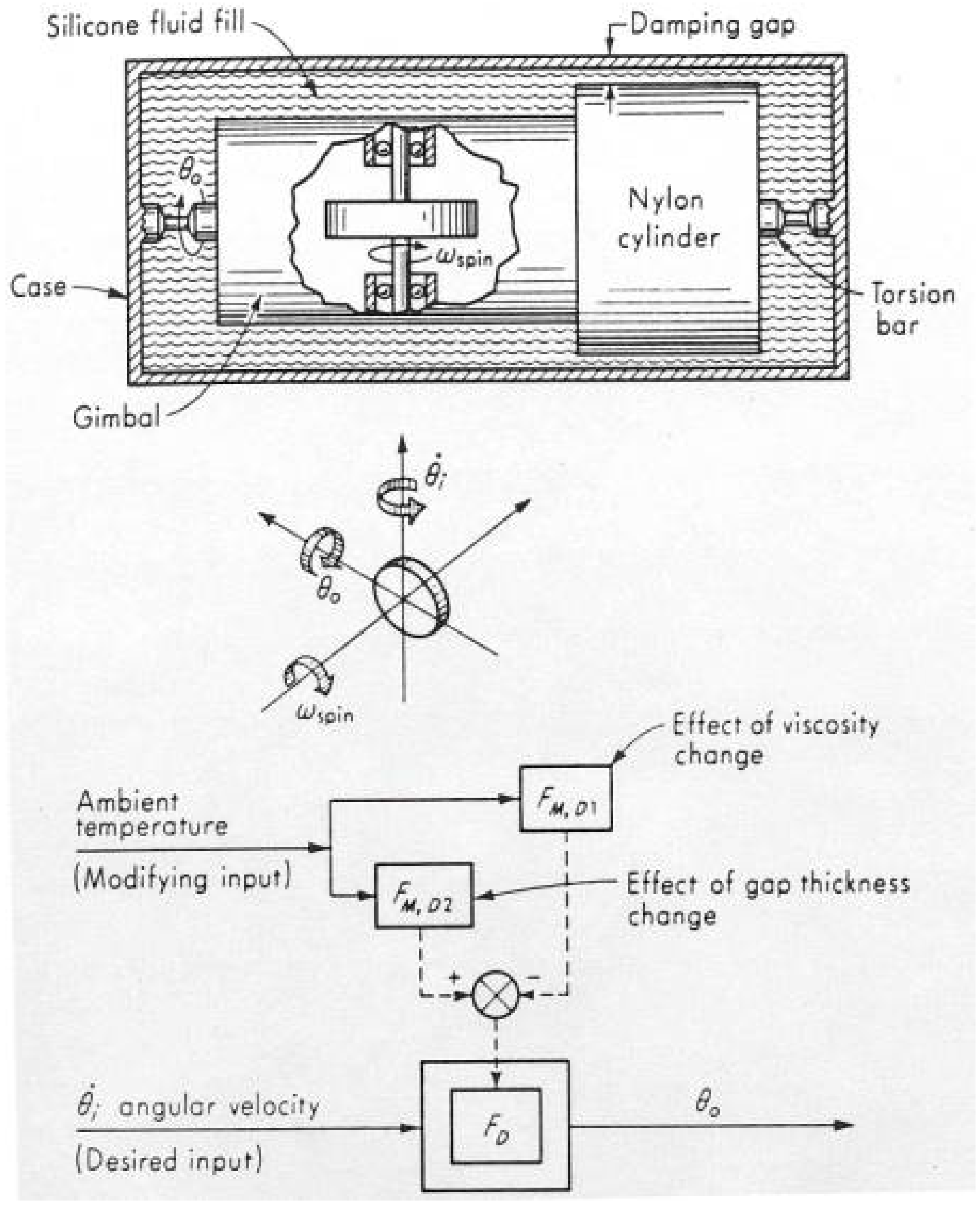
Manometer measures pressure drop across the orifice

Mass flow rate depends on the density of the gas, which varies with pressure and temperature. Variations in gas temperature and pressure yield different mass flow rates for the same orifice pressure drop. Flow rate through the orifice also depends on its flow area.



Rate Gyroscope

To control unwanted oscillations, the gimbal rotation is damped by the shearing action of a viscous fluid in a narrow damping gap. The damping effect varies with the viscosity of the fluid and the thickness of the damping gap. Ambient temperature is a modifying input.



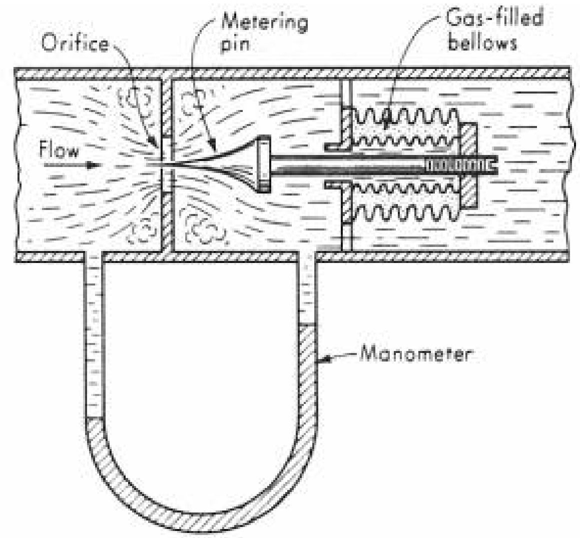
Sensors & Actuators in Mechatronics

MEAE 6960
Summer 2002

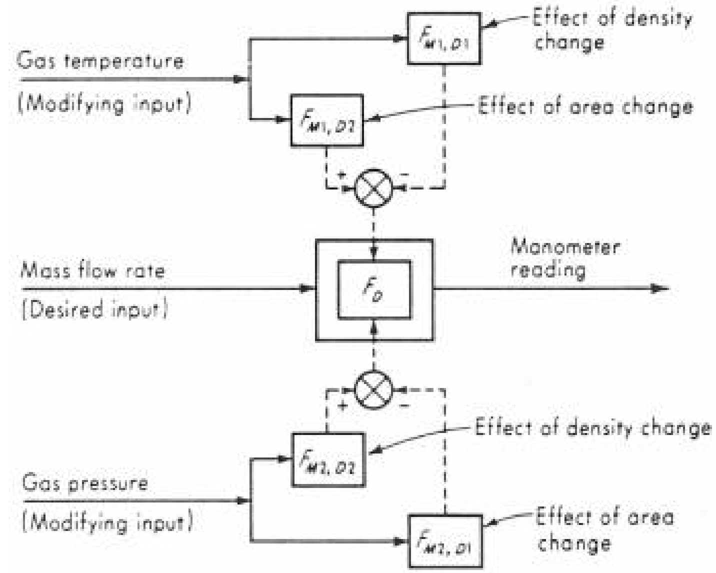
Assignment # 5

- Problem # 1
  - The natural frequency of oscillation of the balance wheel in a watch depends on the moment of inertia of the wheel and the spring constant of the torsional hairspring. A temperature rise results in a reduced spring constant, which lowers the oscillation frequency. Propose a compensating means for this effect. Non-temperature-sensitive hairspring material is not an acceptable solution.

• Problem # 2



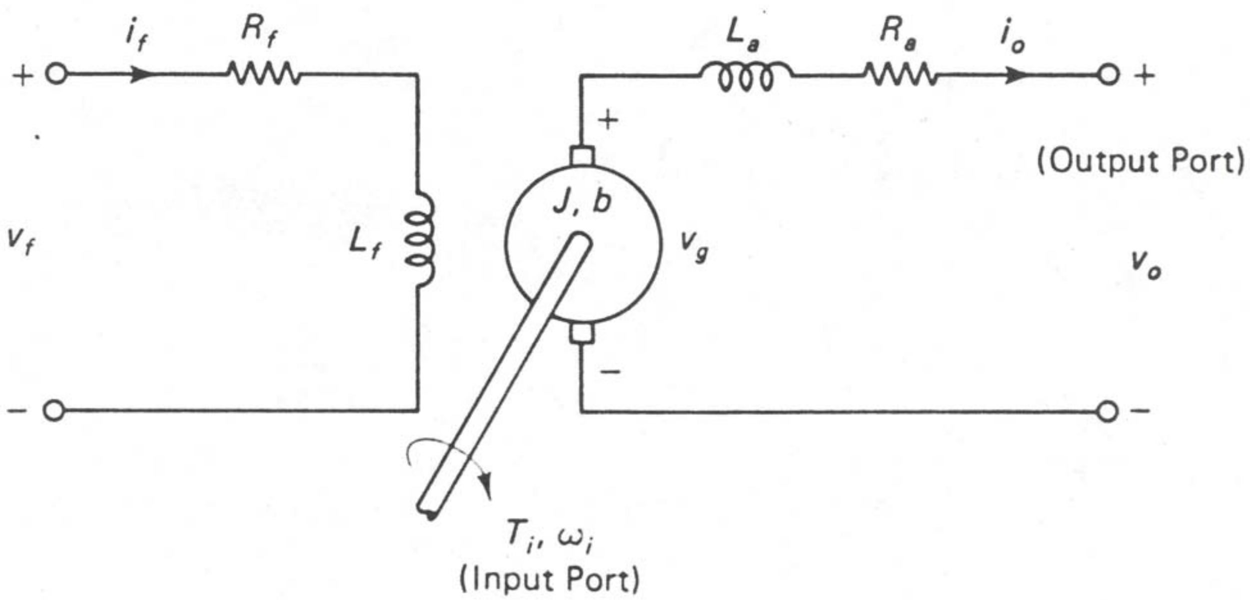
Device for the Measurement of the Mass Flow Rate of Gases



- A device for measurement of the mass flow rate of gases is shown. The mass flow rate of gas through an orifice may be found by measuring the pressure drop across the orifice, perhaps by means of a U-tube manometer. Unfortunately, the mass flow rate also depends on the density of the gas, which varies with pressure and temperature. Thus the pressure-drop measuring device cannot be calibrated to give the mass flow rate, since variations in gas temperature and pressure yield different mass flow rates for the same orifice pressure drop.
- The instrument overcomes this problem in an ingenious fashion. Explain this application of the method of opposing inputs.

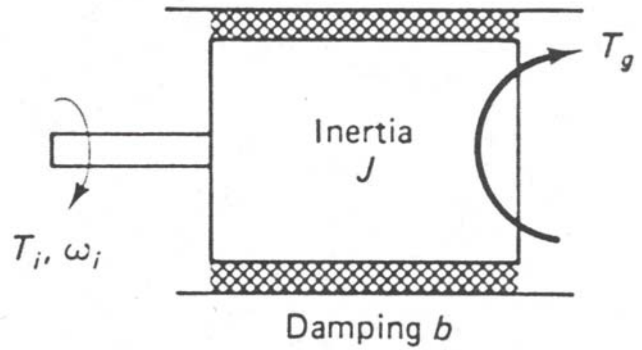
- <u>Problem # 3</u>
  - The tachometer (schematic shown) is a velocity-measuring device (passive) that uses the principle of electromagnetic generation. The field windings are powered by DC voltage v_f. The across variable at the input port is the measured angular speed ω_i. The corresponding torque T_i is the through variable at the input port. The output voltage v_o of the armature circuit is the across variable at the output port. The corresponding current i_o is the through variable at the output port. Obtain a transfer-model for this device.

$$\begin{bmatrix} v_o \\ i_o \end{bmatrix} = \begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix} \begin{bmatrix} \omega_i \\ T_i \end{bmatrix}$$



DC Tachometer

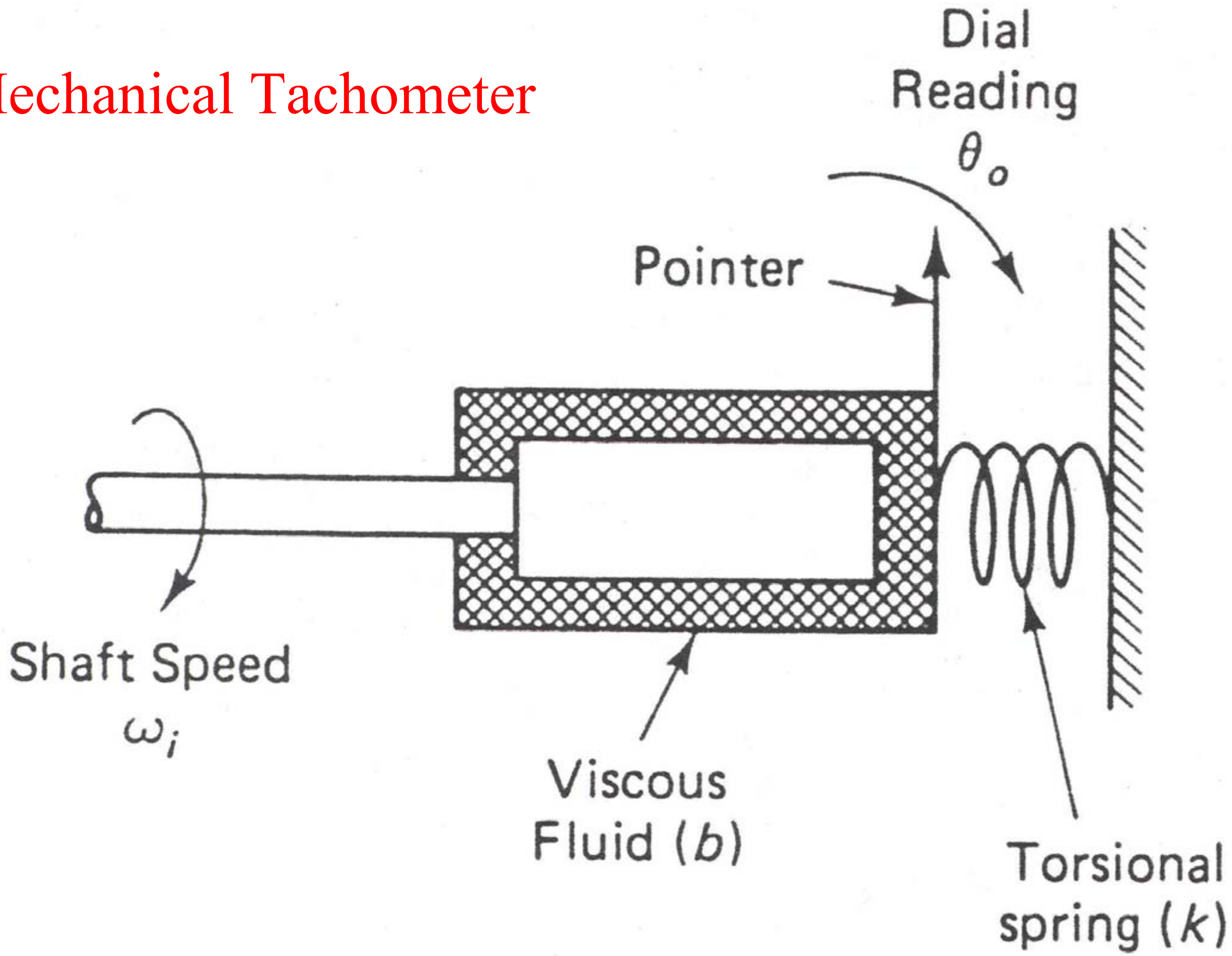
(a) Equivalent Circuit



(b) Armature FBD

- Problem # 4
  - A mechanical device for measuring angular velocity is shown. The main element of the tachometer is a rotary viscous damper (damping constant b) consisting of two cylinders. The outer cylinder carries a viscous fluid within which the inner cylinder rotates. The inner cylinder is connected to the shaft whose speed ω_i is to be measured. The outer cylinder is resisted by a linear torsion spring of stiffness k. The rotation θ_o of the outer cylinder is indicated by a pointer on a suitably calibrated scale. Neglecting the inertia of moving parts, perform a bandwidth analysis on this device. What are the conflicting design requirements?

Mechanical Tachometer



- **Problem # 5**
  - In the study of sensors, there are many terms typically used by manufacturers to describe performance that are most important to understand the meaning of.
  - Define the following:

Linearity	Resolution
Saturation	Drift
Hysteresis	Useful Frequency Range
Sensitivity	Bandwidth
Dynamic Range	Accuracy
Error	Precision
Signal-to-Noise Ratio	Input/Output Impedance

  - What are the characteristics of a perfect measuring device?

Mechatronics: Loading Effects

K. Craig

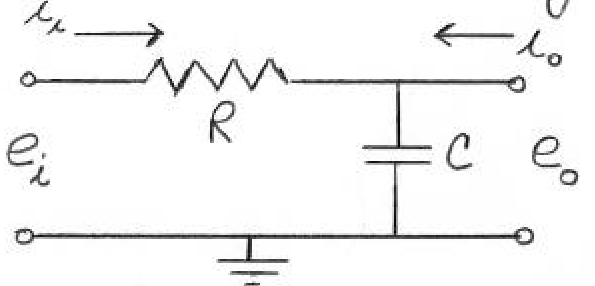
Input Impedance

Output Impedance

I. Examples of Loading Effects

A. Electrical System

- Consider the following electrical system:



Passive RC Low-Pass Filter

$$e_o = \frac{1/Cs}{R + 1/Cs} e_i \qquad \frac{e_o}{e_i} = \frac{1}{RCs+1} = G(s)$$

Here we assume that $i_o = 0$.

Ideal Unloaded Transfer Function

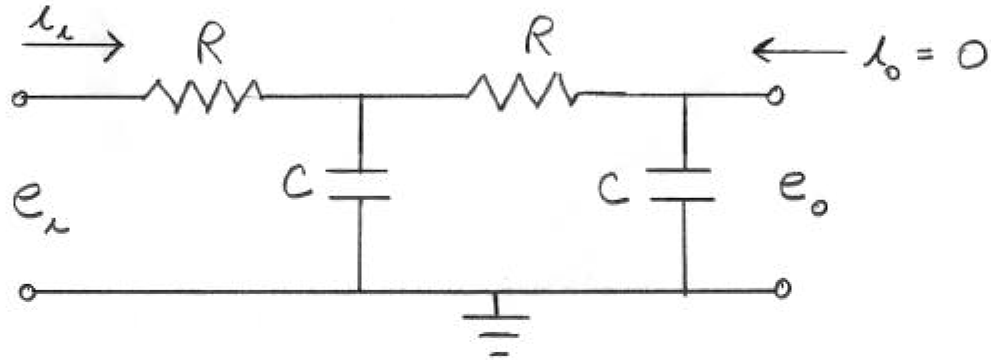
Complete Description:

$$\text{KCL} \Rightarrow i_i + i_o - C\dot{e}_o = 0 \Rightarrow i_i = -i_o + Cs\,e_o$$

$$\text{KVL} \Rightarrow e_i - Ri_i - e_o = 0 \Rightarrow e_i = e_o + Ri_i$$

$$\begin{bmatrix} e_i \\ i_i \end{bmatrix} = \begin{bmatrix} RCs+1 & -R \\ Cs & -1 \end{bmatrix} \begin{bmatrix} e_o \\ i_o \end{bmatrix}$$

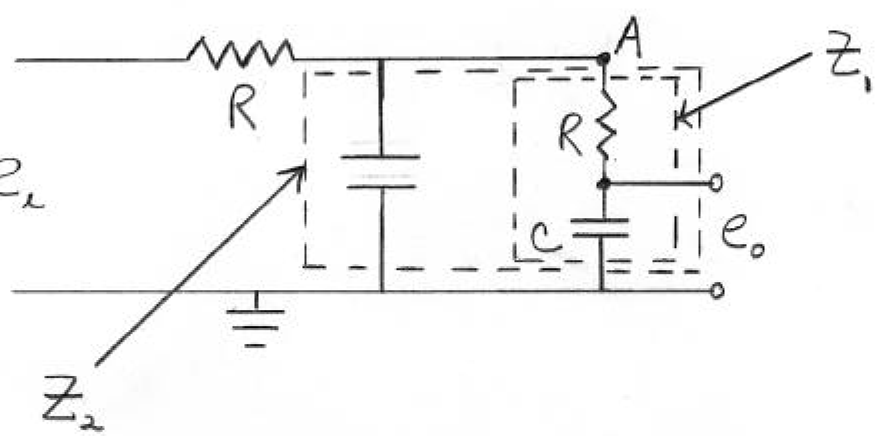
- Now connect 2 identical RC Low-Pass filters in series



Is $\dfrac{e_o}{e_i} = \left(\dfrac{1}{RCs+1}\right)^2$? **NO!**

Let's derive the overall transfer function for the RC-RC Circuit from first principles.

Redraw it to facilitate analysis.



$$\frac{e_o}{e_i} = \frac{e_o}{e_A} \cdot \frac{e_A}{e_i}$$

$$\frac{e_o}{e_A} = \frac{1/Cs}{R + 1/Cs} = \frac{1}{RCs+1}$$

<u>Z_1:</u> R and C in series

$$Z_1 = R + 1/Cs = \frac{RCs+1}{Cs}$$

<u>Z_2:</u> Z_1 and C in parallel

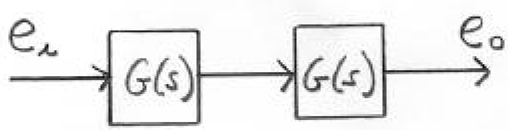
$$Z_2 = \frac{(Z_1)(1/Cs)}{Z_1 + 1/Cs} = \frac{RCs+1}{(RCs+2)(Cs)}$$

$$\frac{e_A}{e_i} = \frac{Z_2}{R + Z_2} = \frac{RCs+1}{(RCs+1)^2 + RCs}$$

$$\frac{e_o}{e_i} = \frac{e_o}{e_A} \cdot \frac{e_A}{e_i} = \frac{1}{(RCs+1)^2 + RCs}$$

$$\neq G(s)\,G(s)$$

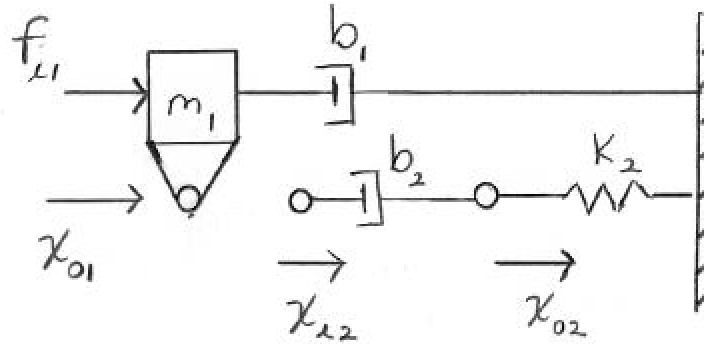
why ?



$$\frac{e_o}{e_i} \neq G(s)\,G(s)$$

B. Mechanical System

- Consider the following mechanical system:

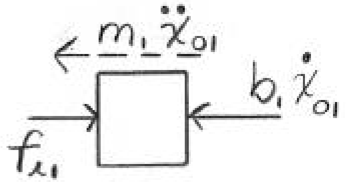


Machine-tool slide positioned by a motor

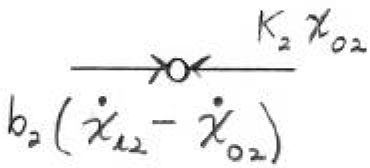
Position-measuring device to be attached to the slide

$x_{o1} = x_{i2}$ when connected

- Consider each system separately:



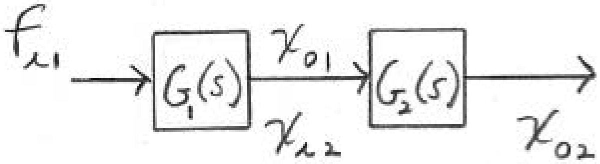
$$m_1 \ddot{x}_{o1} + b_1 \dot{x}_{o1} = f_{i1}$$



$$(\dot{x}_{i2} - \dot{x}_{o2}) b_2 = K_2 x_{o2}$$

$$\frac{x_{o1}}{f_{i1}} = \frac{1}{m_1 s^2 + b_1 s} = G_1(s) \qquad \frac{x_{o2}}{x_{o1}} = \frac{b_2 s}{b_2 s + K_2} = G_2(s)$$

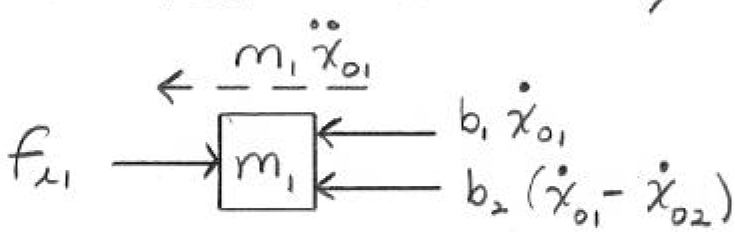
But $x_{o1} = x_{i2}$ when connected



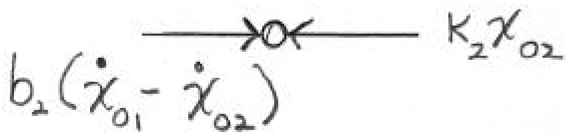
$$\frac{x_{o2}}{f_{i1}} = G_1(s)\,G_2(s) = \frac{\dfrac{b_2}{b_1}K_2}{\dfrac{m_1 b_2}{b_1 K_2}s^2 + \dfrac{m_1 K + b_1 b_2}{b_1 K_2}s + 1}$$

Is This Correct ?

- Consider the two systems connected:



$x_{o1} = x_{i2}$



$$m_1\ddot{x}_{o1} + b_1\dot{x}_{o1} + b_2(\dot{x}_{o1} - \dot{x}_{o2}) = f_{i1}$$

$$b_2(\dot{x}_{o1} - \dot{x}_{o2}) = K_2 x_{o2}$$

Transform and Solve for $\dfrac{x_{o2}}{f_{i1}}$

$$\left[m_1 s^2 + (b_1 + b_2)s\right] x_{o1} - (b_2 s)\, x_{o2} = f_{i1}$$

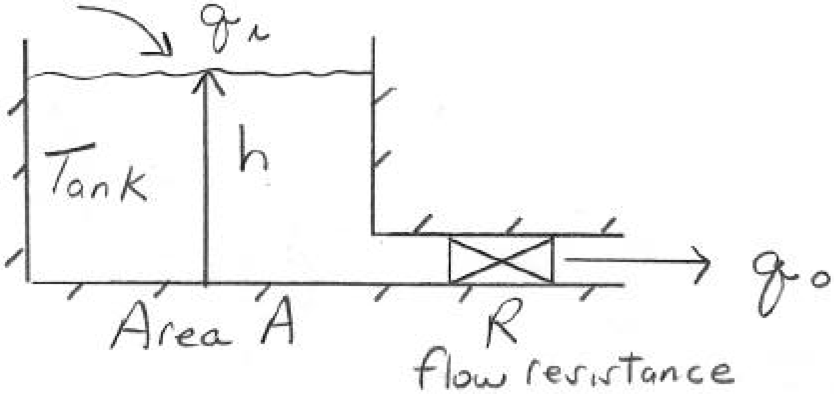
$$b_2 s\, x_{o1} = (b_2 s + k_2)\, x_{o2}$$

Solve for $\dfrac{x_{o2}}{f_{i1}}$

$$\frac{x_{o2}}{f_{i1}} = \frac{\dfrac{b_2}{(b_1 + b_2)k_2}}{\dfrac{m_1 b_2}{(b_1 + b_2)k_2} s^2 + \dfrac{m_1 k_2 + b_1 b_2}{(b_1 + b_2)k_2} s + 1} = G(s)$$

$$G(s) \neq G_1(s)\, G_2(s)$$

C. Hydraulic System

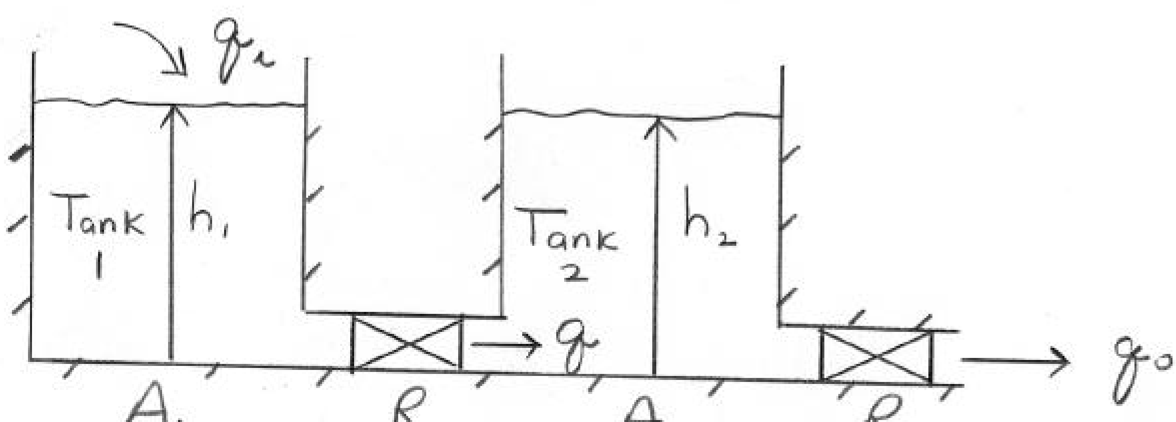


$$q_i - q_o = A\dot{h}$$
$$h = R q_o$$

Transform and Solve for $\frac{q_o}{q_i}$

$$\left.\begin{array}{l} q_i - q_o = Ash \\ h = R q_o \end{array}\right\} \quad \frac{q_o}{q_i} = \frac{1}{RAs+1}$$

- Now consider two tanks interconnected:



Is $\frac{q_o}{q_i} = \frac{1}{(R_1A_1s+1)(R_2A_2s+1)}$?

Write down equations of motion:

$$q_i - q = A_1 \dot{h}_1 \qquad h_1 - h_2 = R_1 q$$

$$q - q_o = A_2 \dot{h}_2 \qquad h_2 = R_2 q_o$$

Transform and Solve for $\dfrac{q_o}{q_i}$

Result:

$$\frac{q_o}{q_i} = \frac{1}{(R_1 A_2 R_2 A_2) s^2 + (R_1 A_1 + R_2 A_2 + R_2 A_1) s + 1}$$

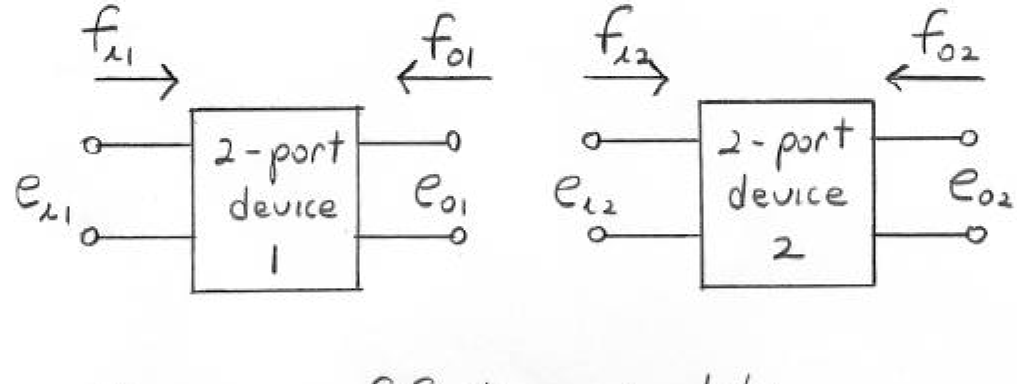
$$= G(s)$$

$$G_1(s) = \frac{1}{R_1 A_1 s + 1} \qquad G_2(s) = \frac{1}{R_2 A_2 s + 1}$$

$$G(s) \neq G_1(s) G_2(s)$$

II. Impedance

- When a second device is coupled to a first device at its output, it will draw some power from the first. Definition of this power is impossible in terms of a single variable. Two variables are required.
- A two-port device is one that exchanges energy with others at only two locations (ports).

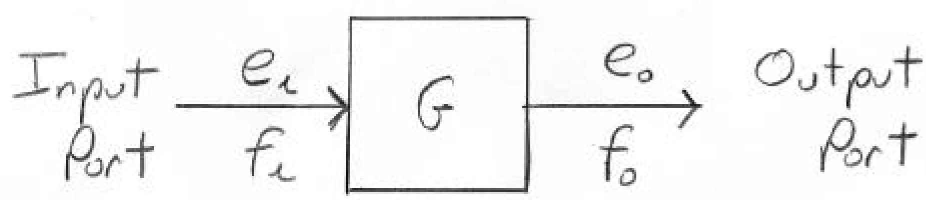


$e \equiv$ effort variable

$f \equiv$ flow variable

Class of System	Effort Variable (e)	Flow Variable (f)
Electrical	Voltage	Current
Mechanical	Force	Velocity
Hydraulic	Pressure	Flow Rate
Thermal	Temperature	Heat-Flow Rate

- The product of the two variables at each port (effort × flow) gives the instantaneous power flowing through the port.



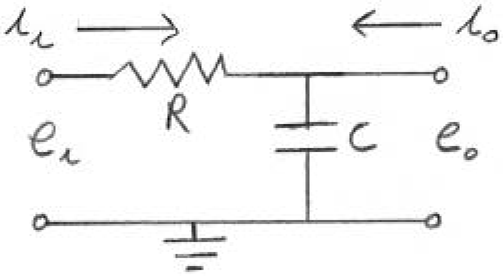
$$\begin{bmatrix} e_i \\ f_i \end{bmatrix} = \begin{bmatrix} G \end{bmatrix} \begin{bmatrix} e_o \\ f_o \end{bmatrix}$$

G = 2×2 Transfer Function

<u>Assumption:</u> Linear Model for device

- Let's Derive this Transfer Function Matrix for the 3 Systems discussed.

- RC Circuit

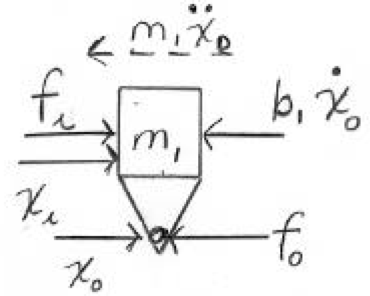


$$\text{KCL} \Rightarrow i_i + i_o - C\dot{e}_o = 0 \Rightarrow i_i = -i_o + Cse_o$$

$$\text{KVL} \Rightarrow e_i - Ri_i - e_o = 0 \Rightarrow e_i = e_o + Ri_i$$

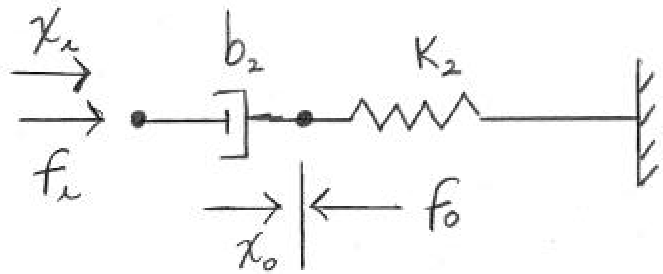
$$\begin{bmatrix} e_i \\ i_i \end{bmatrix} = \begin{bmatrix} RCs+1 & -R \\ Cs & -1 \end{bmatrix} \begin{bmatrix} e_o \\ i_o \end{bmatrix}$$

- Mechanical System



$$\begin{bmatrix} f_i \\ x_i \end{bmatrix} = \begin{bmatrix} 1 & m_1s^2 + b_1s \\ 0 & 1 \end{bmatrix} \begin{bmatrix} f_o \\ x_o \end{bmatrix}$$

$$f_i - f_o - b_1\dot{x}_o - m_1\ddot{x}_o = 0 \qquad x_i = x_o$$



$$f_i - f_o - K_2 x_o = 0 \Rightarrow f_i = f_o + K_2 x_o$$

$$b_2(\dot{x}_i - \dot{x}_o) - f_o - K_2 x_o = 0 \Rightarrow$$

$$b_2 s x_i = b_2 s x_o - K_2 x_o + f_o$$

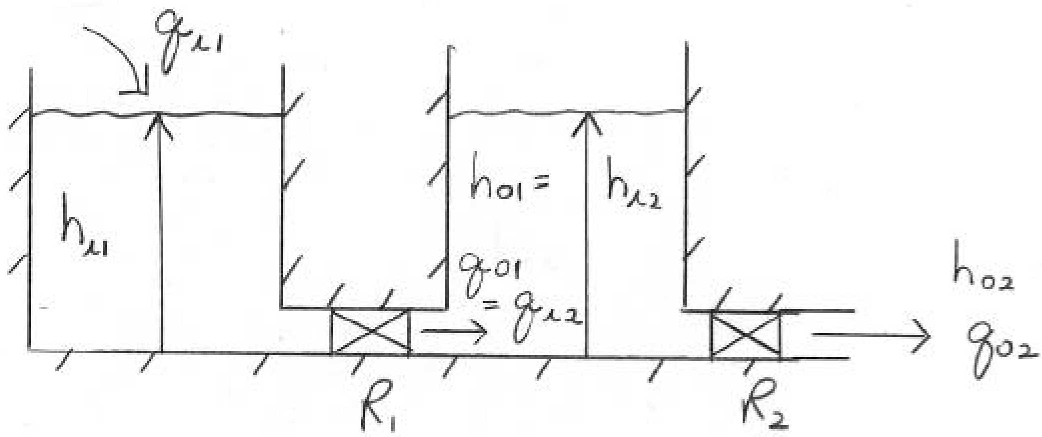
$$\begin{bmatrix} f_i \\ x_i \end{bmatrix} = \begin{bmatrix} 1 & K_2 \\ \frac{1}{b_2 s} & \frac{b_2 s + K_2}{b_2 s} \end{bmatrix} \begin{bmatrix} f_o \\ x_o \end{bmatrix}$$

The overall transfer function matrix is:

$$\begin{bmatrix} f_{i1} \\ x_{i1} \end{bmatrix} = \begin{bmatrix} 1 & m_1 s^2 + b_1 s \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & K \\ \frac{1}{b_2 s} & \frac{b_2 s + K_2}{b_2 s} \end{bmatrix} \begin{bmatrix} f_{o2} \\ x_{o2} \end{bmatrix}$$

$$\begin{bmatrix} f_{i1} \\ x_{i1} \end{bmatrix} = \begin{bmatrix} \frac{m_1 s^2 + b_1 s}{b_2 s} + 1 & K + \frac{(m_1 s^2 + b_1 s)(b_2 s + K_2)}{b_2 s} \\ \frac{1}{b_2 s} & \frac{b_2 s + K_2}{b_2 s} \end{bmatrix} \begin{bmatrix} f_{o2} \\ x_{o2} \end{bmatrix}$$

• Hydraulic System



$$\left.\begin{aligned} h_{i1} - h_{o1} &= R_1 q_{o1} \\ q_{i1} - q_{o1} &= A_1 \dot{h}_{i1} \end{aligned}\right\} \quad \begin{aligned} h_{i1} &= R q_{o1} + h_{o1} \\ q_{i1} &= A_1 (R_1 \dot{q}_{o1} + \dot{h}_{o1}) \\ &\quad + q_{o1} \end{aligned}$$

$$\begin{bmatrix} h_{i1} \\ q_{i1} \end{bmatrix} = \begin{bmatrix} 1 & R_1 \\ A_1 s & A_1 R_1 s + 1 \end{bmatrix} \begin{bmatrix} h_{o1} \\ q_{o1} \end{bmatrix}$$

Similarly

$$\begin{bmatrix} h_{i2} \\ q_{i2} \end{bmatrix} = \begin{bmatrix} 1 & R_2 \\ A_2 s & A_2 R_2 s + 1 \end{bmatrix} \begin{bmatrix} h_{o2} \\ q_{o2} \end{bmatrix}$$

The overall transfer function matrix is:

$$\begin{bmatrix} h_{i1} \\ q_{i1} \end{bmatrix} = \begin{bmatrix} 1 & R_1 \\ A_1 s & A_1 R_1 s + 1 \end{bmatrix} \begin{bmatrix} 1 & R_2 \\ A_2 s & A_2 R_2 s + 1 \end{bmatrix} \begin{bmatrix} h_{o2} \\ q_{o2} \end{bmatrix}$$

$$\begin{bmatrix} h_{i1} \\ q_{i1} \end{bmatrix} = \begin{bmatrix} 1 + R_1 A_2 s & R_2 + R_1 + A_2 R_1 R_2 s \\ (A_1 + A_2) s + A_1 A_2 R_1 s^2 & A_1 R_2 s + (A_1 R_1 s + 1)(A_2 R_2 s + 1) \end{bmatrix} \begin{bmatrix} h_{o2} \\ q_{o2} \end{bmatrix}$$

- In each case we see that we can write:

$$\begin{bmatrix} e_i \\ f_i \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} e_o \\ f_o \end{bmatrix}$$

$$e_i = a_{11} e_o + a_{12} f_o$$

$$f_i = a_{21} e_o + a_{22} f_o$$

- For linear systems, with no internal energy sources, the a_{ii}'s are related by:

$$a_{11} a_{22} - a_{21} a_{12} = 1$$

Therefore it takes only 3 of these quantities to completely describe the terminal behavior of any 2-port device, no matter how complex it might be internally.

- The equations

$$e_i = a_{11} e_o + a_{12} f_o$$
$$f_i = a_{21} e_o + a_{22} f_o$$

are actually a particular pair of a set of 12 possible equations one could write relating the e's and f's of a given 2-port device.

- Since any 2 of the variables may be considered as independent variables, we can write:

$$e_i = e_i(e_o, f_o) \qquad e_o = e_o(e_i, f_i)$$
$$f_i = f_i(e_o, f_o) \qquad f_o = f_o(e_i, f_i)$$

$$f_o = f_o(e_i, e_o) \qquad e_o = e_o(e_i, f_o)$$
$$f_i = f_i(e_i, e_o) \qquad f_i = f_i(e_i, f_o)$$

$$f_o = f_o(e_o, f_i) \qquad e_o = e_o(f_i, f_o)$$
$$e_i = e_i(e_o, f_i) \qquad e_i = e_i(f_i, f_o)$$

All of these are potentially useful.

- For device #1 we can write:

$$e_{o1} = e_{o1}(e_{i1}, f_{o1})$$

$$e_{o1}(s) = \left.\frac{e_{o1}(s)}{e_{i1}(s)}\right|_{f_{o1}(s)=0} e_{i1}(s) + \left.\frac{e_{o1}(s)}{f_{o1}(s)}\right|_{e_{i1}(s)=0} f_{o1}(s)$$

$$\left.\frac{e_{o1}(s)}{e_{i1}(s)}\right|_{f_{o1}(s)=0} \equiv \text{Unloaded Transfer Function } W_u(s)$$

$$\left.\frac{e_{o1}(s)}{f_{o1}(s)}\right|_{e_{i1}(s)=0} \equiv \text{Generalized Output Impedance } Z_{g\,o1}(s)$$

- For device #2 we can write:

$$e_{i2} = e_{i2}(f_{i2}, f_{o2})$$

$$e_{i2}(s) = \left.\frac{e_{i2}(s)}{f_{i2}(s)}\right|_{f_{o2}(s)=0} f_{i2}(s) + \left.\frac{e_{i2}(s)}{f_{o2}(s)}\right|_{f_{i2}(s)=0} f_{o2}(s)$$

$$\left.\frac{e_{i2}(s)}{f_{i2}(s)}\right|_{f_{o2}(s)=0} \equiv \text{Generalized Input Impedance } Z_{gi2}(s)$$

$$\left.\frac{e_{i2}(s)}{f_{o2}(s)}\right|_{f_{i2}(s)=0} \equiv \text{Generalized Transfer Impedance } Z_{gt2}(s)$$

- Suppose device #2 has no third device connected at its output and $f_{o2}(s)=0$.
When the 2 devices are connected $e_{o1}(s) = e_{i2}(s)$ and $f_{o1}(s) = -f_{i2}(s)$.

- Combine Equations:

$$e_{o1}(s) = W_u(s)\, e_{i1}(s) + Z_{go1}(s)\, f_{o1}(s)$$

$$= W_u(s)\, e_{i1}(s) + Z_{go1}(s) \left[\frac{-e_{i2}(s)}{Z_{gi2}(s)} \right]$$

$$= W_u(s)\, e_{i1}(s) + Z_{go1}(s) \left[\frac{-e_{o1}(s)}{Z_{gi2}(s)} \right]$$

$$e_{o1}(s) + \frac{Z_{go1}(s)}{Z_{gi2}(s)}\, e_{o1}(s) = W_u(s)\, e_{i1}(s)$$

$$\frac{e_{o1}(s)}{e_{i1}(s)} = W_u(s) \left[\frac{1}{1 + \dfrac{Z_{go1}(s)}{Z_{gi2}(s)}} \right]$$

This equation clearly shows under what conditions we may couple subsystems accurately using the familiar transfer function (here called the unloaded transfer function) method, and what additional subsystem information (the two impedances) is needed to get accurate coupling when loading is not negligible.

If $Z_{gi2} \gg Z_{go2}$, then the unloaded transfer function is a good approximation to the unloaded transfer function. This approximation may be good over certain ranges of frequency but not others.

- Let's apply this to the 3 examples.

RC-RC Circuit

$$\begin{bmatrix} e_{i1} \\ i_{i1} \end{bmatrix} = \begin{bmatrix} RCs+1 & -R \\ Cs & -1 \end{bmatrix} \begin{bmatrix} e_{o1} \\ i_{o1} \end{bmatrix} \qquad \begin{matrix} e_{o1} = e_{i2} \\ i_{o1} = -i_{i2} \end{matrix}$$

$$\begin{bmatrix} e_{i2} \\ i_{i2} \end{bmatrix} = \begin{bmatrix} RCs+1 & -R \\ Cs & 1 \end{bmatrix} \begin{bmatrix} e_{o2} \\ i_{o2} \end{bmatrix} \qquad i_{o2} = 0$$

$$\begin{bmatrix} e_{i1} \\ i_{i1} \end{bmatrix} = \begin{bmatrix} R^2C^2s^2 + 3RCs + 1 & -R^2Cs - 2R \\ RC^2s^2 + 2Cs & -RCs - 1 \end{bmatrix} \begin{bmatrix} e_{o2} \\ i_{o2} \end{bmatrix}$$

With $i_{o2} = 0$

$$\frac{e_{o2}}{e_{i1}} = \frac{1}{(RCs+1)^2 + RCs}$$

Using Impedances:

$$\frac{e_{o1}}{e_{i1}} = W_u(s) \left[\frac{1}{1 + \frac{Z_{go1}(s)}{Z_{gi2}(s)}} \right]$$

$$W_u(s) = \left. \frac{e_{o1}(s)}{e_{i1}(s)} \right|_{f_{o1}(s)=0} = \frac{1}{RCs + 1}$$

$$Z_{go1}(s) = \left. \frac{e_{o1}(s)}{f_{o1}(s)} \right|_{e_{i1}(s)=0} = \frac{R}{RCs + 1}$$

$$Z_{gi2}(s) = \left. \frac{e_{i2}(s)}{f_{i2}(s)} \right|_{f_{o2}(s)=0} = \frac{RCs+1}{Cs}$$

Therefore

$$\frac{e_{o1}}{e_{i1}} = \frac{1}{RCs+1}\left[\frac{1}{1+\frac{RCs}{(RCs+1)^2}}\right]$$

This is the loaded transfer function for the first RC Circuit. To get the overall transfer function we multiply this by the ideal transfer function of the second RC Circuit since it is unloaded.

$$\frac{e_{o2}}{e_{i1}} = \frac{1}{RCs+1}\left[\frac{1}{1+\frac{RCs}{(RCs+1)^2}}\right]\frac{1}{RCs+1}$$

$$= \frac{1}{(RCs+1)^2+RCs}$$

Same as found by direct analysis.

Mechanical System

$$\begin{bmatrix} f_{i1} \\ x_{i1} \end{bmatrix} = \begin{bmatrix} 1 & m_1 s^2 + b_1 s \\ 0 & 1 \end{bmatrix} \begin{bmatrix} f_{o1} \\ x_{o1} \end{bmatrix}$$

$$\begin{bmatrix} f_{i2} \\ x_{i2} \end{bmatrix} = \begin{bmatrix} 1 & K_2 \\ \frac{1}{b_2 s} & \frac{b_2 s + K_2}{b_2 s} \end{bmatrix} \begin{bmatrix} f_{o2} \\ x_{o2} \end{bmatrix}$$

Desired Transfer Function: $\frac{x_{o2}}{f_{i1}}$

$\left.\begin{matrix} x_{o1} = x_{i2} \\ f_{o1} = f_{i2} \end{matrix}\right\}$ when connected

Assume $f_{o2} = 0$, i.e., there is no system connected to the output of system 2.

$$\frac{x_{o2}}{f_{i1}} = \left[\frac{x_{o1}}{f_{i1}}\right]_{\text{Loaded}} \left[\frac{x_{o2}}{x_{i2}}\right]_{\text{Unloaded}}$$

$$\chi_{01} = \chi_{01}(f_{i1}, f_{01})$$

$$= \left.\frac{\chi_{01}}{f_{i1}}\right|_{f_{01}=0} f_{i1} + \left.\frac{\chi_{01}}{f_{01}}\right|_{f_{i1}=0} f_{01}$$

$$\left.\begin{array}{l} f_{01} = f_{12} \\ \chi_{01} = \chi_{12} \end{array}\right\}$$ systems are connected.

$$f_{12} = f_{12}(f_{02}, \chi_{12})$$

$$= \left.\frac{f_{12}}{f_{02}}\right|_{\chi_{12}=0} f_{02} + \left.\frac{f_{12}}{\chi_{12}}\right|_{f_{02}=0} \chi_{12}$$

$$= \left.\frac{f_{12}}{\chi_{12}}\right|_{f_{02}=0} \chi_{12}$$ since $f_{02} = 0$

$$= \left.\frac{f_{12}}{\chi_{12}}\right|_{f_{02}=0} \chi_{01}$$ since $\chi_{01} = \chi_{12}$

$$= f_{01}$$ since $f_{01} = f_{12}$

Therefore

$$x_{o1} = \left.\frac{x_{o1}}{f_{i1}}\right|_{f_{o1}=0} f_{i1} + \left.\frac{x_{o1}}{f_{o1}}\right|_{f_{i1}=0} \left.\frac{f_{i2}}{x_{i2}}\right|_{f_{o2}=0} x_{o1}$$

$$x_{o1} = \left.\frac{x_{o1}}{f_{i1}}\right|_{f_{o1}=0} f_{i1} + \frac{\left.\frac{x_{o1}}{f_{o1}}\right|_{f_{i1}=0}}{\left.\frac{x_{i2}}{f_{i2}}\right|_{f_{o2}=0}} x_{o1}$$

Solve for $\frac{x_{o1}}{f_{i1}}$:

$$\frac{x_{o1}}{f_{i1}} = \left.\frac{x_{o1}}{f_{i1}}\right|_{f_{o1}=0} \left[\frac{1}{1 - \frac{\left.\frac{x_{o1}}{f_{o1}}\right|_{f_{i1}=0}}{\left.\frac{x_{i2}}{f_{i2}}\right|_{f_{o2}=0}}} \right]$$

Apply this to systems 1 and 2:

$$\begin{bmatrix} f_{i1} \\ x_{i1} \end{bmatrix} = \begin{bmatrix} 1 & m_1 s^2 + b_1 s \\ 0 & 1 \end{bmatrix} \begin{bmatrix} f_{o1} \\ x_{o1} \end{bmatrix}$$

$$\begin{bmatrix} f_{i2} \\ x_{i2} \end{bmatrix} = \begin{bmatrix} 1 & K_2 \\ \frac{1}{b_2 s} & \frac{b_2 s + K_2}{b_2 s} \end{bmatrix} \begin{bmatrix} f_{o2} \\ x_{o2} \end{bmatrix}$$

$$\left.\frac{x_{o1}}{f_{i1}}\right|_{f_{o1}=0} = \frac{1}{m_1 s^2 + b_1 s}$$

$$\left.\frac{x_{o1}}{f_{o1}}\right|_{f_{i1}=0} = \frac{-1}{m_1 s^2 + b_1 s}$$

$$\left.\frac{x_{i2}}{f_{i2}}\right|_{f_{o2}=0} = \frac{b_2 s + K_2}{K_2 b_2 s}$$

Combine:

$$\frac{x_{o1}}{f_{i1}} = \frac{1}{m_1 s^2 + b_1 s}\left[\frac{1}{1 - \frac{-1/m_1 s^2 + b_1 s}{b_2 s + K_2 / K_2 b_2 s}}\right]$$

$$= \frac{b_2 s + K_2}{(m_1 s^2 + b_1 s)(b_2 s + K_2) + b_2 K_2 s}$$

= Loaded Transfer Function for System #1

$$\left.\frac{x_{o2}}{x_{i2}}\right|_{f_{o2}=0} = \frac{b_2 s}{b_2 s + K_2}$$

= Unloaded Transfer Function for System #2

Result:

$$\frac{x_{o2}}{f_{i1}} = \frac{b_2 s + K_2}{(m_1 s^2 + b_1 s)(b_2 s + K_2) + b_2 K_2 s} \cdot \frac{b_2 s}{b_2 s + K_2}$$

$$= \frac{\dfrac{b_2}{(b_1+b_2)K_2}}{\dfrac{m_1 b_2}{(b_1+b_2)K_2} s^2 + \dfrac{m_1 K_2 + b_1 b_2}{(b_1+b_2)K_2} s + 1}$$

Same as obtained previously.

Hydraulic System

$$\begin{bmatrix} h_{i1} \\ q_{i1} \end{bmatrix} = \begin{bmatrix} 1 & R_1 \\ A_1 s & A_1 R_1 s + 1 \end{bmatrix} \begin{bmatrix} h_{o1} \\ q_{o1} \end{bmatrix}$$

$$\begin{bmatrix} h_{i2} \\ q_{i2} \end{bmatrix} = \begin{bmatrix} 1 & R_2 \\ A_2 s & A_2 R_2 s + 1 \end{bmatrix} \begin{bmatrix} h_{o2} \\ q_{o2} \end{bmatrix}$$

$$\left.\begin{array}{l} q_{o1} = q_{i2} \\ h_{o1} = h_{i2} \end{array}\right\} \text{ when connected}$$

$h_{o2} = 0 \Rightarrow$ unloaded

Desired Transfer function: $\dfrac{q_{o2}}{q_{i1}}$

From previous analysis:

$$\left.\frac{q_{o2}}{q_{i1}}\right|_{h_{o2}=0} = \frac{1}{(A_1 R_1 s + 1)(A_2 R_2 s + 1) + A_1 R_2 s}$$

Now Use Impedances.

$$q_{01} = q_{01}(q_{l1}, h_{01})$$

$$= \left.\frac{q_{01}}{q_{l1}}\right|_{h_{01}=0} q_{l1} + \left.\frac{q_{01}}{h_{01}}\right|_{q_{l1}=0} h_{01}$$

$$q_{l2} = q_{l2}(h_{l2}, h_{02})$$

$$= \left.\frac{q_{l2}}{h_{l2}}\right|_{h_{02}=0} h_{l2} + \left.\frac{q_{l2}}{h_{02}}\right|_{h_{l2}=0} h_{02}$$

But: $h_{02} = 0$

$q_{l2} = q_{01}$

$h_{01} = h_{l2}$

$$q_{01} = \left.\frac{q_{01}}{q_{l1}}\right|_{h_{01}=0} q_{l1} + \left.\frac{q_{01}}{h_{01}}\right|_{q_{l1}=0} \frac{q_{01}}{\left.\frac{q_{l2}}{h_{l2}}\right|_{h_{02}=0}}$$

Solve for $\dfrac{q_{o1}}{q_{i1}}$:

$$\frac{q_{o1}}{q_{i1}} = \left.\frac{q_{o1}}{q_{i1}}\right|_{h_{o1}=0} \left[\frac{1}{1 - \dfrac{\left.\dfrac{q_{o1}}{h_{o1}}\right|_{q_{i1}=0}}{\left.\dfrac{q_{i2}}{h_{i2}}\right|_{h_{o2}=0}}} \right]$$

$$\left.\frac{q_{o1}}{q_{i1}}\right|_{h_{o1}=0} = \frac{1}{A_1 R_1 s + 1}$$

$$\left.\frac{q_{o1}}{h_{o1}}\right|_{q_{i1}=0} = \frac{-A_1 s}{A_1 R_1 s + 1}$$

$$\left.\frac{q_{i2}}{h_{i2}}\right|_{h_{o2}=0} = \frac{A_2 R_2 s + 1}{R_2}$$

$$\frac{q_{o1}}{q_{i1}} = \frac{1}{A_1R_1s+1}\left[\frac{1}{1-\frac{\frac{-A_1s}{A_1R_1s+1}}{\frac{A_2R_2s+1}{R_2}}}\right]$$

$$\frac{q_{o2}}{q_{i1}} = \left.\frac{q_{o1}}{q_{i1}}\right|_{Loaded} \left.\frac{q_{o2}}{q_{i2}}\right|_{Unloaded}$$

Note that $q_{o1} = q_{i2}$

$$\left.\frac{q_{o2}}{q_{i2}}\right|_{h_{o2}=0} = \frac{1}{A_2R_2s+1}$$

Result:

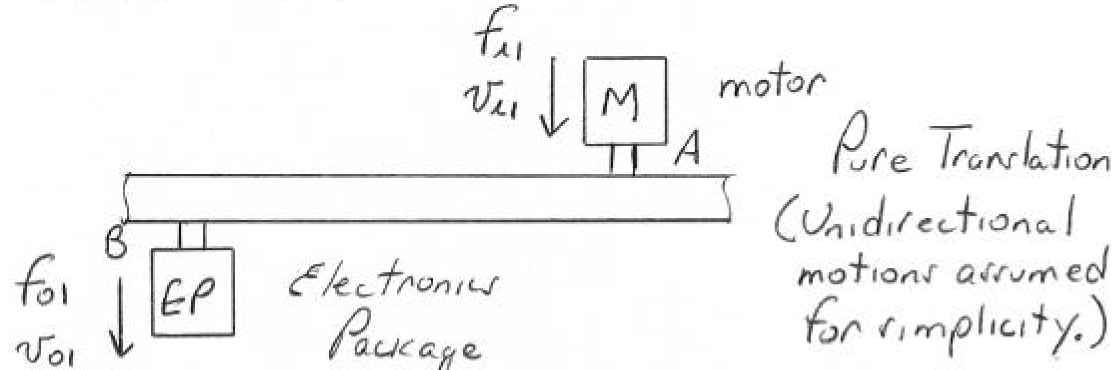
$$\frac{q_{o2}}{q_{i1}} = \frac{1}{A_1R_1s+1}\left[\frac{1}{1+\frac{A_1R_2s}{(A_1R_1s+1)(A_2R_2s+1)}}\right]\frac{1}{A_2R_2s+1}$$

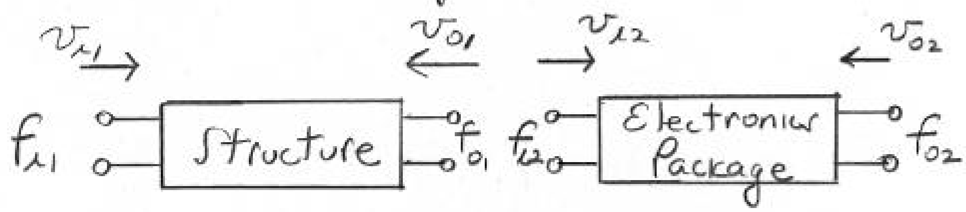
Same as previously determined.

III. Practical Application

- In equipment involving diverse technologies, major subsystems may be manufactured by different contractors at remote locations, with subsystems being brought together at a single final assembly point only after each has been individually completed.
- Experimental studies of the complete system cannot be performed until final assembly. Discovery of design faults at this late stage can cause severe economic and scheduling problems.
- Capability for experimental testing of each subsystem at the respective manufacturer's facility and proper coupling of these results to predict behavior of the assembled system can be a valuable tool.

- Consider the following physical situation:





$f \equiv$ force

$v \equiv$ velocity

- Motor produces vibration-exciting dynamic forces f_{i1} at location A.
- Electronic package at location B can withstand only limited vibration.
- Determine the force f_{i2} that will be applied to the electronics package when it has been connected.

- It is preferable to run separate vibration tests on each subsystem and then calculate from these measurements what the force will be.
- Assume that the frequency spectrum of the input force f_{i1} is known from theory or experiment.

<u>Problem:</u> Find $\frac{f_{i2}}{f_{i1}}(s)$ for the loaded condition

$f_{i2} = f_{o1}$ when the subsystems are joined.

$f_{o2} = 0$ since the electronics package is allowed to vibrate freely.

$v_{o1} = -v_{i2}$ when the subsystems are joined.

$$f_{o1} = f_{o1}(f_{i1}, v_{o1})$$

$$f_{o1} = \left.\frac{f_{o1}}{f_{i1}}\right|_{v_{o1}=0} f_{i1} + \left.\frac{f_{o1}}{v_{o1}}\right|_{f_{i1}=0} v_{o1}$$

$$f_{i2} = f_{i2}(v_{i2}, f_{o2})$$

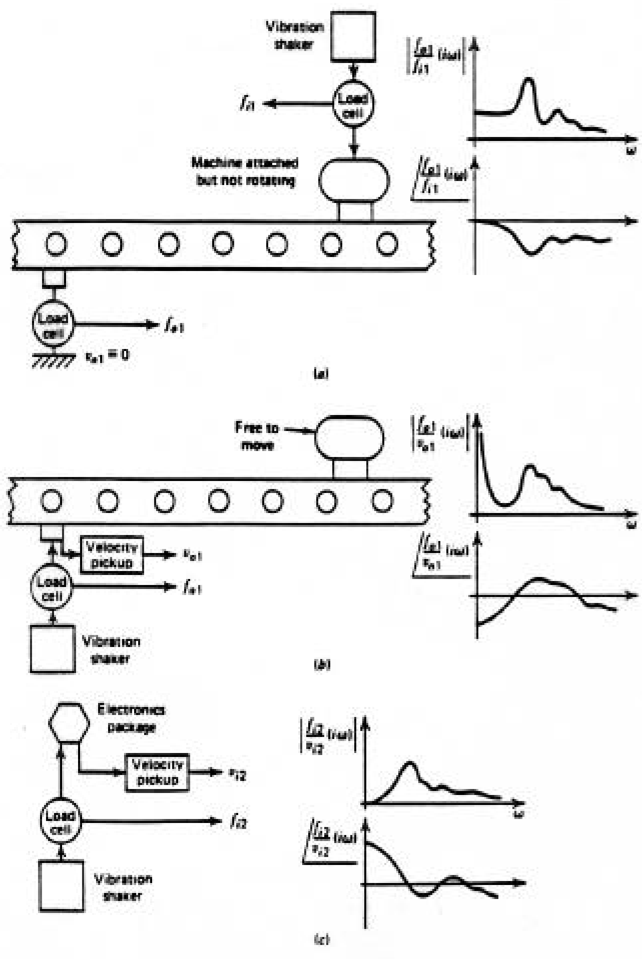
$$f_{i2} = \left.\frac{f_{i2}}{v_{i2}}\right|_{f_{o2}=0} v_{i2} + \left.\frac{f_{i2}}{f_{o2}}\right|_{v_{i2}=0} f_{o2}$$

Combine:

$$f_{o1} = \left.\frac{f_{o1}}{f_{i1}}\right|_{v_{o1}=0} f_{i1} + \left.\frac{f_{o1}}{v_{o1}}\right|_{f_{i1}=0} \left[\frac{-f_{i2}}{\left.\frac{f_{i2}}{v_{i2}}\right|_{f_{o2}=0}} \right]$$

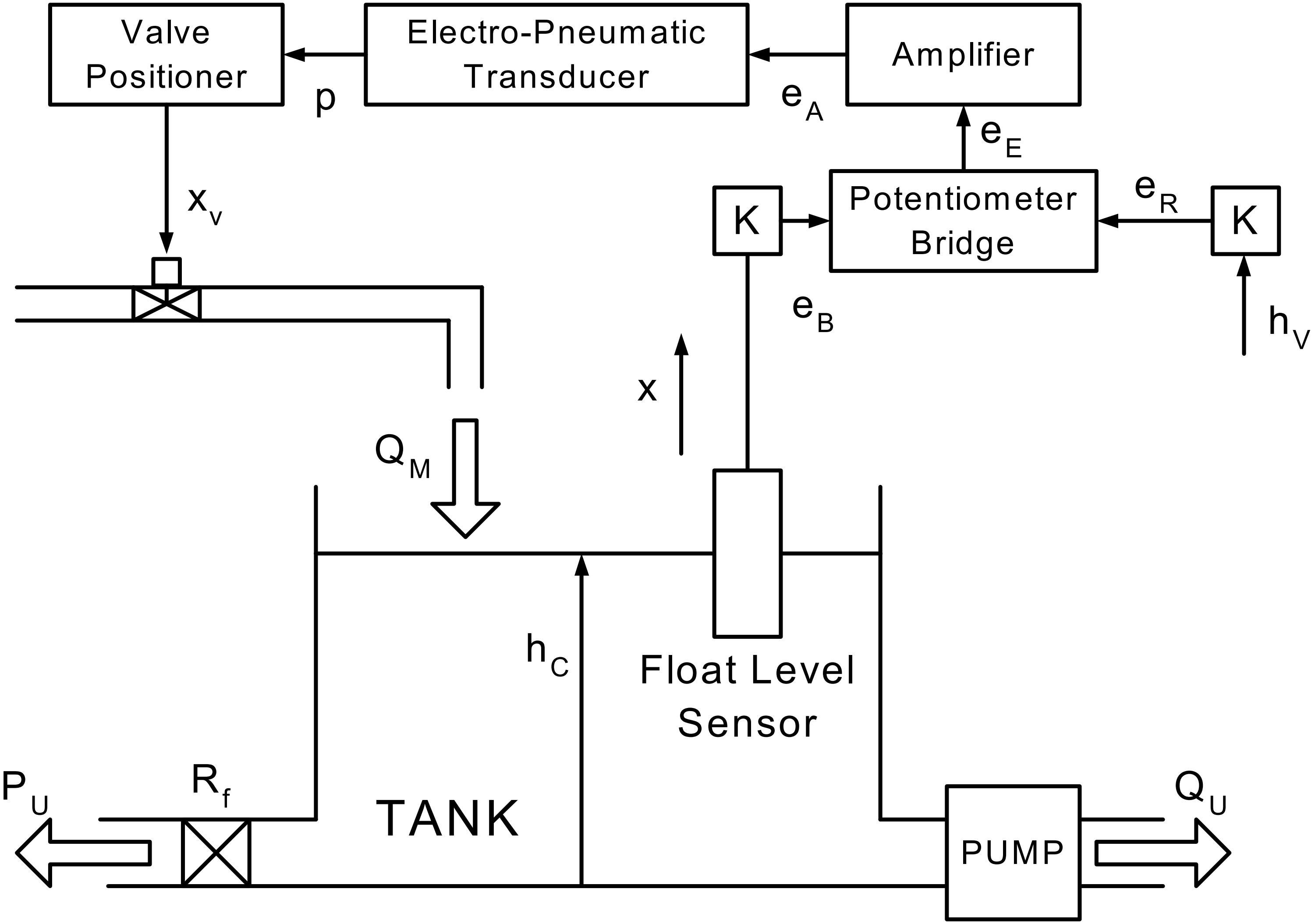
But $f_{i2} = f_{o1}$

$$\frac{f_{o1}}{f_{i1}} = \left.\frac{f_{o1}}{f_{i1}}\right|_{v_{o1}=0} \left[\frac{1}{1 + \dfrac{\left.\dfrac{f_{o1}}{v_{o1}}\right|_{f_{i1}=0}}{\left.\dfrac{f_{i2}}{v_{i2}}\right|_{f_{o2}=0}}} \right]$$



Experimental Measurements needed to predict coupled-system response.

Control of a Liquid-Level Process



- **Objective**
  - Maintain tank level h_C at the desired level h_V in the face of disturbances pressure $P_U(t)$ (psig) and volume flow rate $Q_U(t)$ (ft^3/sec). R_f is a linearized flow resistance with units psi/(ft^3/sec).
- **Equilibrium Operating Point**
  - All variables are steady
  - Inflow Q_M exactly matches the two outflows
  - $h_C = h_V$ and $e_E = 0$
    - When $e_E = 0$, Q_M can be nonzero since the electropneumatic transducer has a zero adjustment and the valve positioner has a zero adjustment, e.g., p = 9 psig and the valve opening corresponds to equilibrium flow Q_M.
    - We will deal with small perturbations in all variables away from the initial steady state.

- Assumptions and Equations of Motion
- Tank Process Dynamics
  - Density of fluid ρ is constant.

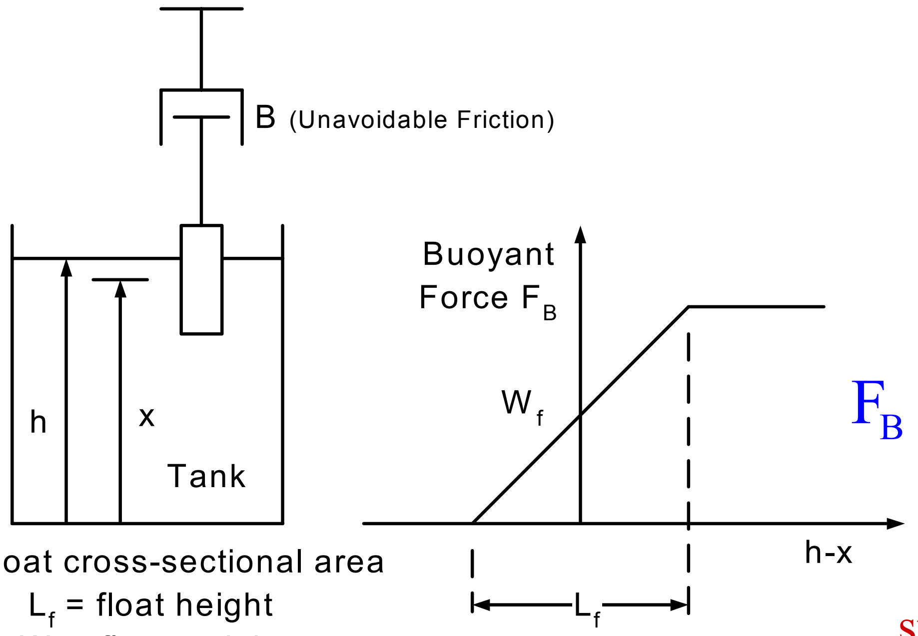
$$Q_M - Q_U - \frac{\rho g h_C - P_U}{R_f} = A_T \frac{dh_C}{dt}$$

$$\left(\tau_p s + 1\right) h_C = \frac{R_f}{\rho g} Q_M + \frac{1}{\rho g} P_U - \frac{R_f}{\rho g} Q_U$$

Conservation of Mass

$$\tau_p = \frac{A_T R_f}{\rho g}$$ process time constant

- Float Level Sensor
  - Assume a zero-order dynamic model, i.e., the dynamics are negligible relative to the process time constant τ_p since the cross-sectional area of the tank is assumed large.
  - Consider the actual dynamics to justify this assumption:



A_f = float cross-sectional area
L_f = float height
W_f = float weight

$$F_B = W_f + A_f \rho g (h - x)$$

steady condition: h = x

– Equation of Motion

$$F_B - W_f - B\frac{dx}{dt} = M_f\frac{d^2x}{dt^2}$$

$$M_f\frac{d^2x}{dt^2} + B\frac{dx}{dt} = A_f\rho g\left(h - x\right)$$

$$M_f\frac{d^2x}{dt^2} + B\frac{dx}{dt} + A_f\rho g x = A_f\rho g h$$

$$\left[\frac{D^2}{\omega_n^2} + \frac{2\zeta D}{\omega_n} + 1\right]x = Kh$$

$$\omega_n = \sqrt{\frac{A_f\rho g}{M_f}}$$

$$D \equiv \frac{d}{dt}$$

$$D^2 \equiv \frac{d^2}{dt^2}$$

Differential Operator

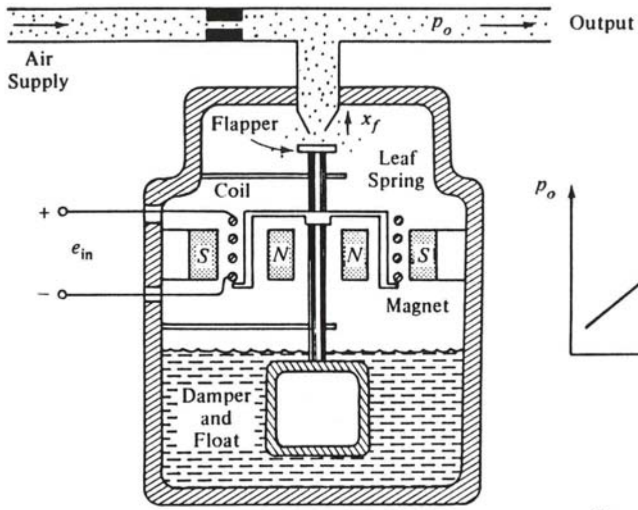
$$\zeta = \frac{B}{2\sqrt{A_f M_f \rho g}}$$

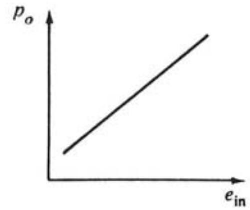
$$K = 1.0$$

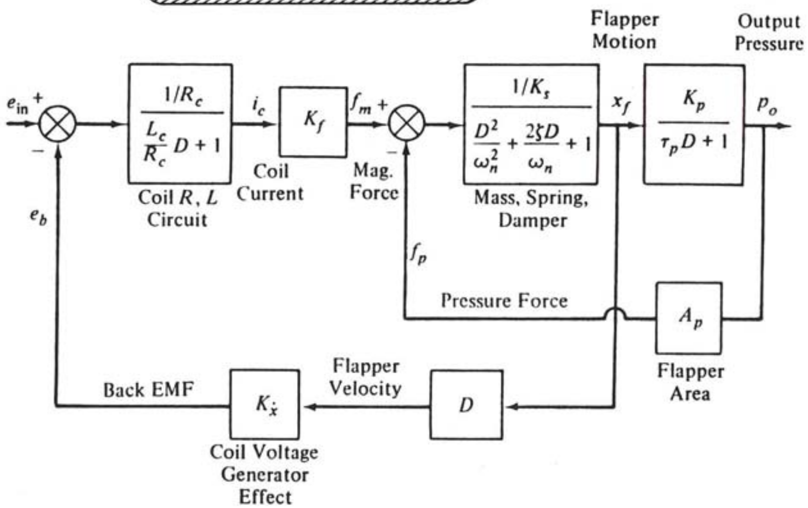
- To measure rapid changes in h accurately, ω_n must be sufficiently large. The specific weight of the fluid (ρg) is not a design variable, so strive for large values of A_f/M_f (i.e., hollow floats).
- In our case, the tank has a large diameter and if the inflow and outflow rates are modest, h cannot change rapidly and so a zero-order model is justified.

- <u>Potentiometer Bridge and Electronic Amplifier</u>
  - Obviously these two components are fast enough to be treated as zero order in this system.

• Electropneumatic Transducer







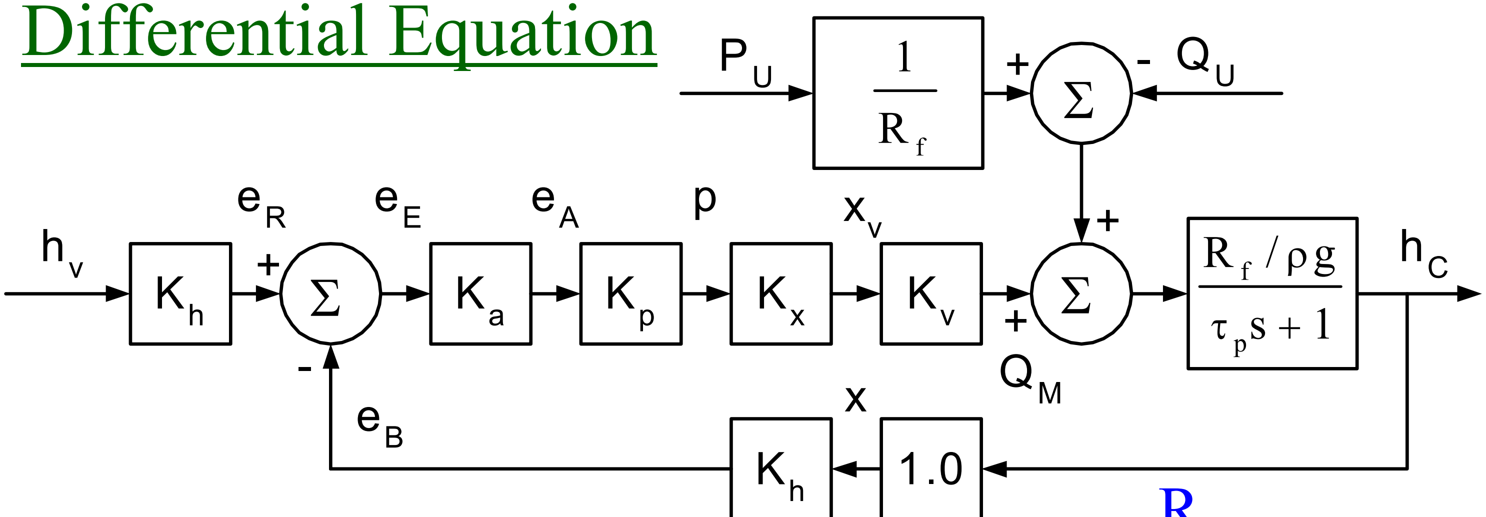
This device produces a pneumatic output signal closely proportional (± 5% nonlinearity) to an electrical input (± 5V and 3-15 psig).

We are concerned with overall dynamics from e_A to p. The block diagram shows a 4^{th}-order closed-loop differential equation.

However, experimental frequency response tests show typically a flat amplitude ratio out to about 5 Hz. This response is very fast relative to τ_p so we model the electropneumatic transducer as zero order.

- Pneumatic Valve Positioner
  - We are only interested in the overall dynamics relating x_v to p. These are again quite fast relative to τ_p, so we model the component as zero order.
  - The valve positioner allows one to "characterize" the static calibration curve between p and x_v and thus obtain desired linear or nonlinear relationships between p and manipulated flowrate Q_M.
- Relation between Q_M and x_v
  - This relationship is assumed to be statically linear and dynamically instantaneous and thus a zero-order model.
  - Although the dynamic response of Q_M to x_v is not instantaneous due to fluid inertia and compliance, the response is much faster than the tank-filling dynamics.

- ## Closed-Loop System Block Diagram and Differential Equation



$$\left[\left(K_h h_V - K_h h_C\right) K_a K_p K_x K_v + \frac{1}{R_f} P_U - Q_U\right] \frac{\frac{R_f}{\rho g}}{\tau_p s + 1} = h_C$$

$$\left(\tau_s s + 1\right) h_C = \frac{K}{K+1} h_V + \frac{1}{\rho g (K+1)} P_U - \frac{R_f}{\rho g (K+1)} Q_U$$

$$\tau_s = \frac{\tau_p}{K+1}$$ Closed-Loop System Time Constant

$$K = \frac{1}{\rho g}\left(K_h K_a K_p K_x K_v R_f\right)$$ System Loop Gain

- ## Speed of Response
  - Response for a step input in h_v (hold perturbations P_U and Q_U at zero)

$$h_C = \frac{K}{K+1} h_{V_s} \left(1 - e^{\frac{-t}{\tau_s}} \right)$$

  - Response for a step input in disturbances P_U and Q_U (hold $h_V = 0$)

$$h_C = \frac{1}{\rho g (K+1)} P_{U_s} \left(1 - e^{\frac{-t}{\tau_s}} \right) \qquad h_C = \frac{-R_f}{\rho g (K+1)} Q_{U_s} \left(1 - e^{\frac{-t}{\tau_s}} \right)$$

  - Increasing loop gain K increases the speed of response

$$\tau_s = \frac{\tau_p}{K+1} = \text{closed-loop system time constant}$$

- <u>Steady-State Errors</u>
  - A procedure generally useful for all types of systems and inputs is to rewrite the closed-loop system differential equation with system error (V-C), rather than the controlled variable C, as the unknown.
  - In this case we have:

$$\left(\tau_s s+1\right)h_C = \frac{K}{K+1}h_V + \frac{1}{\rho g\left(K+1\right)}P_U - \frac{R_f}{\rho g\left(K+1\right)}Q_U$$

$$h_E = h_V - h_C$$

$$\left(\tau_s s+1\right)\left(h_V - h_E\right) = \frac{K}{K+1}h_V + \frac{1}{\rho g\left(K+1\right)}P_U - \frac{R_f}{\rho g\left(K+1\right)}Q_U$$

$$\left(\tau_s s+1\right)h_E = \left(\tau_s s + \frac{1}{K+1}\right)h_V - \frac{1}{\rho g\left(K+1\right)}P_U + \frac{R_f}{\rho g\left(K+1\right)}Q_U$$

- For any chosen commands or disturbances, the steady-state error will just be the particular solution of the equation:

$$\left(\tau_s s + 1\right) h_E = \left(\tau_s s + \frac{1}{K+1}\right) h_V - \frac{1}{\rho g \left(K+1\right)} P_U + \frac{R_f}{\rho g \left(K+1\right)} Q_U$$

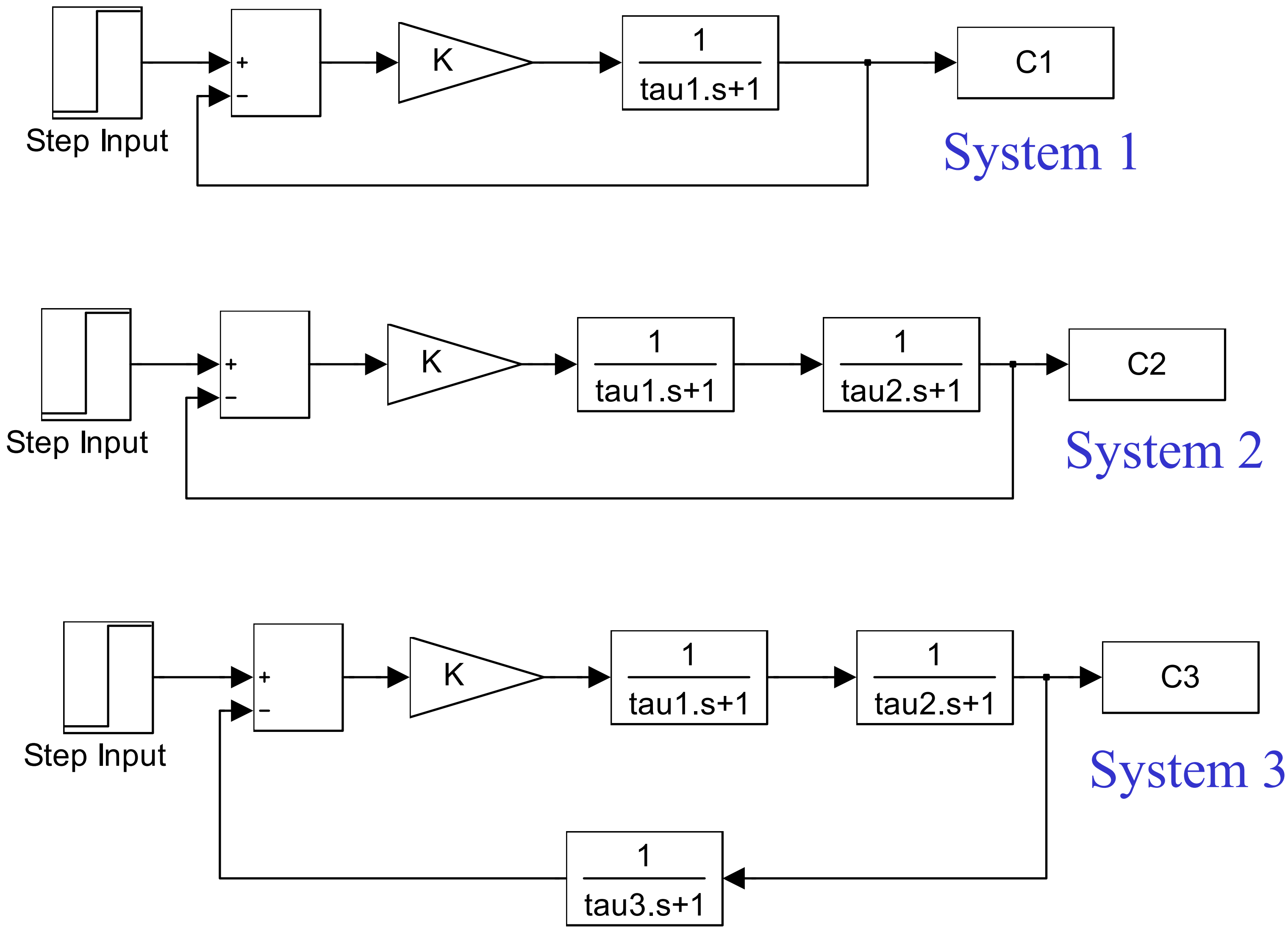
- We see that the steady-state error is improved if we increase the loop gain K.
- For any initial equilibrium condition we can "trim" the system for zero error but subsequent steady commands and/or disturbances must cause steady-state errors.
- Ramp inputs would cause steady-state errors that increase linearly with time, the rate of increase being proportional to ramp slope and inversely proportional to K+1.

- Stability
  - All aspects of system behavior are improved by increasing loop gain – up to a point! – instability may result, but our present model gives no warning of this. Why?
  - We neglected dynamics in some components and a general rule is:

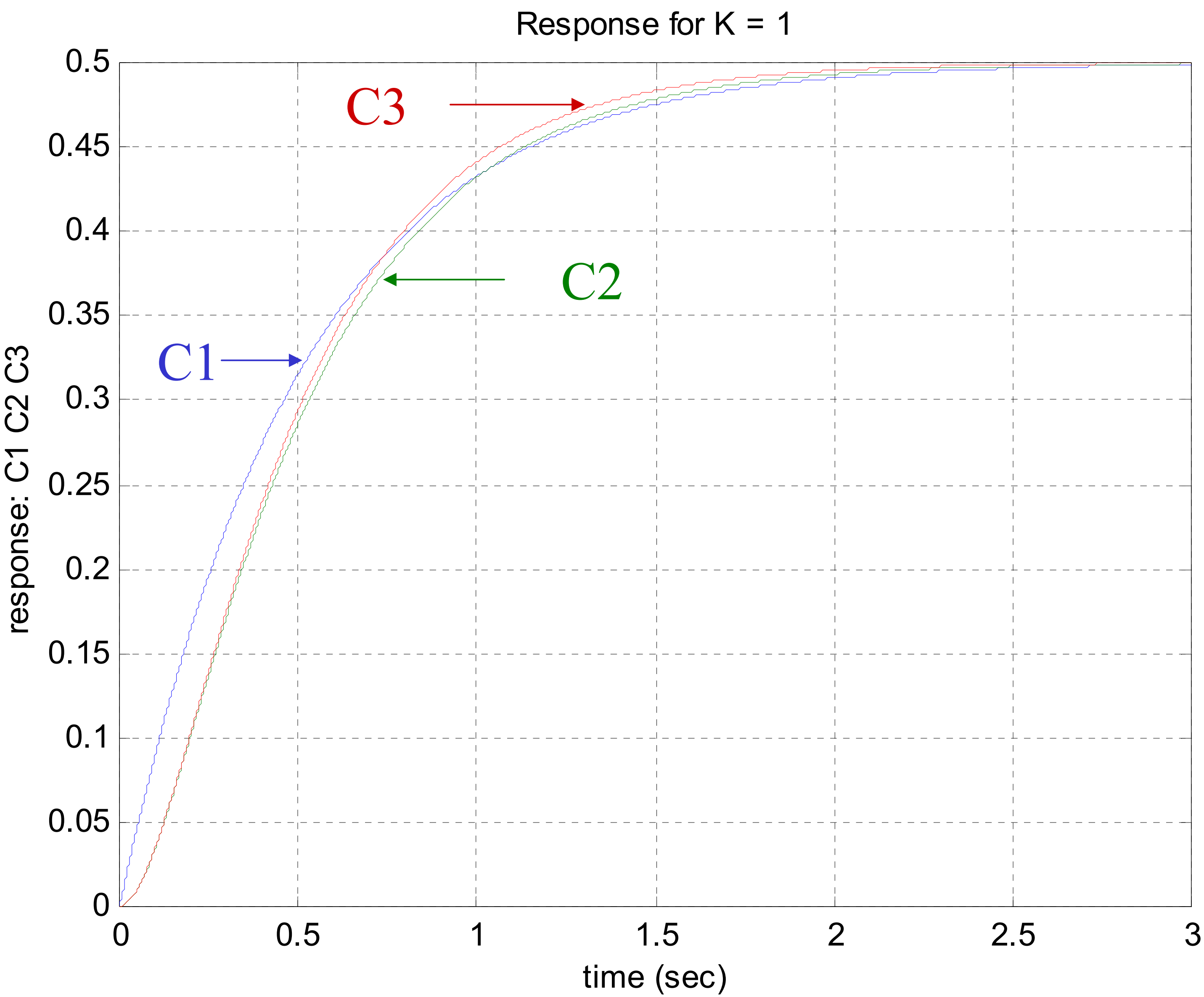
    If we want to make valid stability predictions we must include enough dynamics in our system so that the closed-loop system differential equation is at least third order. The one exception is systems with dead times where instability can occur even when dynamics are zero, first, or second order.
  - Is our model then useless?

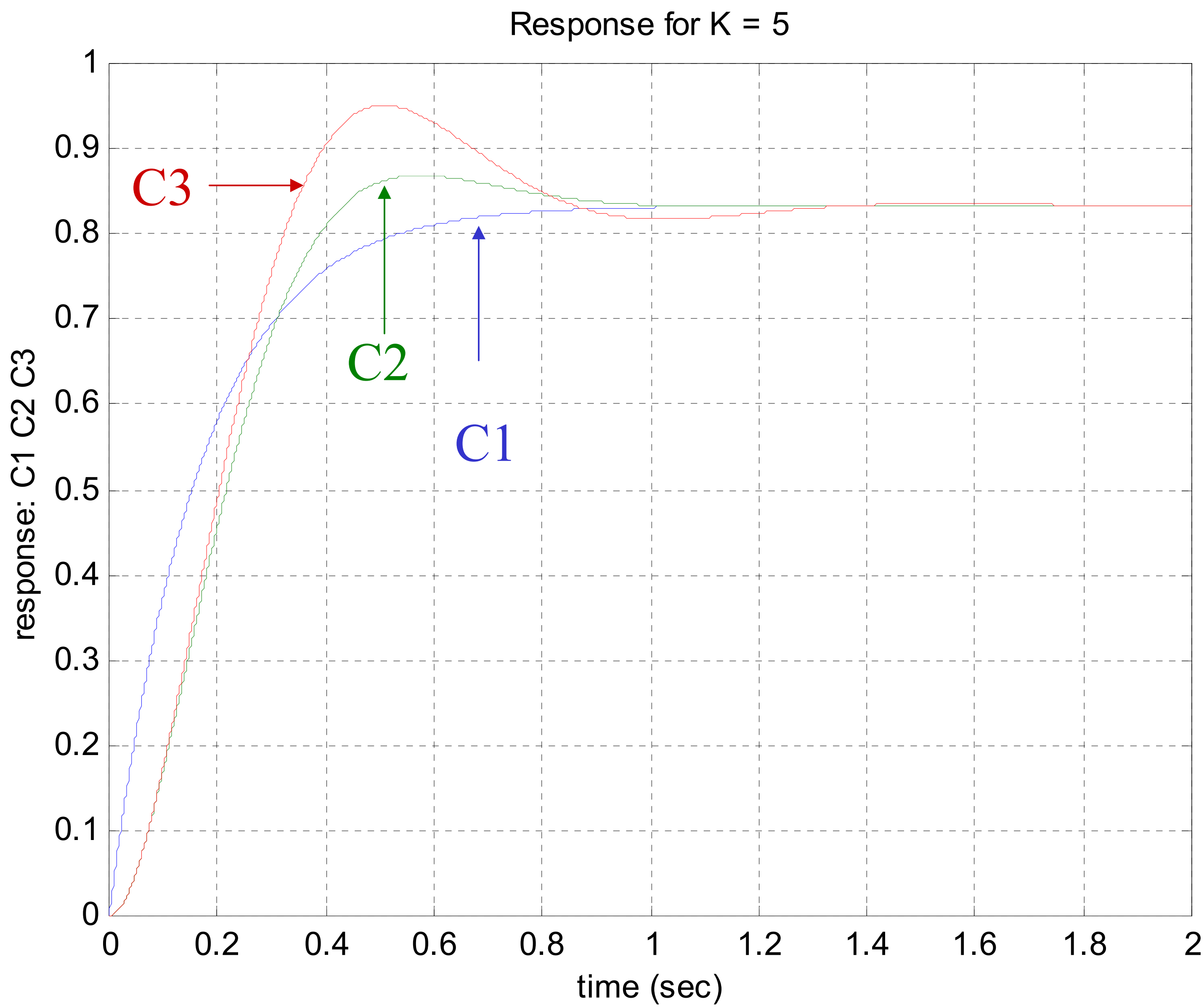
- No! It does correctly predict system behavior as long as the loop gain is not made "too large." As K is increased, the closed-loop system response gets faster and faster. At some point, the neglected dynamics are no longer negligible and the model becomes inaccurate.
- We neglected dynamics relative to τ_p, but in the closed-loop system, response speed is determined by τ_s.

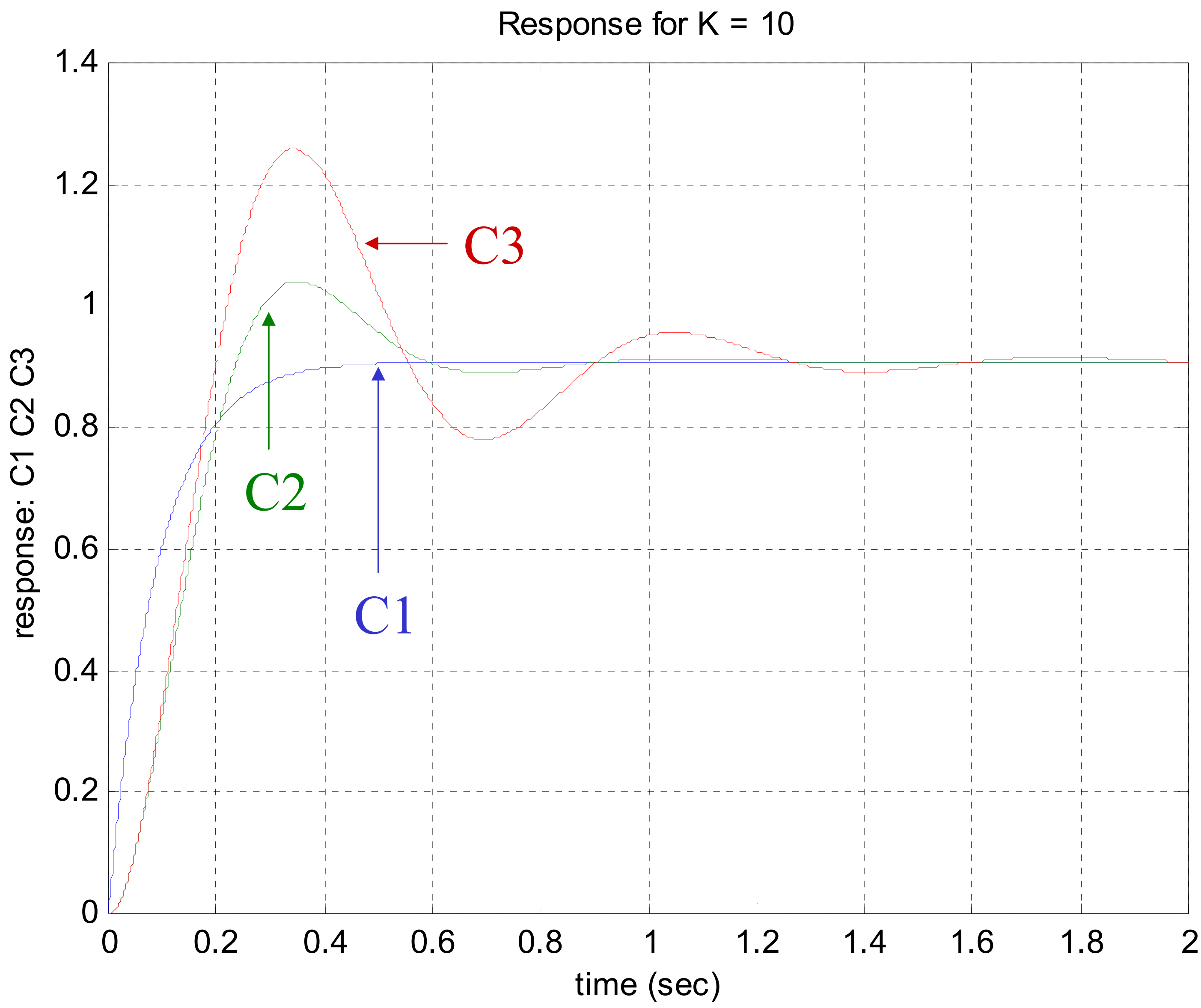
• <u>Exercise</u>:
  - Compare the responses of the following 3 systems for K = 1, 5, 10 with $\tau_1 = 1.0$, $\tau_2 = 0.1$, and $\tau_3 = 0.05$. The input is a unit step.
  - Examine speed of response, steady-state error, and stability predictions.

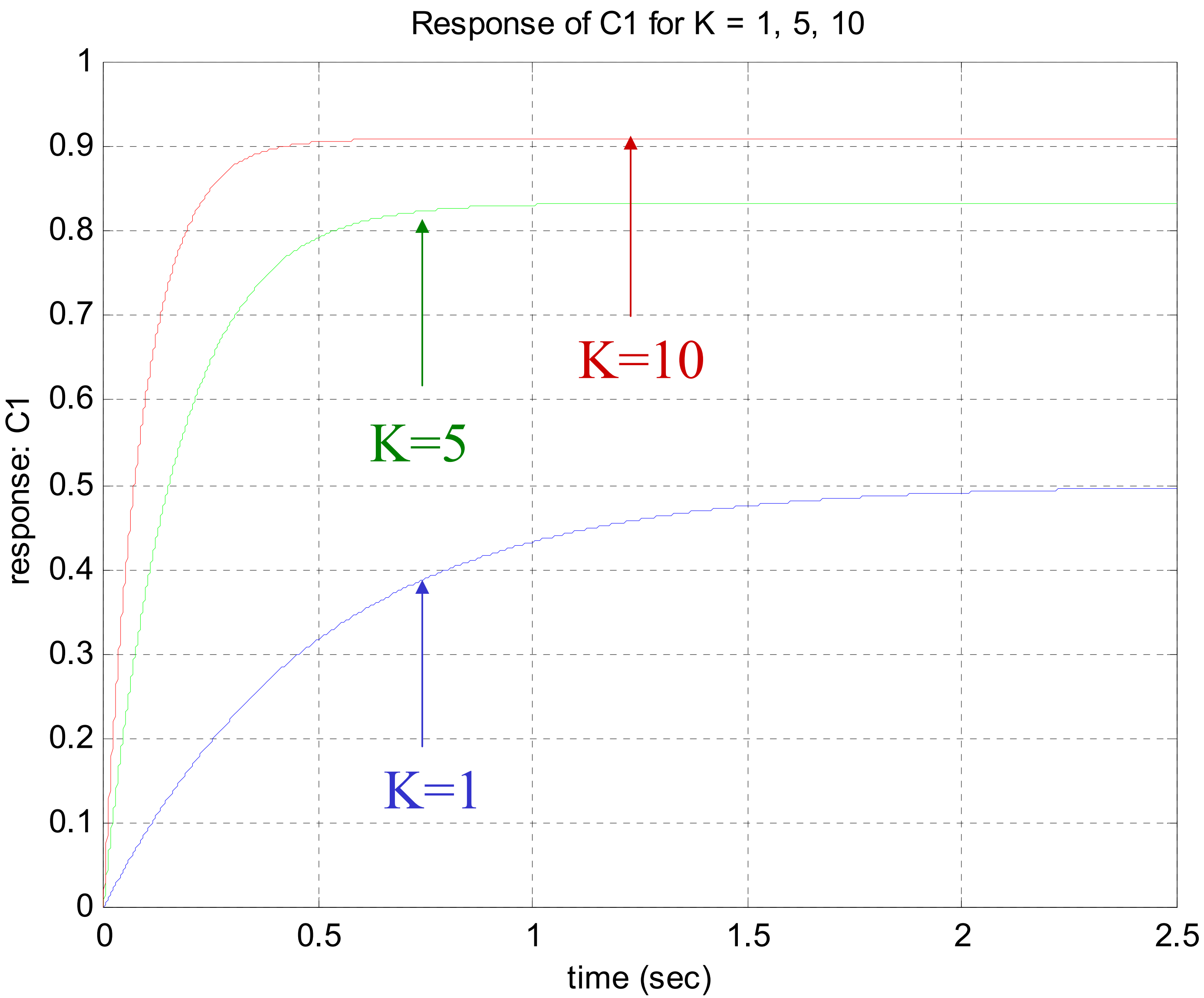


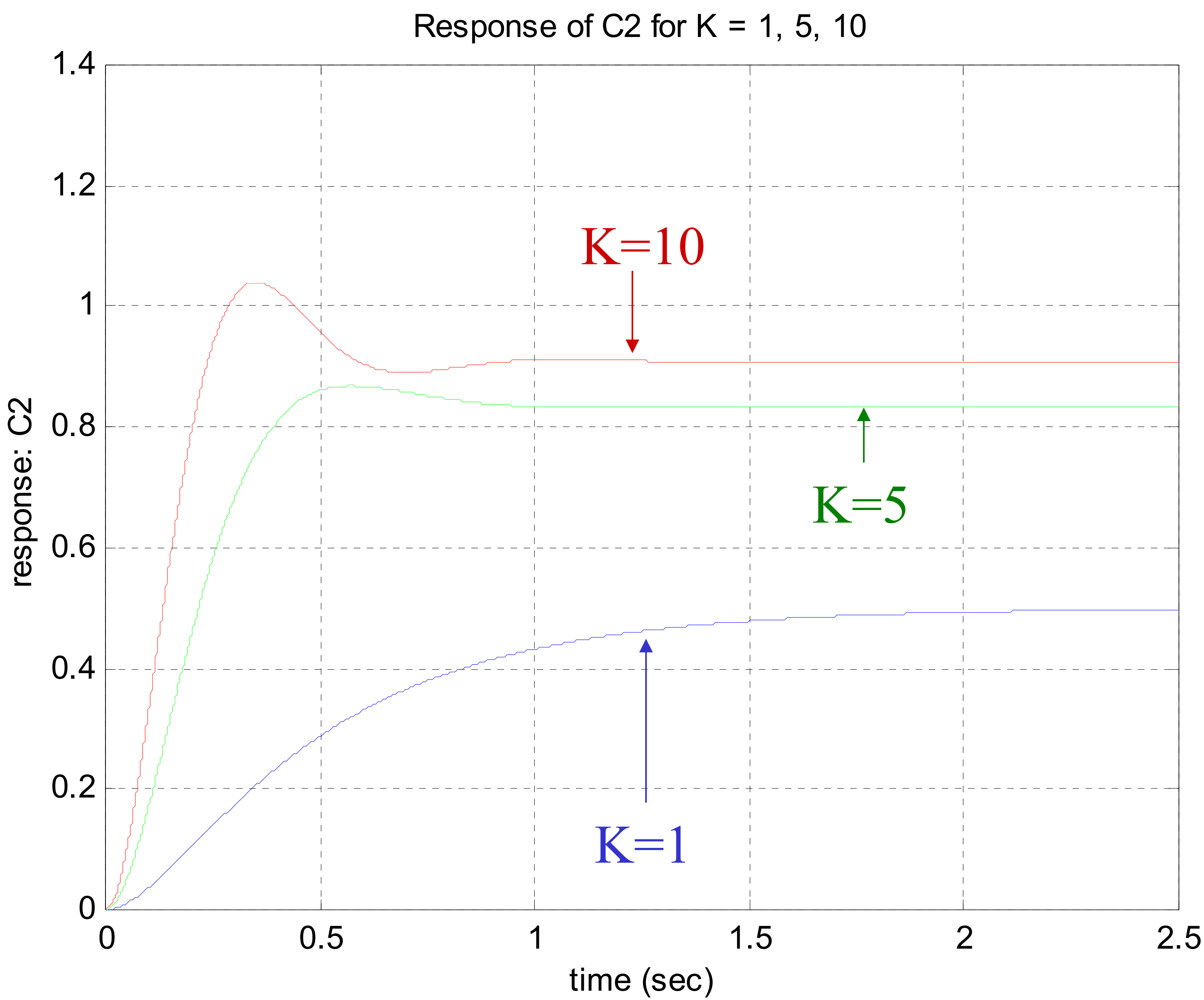
MatLab / Simulink Diagram

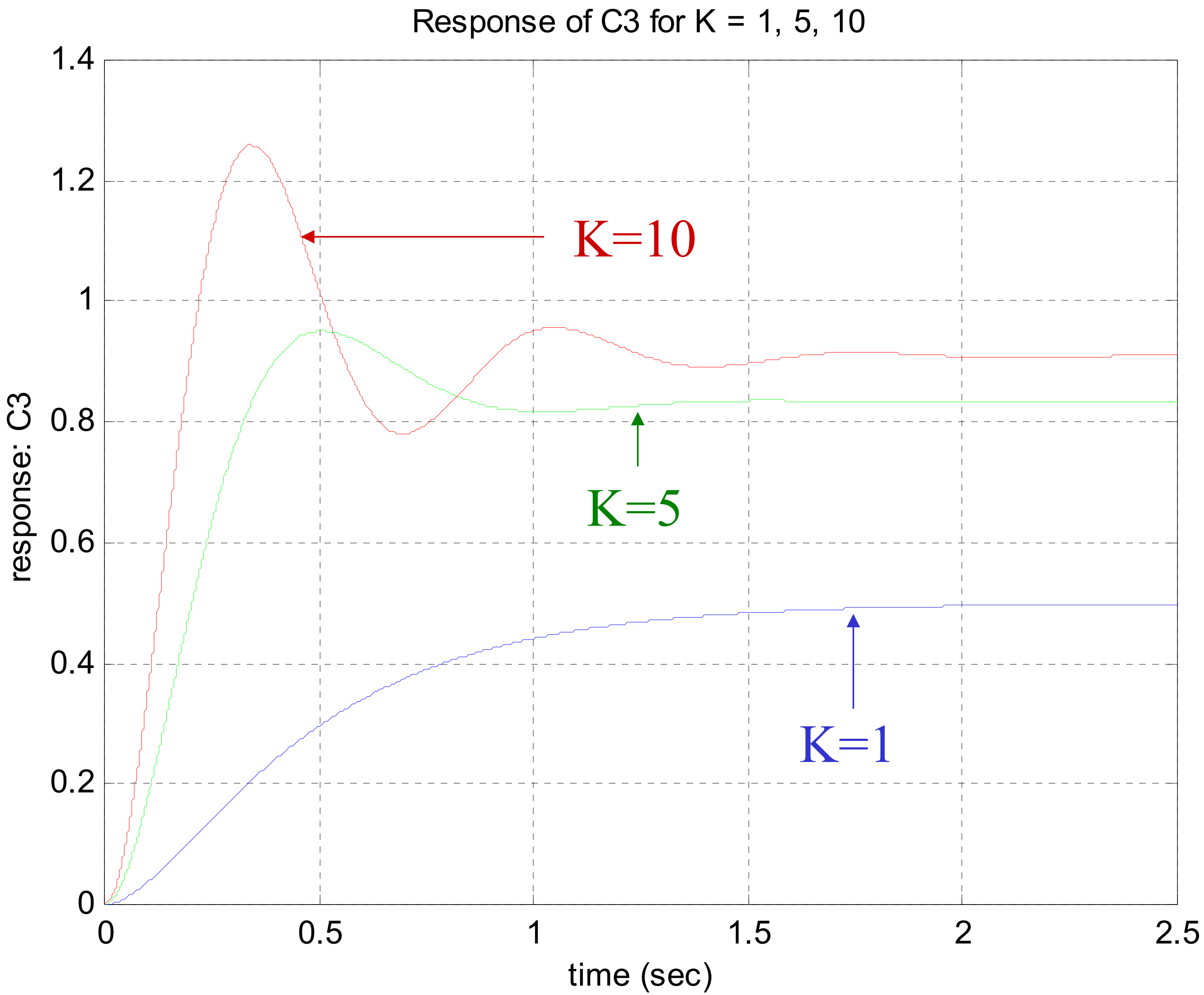

Response for K = 1
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
response: C1 C2 C3
C1
C2
C3
0
0.5
1
1.5
2
2.5
3
time (sec)


Response for K = 5
response: C1 C2 C3
time (sec)
C3
C2
C1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2


Response for K = 10
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
time (sec)
response: C1 C2 C3
C3
C2
C1


Response of C1 for K = 1, 5, 10
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
response: C1
0
0.5
1
1.5
2
2.5
time (sec)
K=10
K=5
K=1


Response of C2 for K = 1, 5, 10
1.4
1.2
1
0.8
0.6
0.4
0.2
0
response: C2
0
0.5
1
1.5
2
2.5
time (sec)
K=10
K=5
K=1


Response of C3 for K = 1, 5, 10
1.4
1.2
1
0.8
0.6
0.4
0.2
0
response: C3
0
0.5
1
1.5
2
2.5
time (sec)
K=10
K=5
K=1

– Note that as loop gain K is increased, the speed of response is increased and the steady-state error is reduced.

– For what value of loop gain K will any of these systems go unstable?

– Let's look at the closed-loop system transfer functions and characteristic equations:

Transfer Functions
(The characteristic equation is obtained by setting the denominator polynomial equal to zero.)

$$\frac{C1}{V} = \frac{K}{\tau_1 s + 1 + K}$$

$$\frac{C2}{V} = \frac{K}{\tau_1 \tau_2 s^2 + \left(\tau_1 + \tau_2\right) s + 1 + K}$$

$$\frac{C3}{V} = \frac{K\left(\tau_3 + 1\right)}{\tau_1 \tau_2 \tau_3 s^3 + \left(\tau_1 \tau_2 + \tau_2 \tau_3 + \tau_1 \tau_3\right) s^2 + \left(\tau_1 + \tau_2 + \tau_3\right) s + 1 + K}$$

$$\frac{C1}{V} = \frac{K}{s+1+K}$$

$$\frac{C2}{V} = \frac{K}{0.1s^2 + 1.1s + 1 + K}$$

$$\frac{C3}{V} = \frac{K(0.05s+1)}{0.005s^3 + 0.155s^2 + 1.15s + 1 + K}$$

– The only system which will go unstable as the loop gain K is increased is the third system; its characteristic equation is third order. The first two systems will continue to show improved speed of response and reduction of steady-state error without any hint of instability!

– Let's apply the three methods of determining closed-loop system stability – Routh, Nyquist, and Root-Locus - to the third system and determine the value of K for which this system becomes marginally stable.

– Routh Stability Criterion

- Closed-Loop System Characteristic Equation

$$0.005s^3 + 0.155s^2 + 1.15s + 1 + K = 0$$

- Routh Array

$$\begin{array}{cc} 0.005 & 1.15 \\ 0.155 & 1+K \\ \dfrac{(0.155)(1.15)-(1+K)(0.005)}{0.155} & 0 \\ 1+K & 0 \end{array}$$

- For stability we see that:

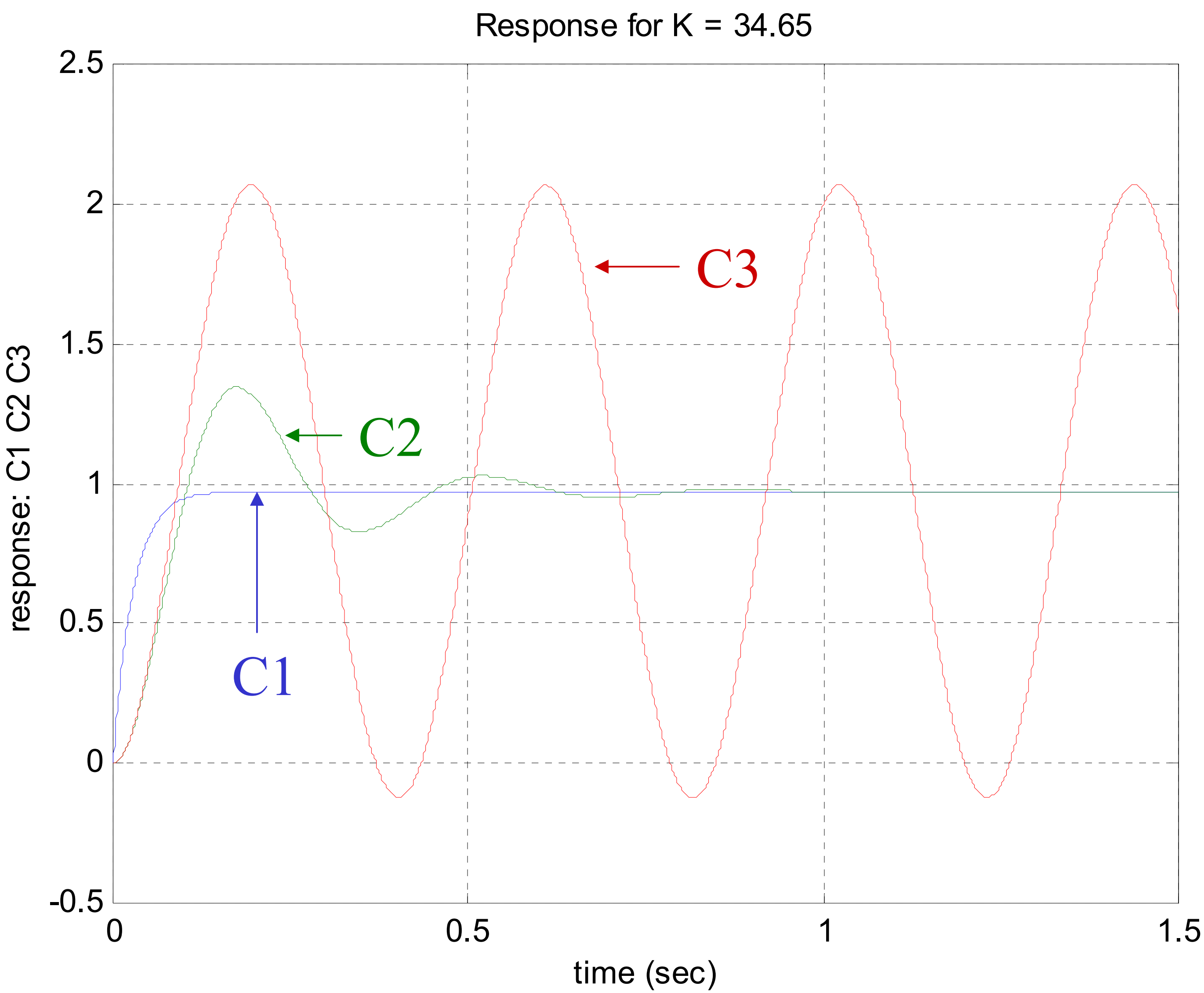
$$(0.155)(1.15) - (1+K)(0.005) > 0$$

$$1 + K > 0$$

- This leads to the result that for absolute stability:

$$-1 < K < 34.65$$

- A simulation with the loop gain set to K = 34.65 should verify this result. The value of gain K = -1 will give the closed-loop system characteristic equation a root at the origin but that value is of less interest, since we rarely use negative gain values.
- Note that at the loop gain value of 34.65, only system 3 is marginally stable. Systems 1 and 2 show no signs of instability, only improved speed of response and reduced steady-state error.


Response for K = 34.65
2.5
2
1.5
1
0.5
0
-0.5
0
0.5
1
1.5
time (sec)
response: C1 C2 C3
C3
C2
C1

– Root-Locus Interpretation of Stability

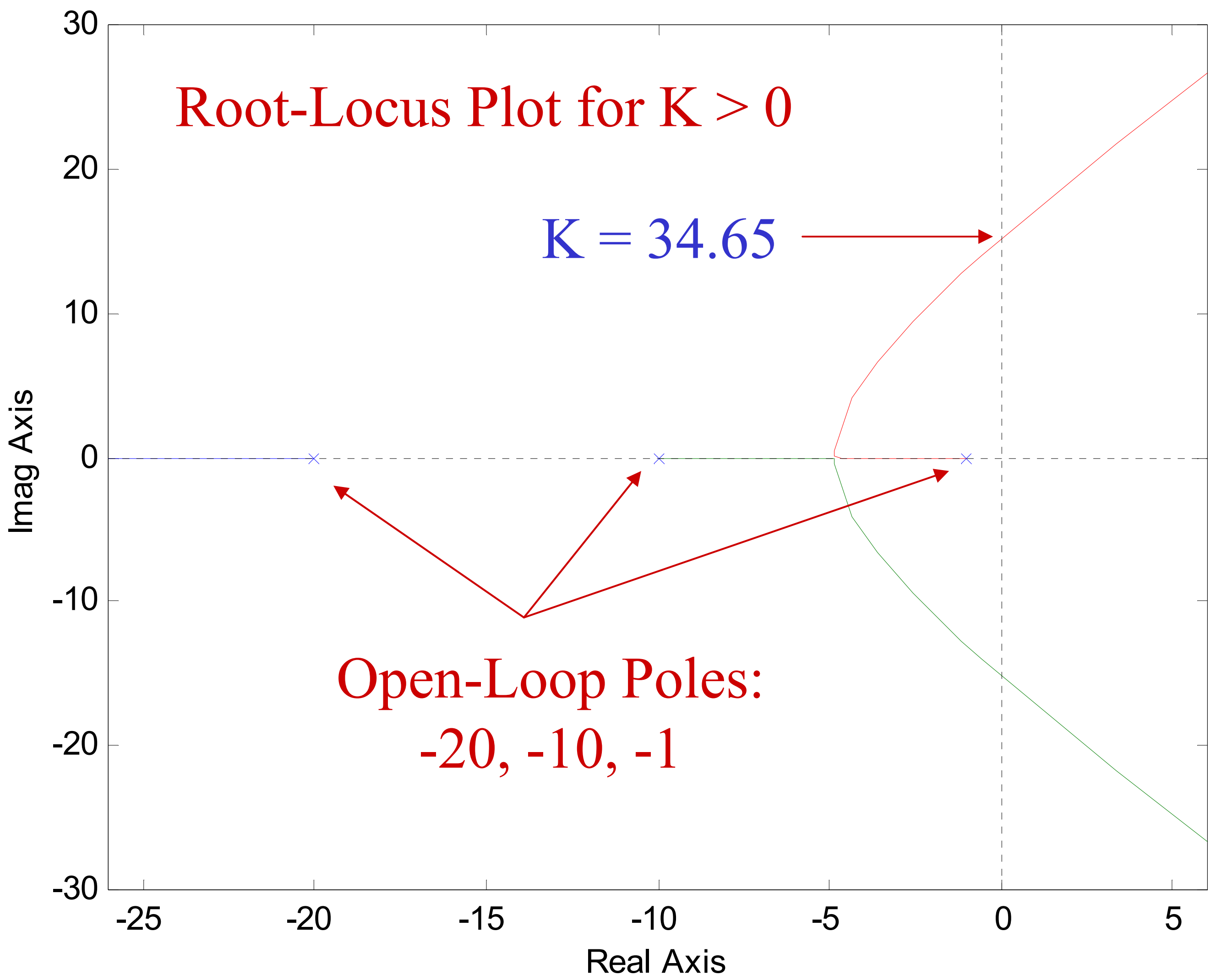
- The open-loop transfer function is:

$$\frac{K}{(\tau_1 s+1)(\tau_2 s+1)(\tau_3 s+1)}$$

$$=\frac{K}{\tau_1\tau_2\tau_3 s^3+(\tau_1\tau_2+\tau_2\tau_3+\tau_1\tau_3)s^2+(\tau_1+\tau_2+\tau_3)s+1}$$

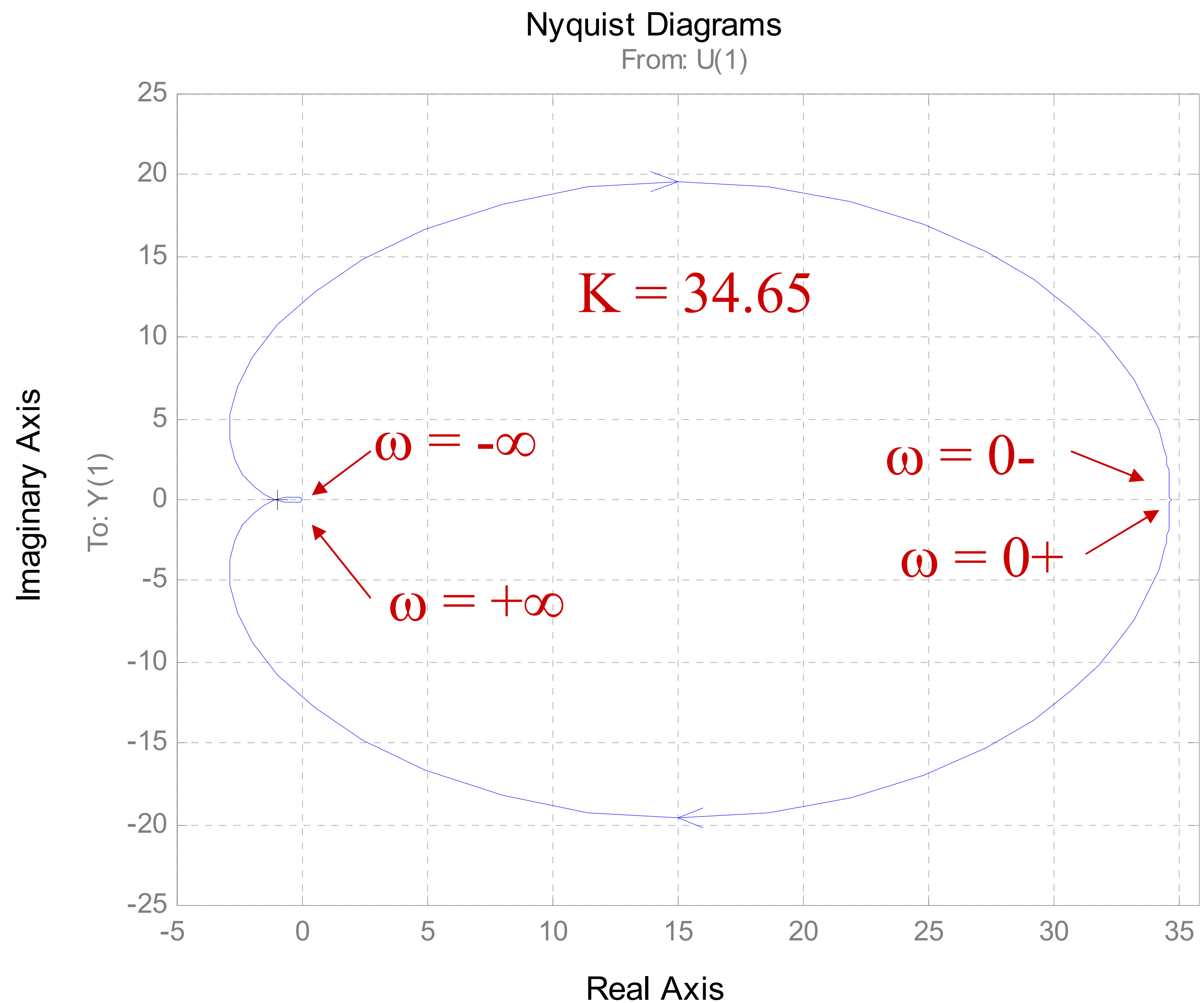
$$=\frac{K}{0.005s^3+0.155s^2+1.15s+1}$$

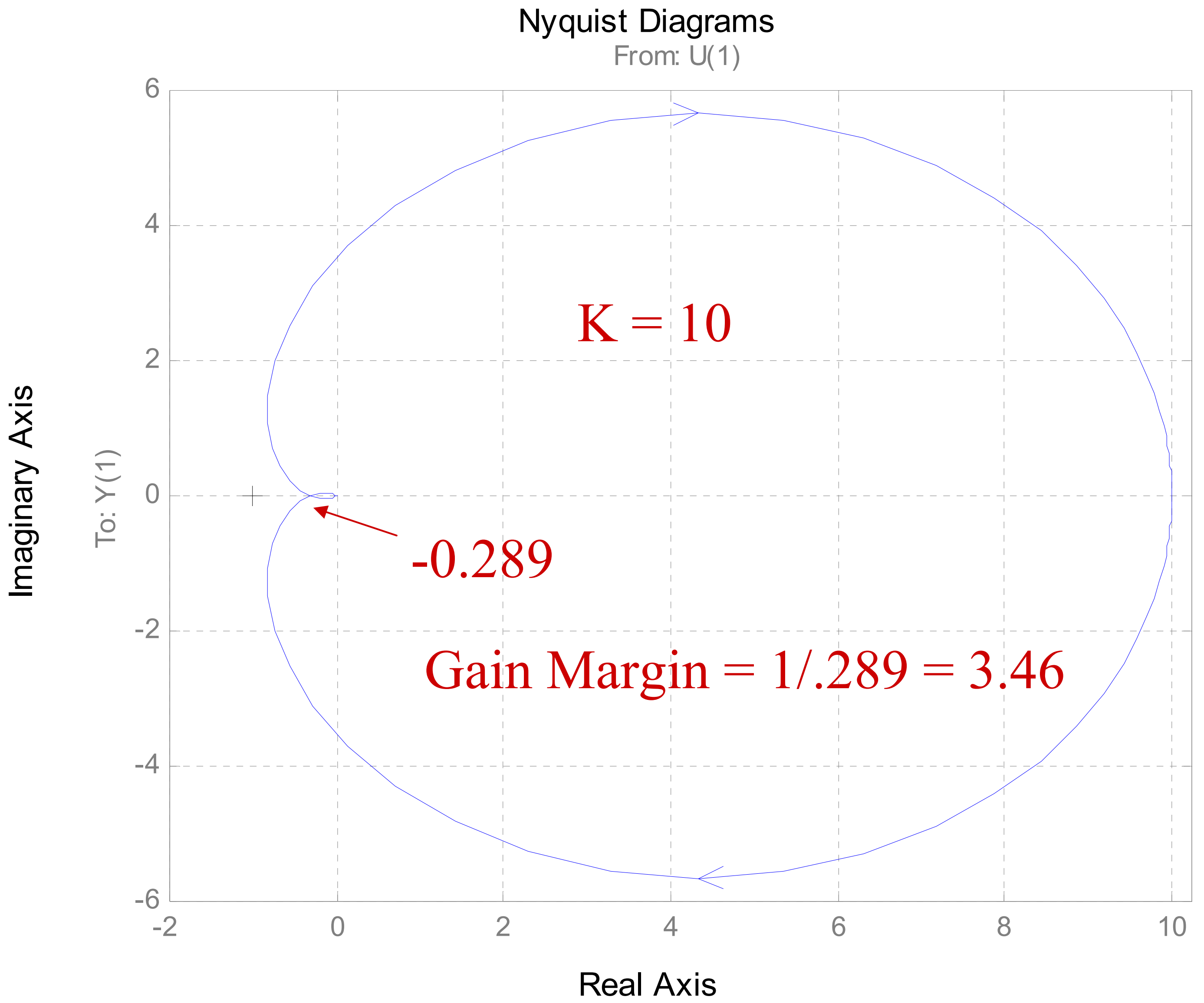
- The root-locus plot shows that when K = 34.65, the system is marginally stable. For that value of K, the closed-loop poles are at: -31 and ± 15.1658i.


Root-Locus Plot for K > 0
K = 34.65
Open-Loop Poles:
-20, -10, -1
Imag Axis
Real Axis
30
20
10
0
-10
-20
-30
-25
-20
-15
-10
-5
0
5

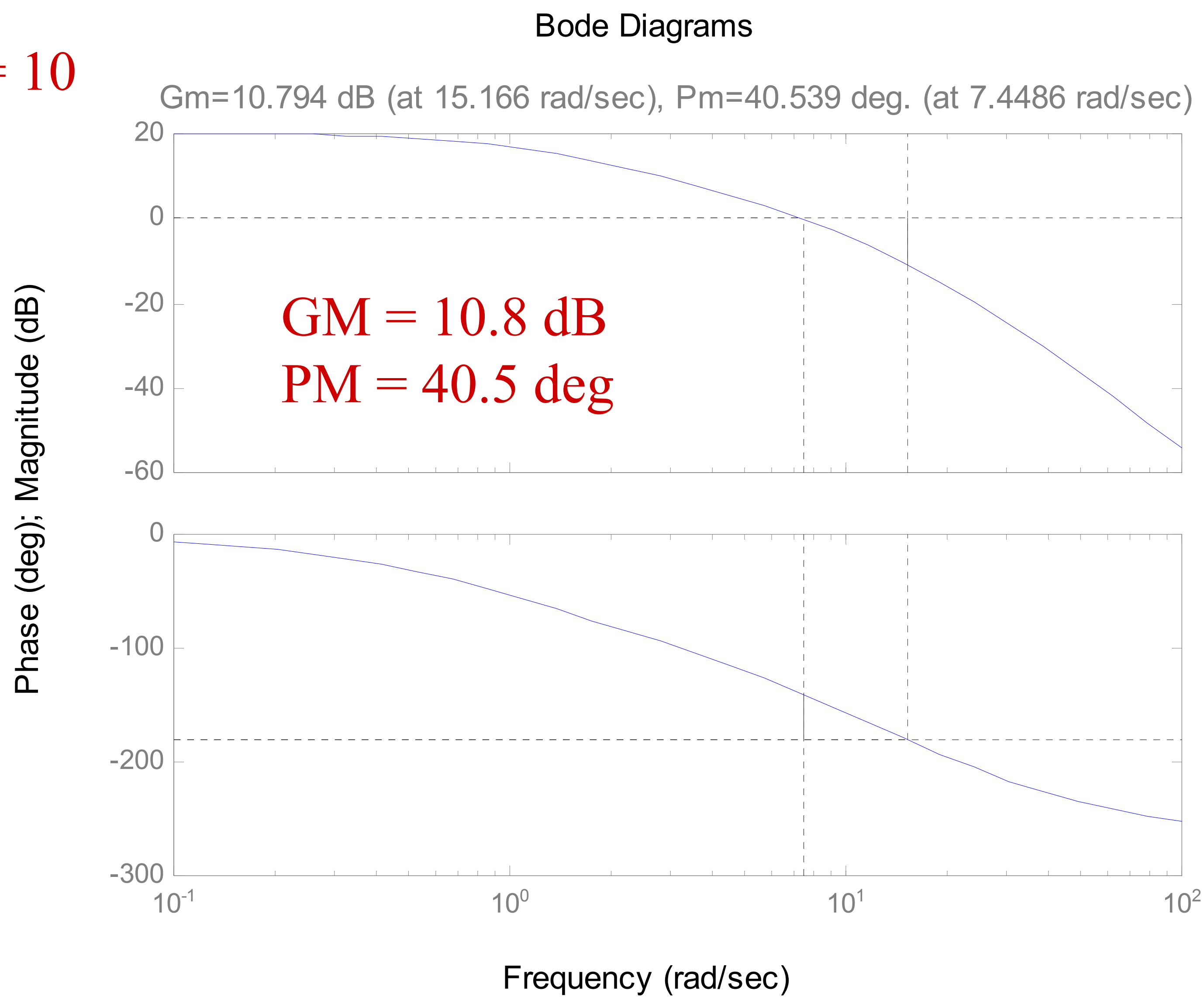
– Nyquist Stability Criterion

- A polar plot of the open-loop transfer function for the gain K = 34.65 goes through the point –1, indicating marginal stability of the closed-loop system.
- A polar plot of the open-loop transfer function for the gain K = 10 shows a gain margin = 3.46.
- The Bode plots for a gain K = 10 show a gain margin = 10.794 dB = 3.46 and a phase margin = 40.5 degrees.


Nyquist Diagrams
From: U(1)
K = 34.65
ω = -∞
ω = +∞
ω = 0-
ω = 0+
Imaginary Axis
To: Y(1)
Real Axis
25
20
15
10
5
0
-5
-10
-15
-20
-25
-5
0
5
10
15
20
25
30
35

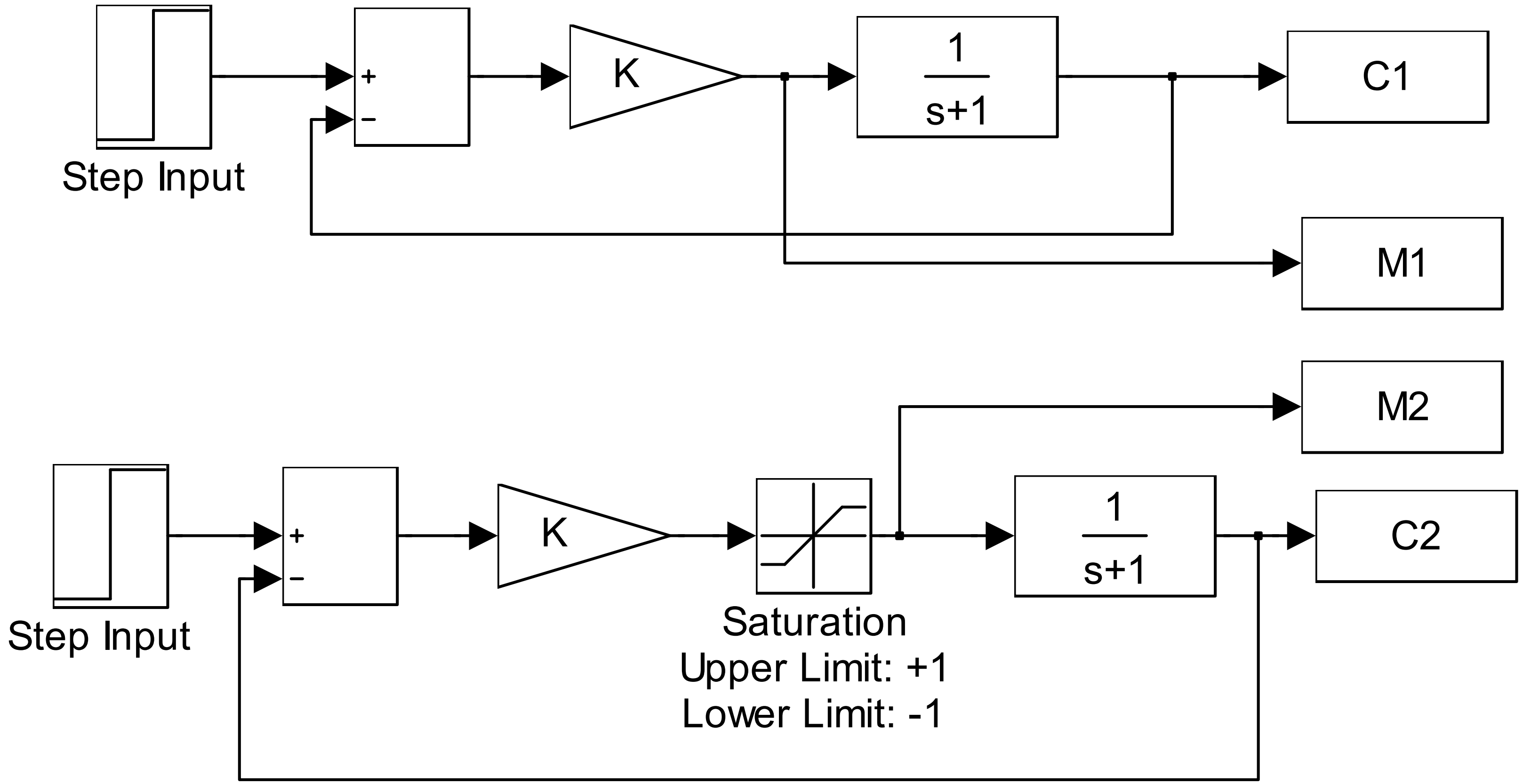

Nyquist Diagrams
From: U(1)
Imaginary Axis
To: Y(1)
Real Axis
6
4
2
0
-2
-4
-6
-2
0
2
4
6
8
10
K = 10
-0.289
Gain Margin = 1/.289 = 3.46

K = 10

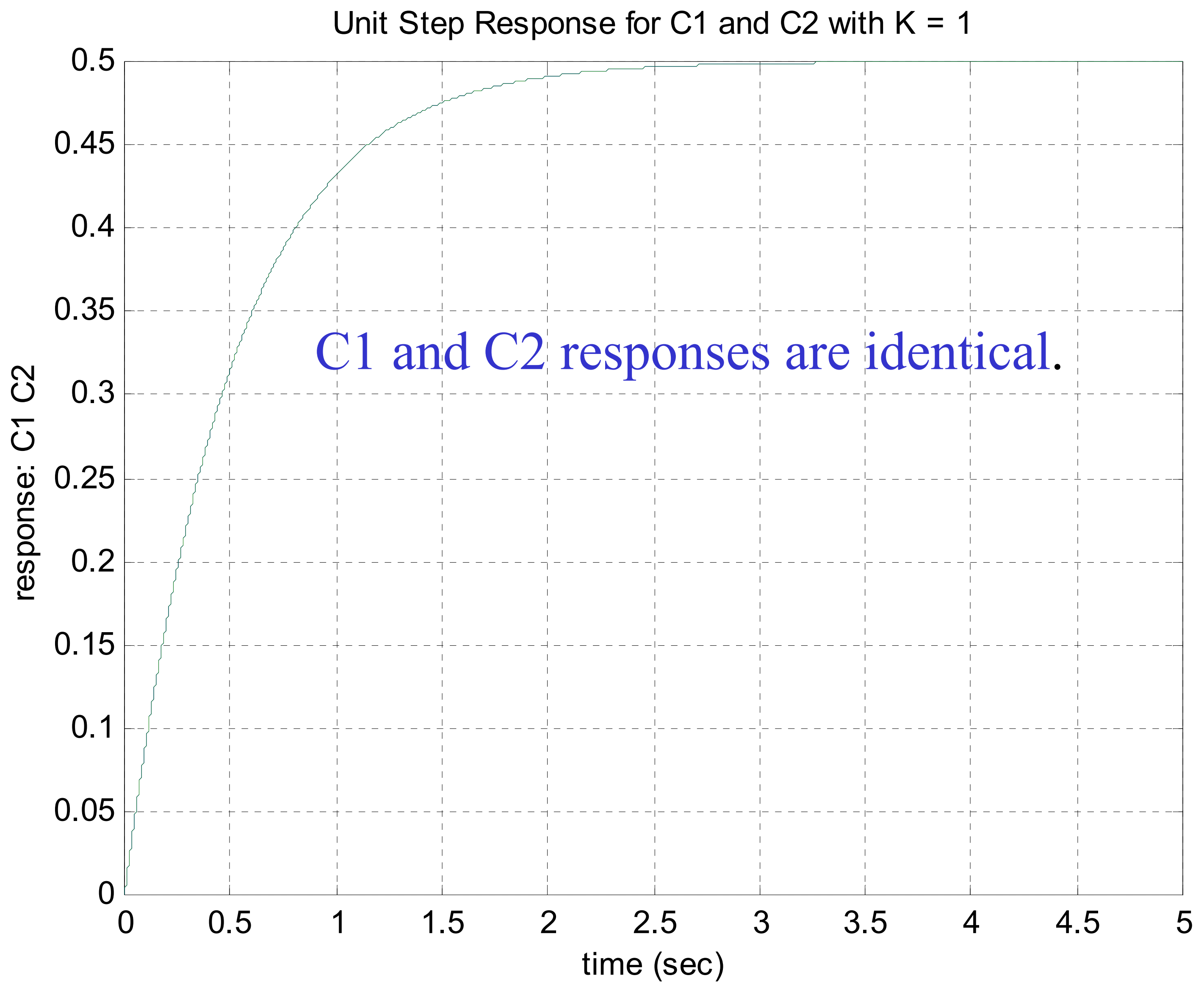


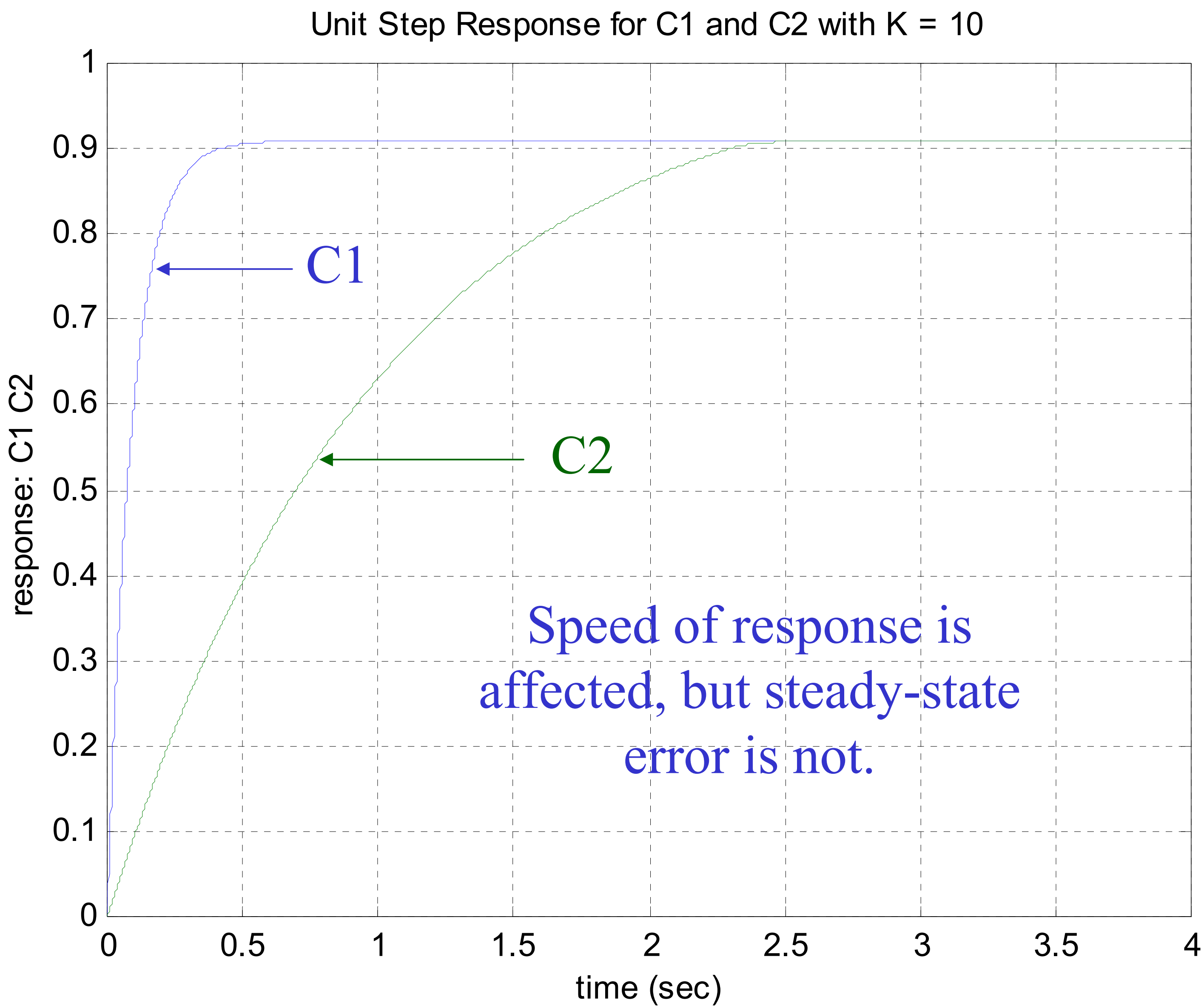
Bode Diagrams
Gm=10.794 dB (at 15.166 rad/sec), Pm=40.539 deg. (at 7.4486 rad/sec)
20
0
-20
-40
-60
0
-100
-200
-300
Phase (deg); Magnitude (dB)
10^{-1}
10^{0}
10^{1}
10^{2}
Frequency (rad/sec)
GM = 10.8 dB
PM = 40.5 deg

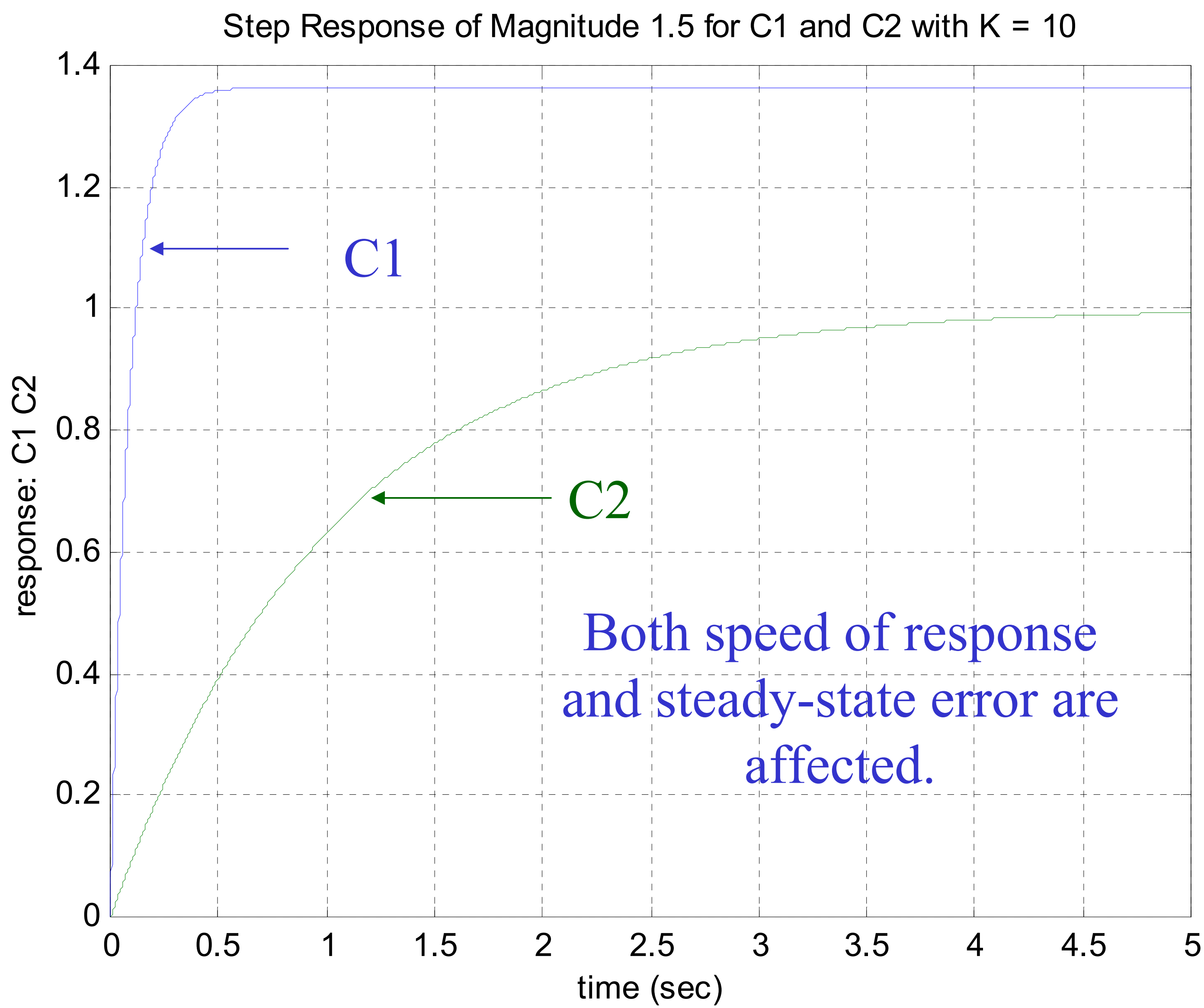
- Saturation
  - The amplifier, electropneumatic transducer, and valve positioner all exhibit saturation, limiting their output when the input becomes too large.
  - All real systems must exhibit such power limitations and one of the consequences is that the closed-loop response speed improvement will not be realized for large signals.
- Exercise:
  - Simulate the following two systems for gains of K = 1, 10 and a unit step input.
  - Examine speed of response and steady-state error.



MatLab / Simulink Diagram


Unit Step Response for C1 and C2 with K = 1
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
response: C1 C2
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
time (sec)
C1 and C2 responses are identical.


Unit Step Response for C1 and C2 with K = 10
response: C1 C2
time (sec)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.5
1
1.5
2
2.5
3
3.5
4
C1
C2
Speed of response is affected, but steady-state error is not.


Step Response of Magnitude 1.5 for C1 and C2 with K = 10
1.4
1.2
1
0.8
0.6
0.4
0.2
0
response: C1 C2
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
time (sec)
C1
C2
Both speed of response
and steady-state error are
affected.

- Let's make the system model more realistic by modeling the pneumatic valve positioner as a first-order system:

$$\frac{x_v}{p}(D) = \frac{K_x}{\tau_{vp} D + 1}$$

  – The open-loop system is now second order. The closed-loop system differential equation is now:

$$\left(\frac{D^2}{\omega_n^2} + \frac{2\zeta D}{\omega_n} + 1 \right) h_C = \frac{K}{K+1} h_V + \frac{\tau_{vp} D + 1}{\rho g (K+1)} P_U - \frac{R_f (\tau_{vp} D + 1)}{\rho g (K+1)} Q_U$$

$$\omega_n = \sqrt{\frac{K+1}{\tau_p \tau_{vp}}} \qquad \zeta = \frac{\tau_p + \tau_{vp}}{2\sqrt{\tau_p \tau_{vp} (K+1)}} \qquad K = \frac{1}{\rho g}\left(K_h K_a K_p K_x K_v R_f \right)$$

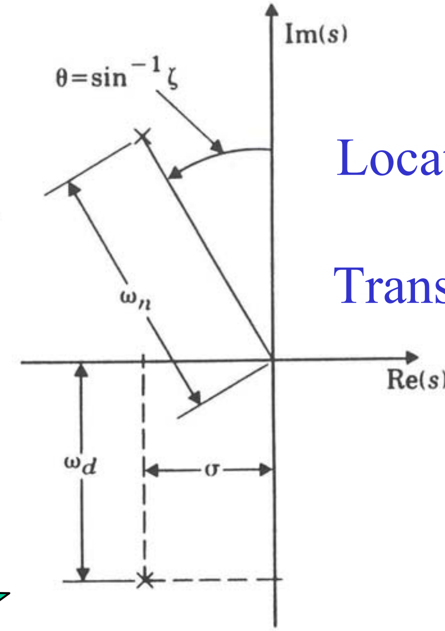
- To get fast response (large ω_n) for given lags τ_p and τ_{vp}, we must increase loop gain K. How does K affect ζ ?
- If τ_p = 60 sec and τ_{vp} = 1.0 sec and we desire ζ = 0.6, what is K? What is ω_n ? Is absolute instability possible with this model? What does the Nyquist plot show as K is increased? What does the root-locus plot show as K is increased?

- Consider Gain Distribution
  - How does gain distribution affect stability and dynamic response of the closed-loop system?
  - Are steady-state errors for disturbances sensitive to gain distribution?
  - Should one optimize the distribution of gain so as to minimize steady-state errors?

$$G(s) = \frac{K\omega_n^2}{s^2 + 2\varsigma\omega_n s + \omega_n^2}$$

$$s_{1,2} = -\varsigma\omega_n \pm i\omega_n\sqrt{1-\varsigma^2}$$

$$s_{1,2} = -\sigma \pm i\omega_d$$



Location of Poles
Of
Transfer Function

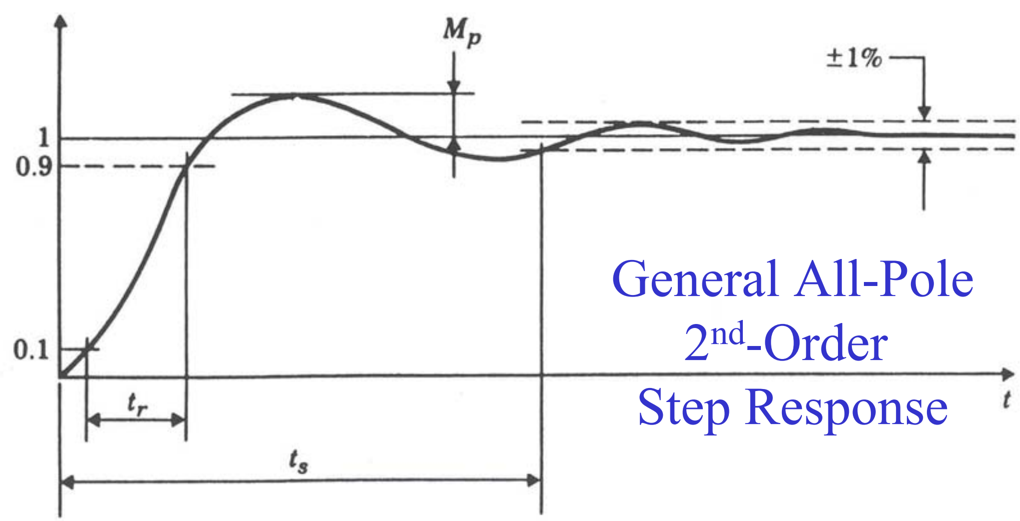
$$y(t) = 1 - e^{-\sigma t}\left(\cos\omega_d t + \frac{\sigma}{\omega_d}\sin\omega_d t\right)$$

$$t_r \approx \frac{1.8}{\omega_n} \quad \text{rise time}$$

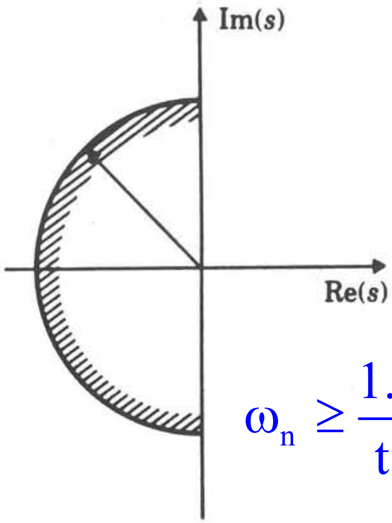
$$t_s \approx \frac{4.6}{\varsigma\omega_n} \quad \text{settling time}$$

$$M_p = e^{\frac{-\pi\varsigma}{\sqrt{1-\varsigma^2}}} \quad (0 \le \zeta < 1) \quad \text{overshoot}$$

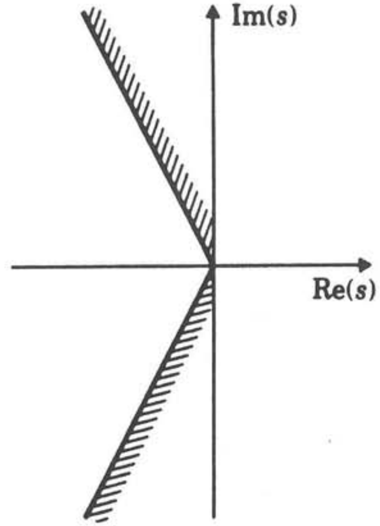
$$\approx \left(1 - \frac{\zeta}{0.6}\right) \quad (0 \le \zeta \le 0.6)$$

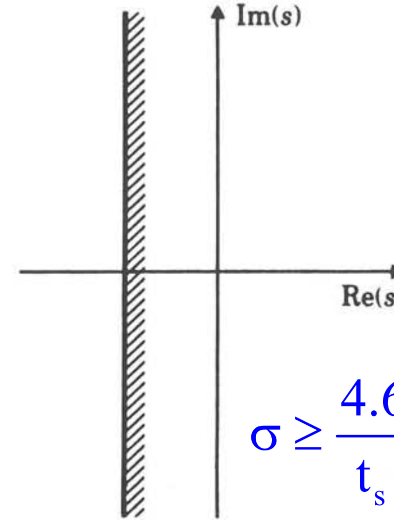


General All-Pole
2^{nd}-Order
Step Response



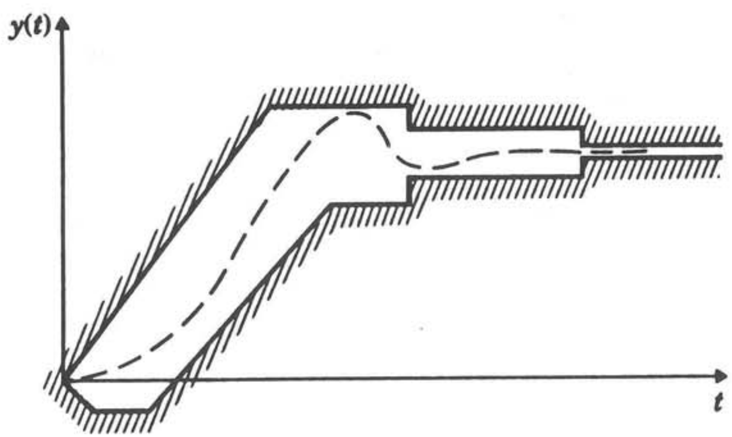
$$\omega_n \geq \frac{1.8}{t_r}$$

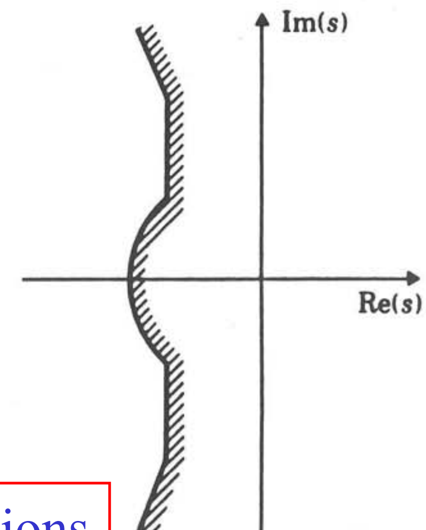




$$\sigma \geq \frac{4.6}{t_s}$$

$$\zeta \geq 0.6\left(1 - M_p\right) \qquad 0 \leq \zeta \leq 0.6$$





Time-Response Specifications vs. Pole-Location Specifications

- <u>Integral Control</u>
  - Consider integral control of this liquid-level process. Replace the amplifier block K_a by K_I/s.
  - The closed-loop system differential equation is:

$$\left[\left(h_V - h_C\right)\frac{K_h K_I K_p K_x K_v}{s} + \frac{1}{R_f}P_U - Q_U\right]\frac{\frac{R_f}{\rho g}}{\tau_p s + 1} = h_C$$

$$\left(\tau_p D^2 + D + K\right)h_C = K h_V + \frac{1}{\rho g}DP_U - \frac{R_f}{\rho g}DQ_U$$

$$\left(\tau_p D^2 + D + K\right)h_E = -\left(\tau_p D^2 + D\right)h_V + \frac{1}{\rho g}DP_U - \frac{R_f}{\rho g}DQ_U$$

$$K = \frac{K_h K_I K_p K_x K_v R_f}{\rho g}$$ loop gain

$$h_E = h_V - h_C$$ system error

– Note the following:

– Step changes (constant values) of h_V, P_U, and/or Q_U give zero steady-state errors.

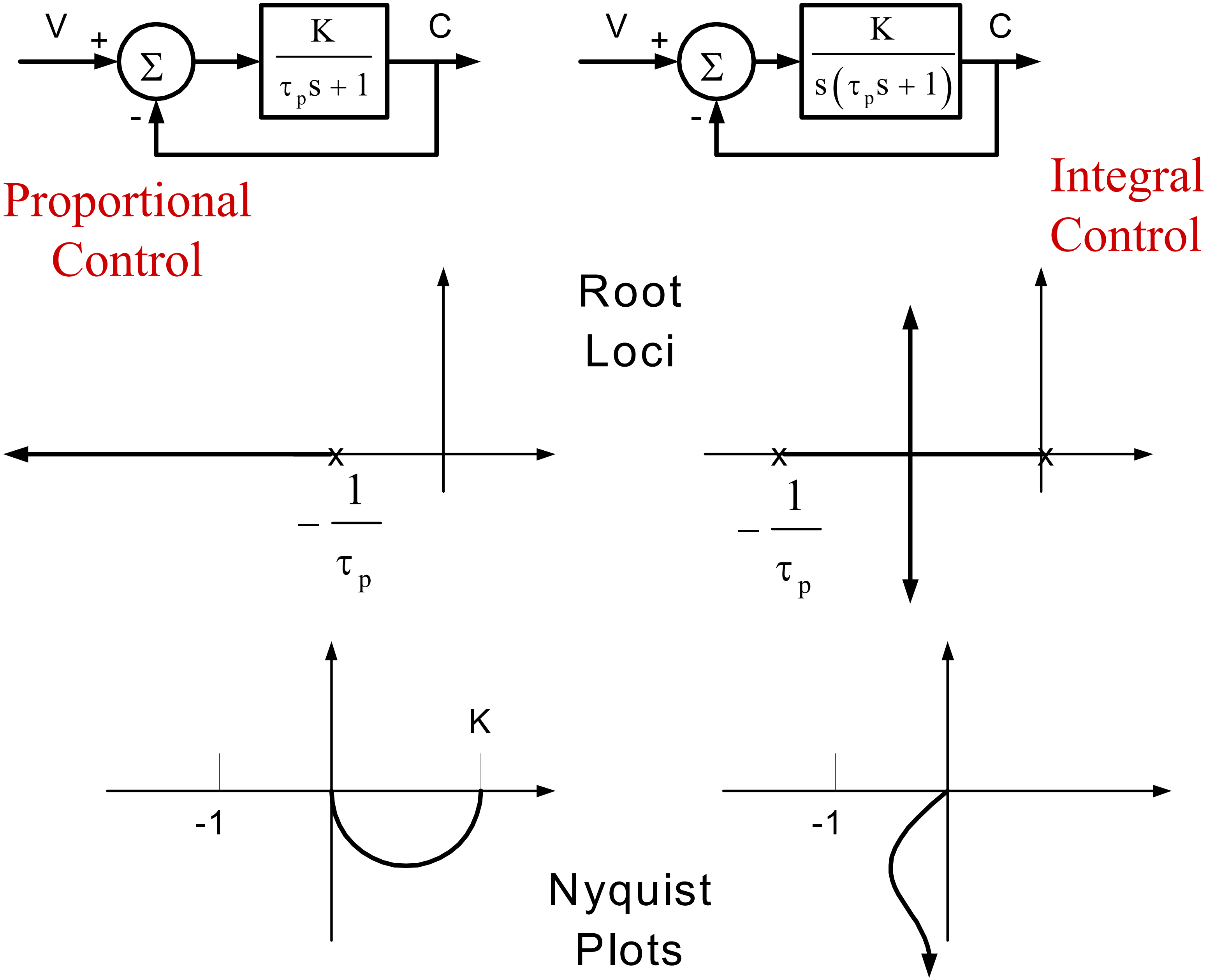
– For ramp inputs, we now have constant, nonzero steady-state errors whose magnitudes can be reduced by increasing K.

– The characteristic equation is second order, so define:

$$\omega_n = \sqrt{\frac{K}{\tau_p}} \qquad \zeta = \frac{1}{2\sqrt{K\tau_p}}$$

– If we take τ_p as unavailable for change, we see that an increase in K to gain response speed or decrease ramp steady-state errors will be limited by loss of relative stability (low ζ).

– If we design for a desired ζ, the needed K is easily found and ω_n is then fixed.
– Absolute instability is not predicted; the model is too simple.
– See the comparison between proportional control and integral control: root locus plots and Nyquist plots. Note the destabilizing effects of integral control.



V
+
-
Σ
$\frac{K}{\tau_p s + 1}$
C
Proportional Control
V
+
-
Σ
$\frac{K}{s(\tau_p s + 1)}$
C
Integral Control
Root Loci
$-\frac{1}{\tau_p}$
$-\frac{1}{\tau_p}$
K
-1
-1
Nyquist Plots

- What if we change the controlled process by closing off the pipe on the left side of the tank. This not only deletes P_U as a disturbance but also causes a significant change in process dynamics.
- Take $R_f = \infty$. The from Conservation of Mass we have:

$$h_C = \frac{1}{A_T s}\left(Q_M - Q_U\right)$$

- The original tank process had self-regulation.
  - If one changes Q_M and/or Q_U the tank will itself in time find a new equilibrium level since the flow through R_f varies with level.

- With R_f not present, the tiniest difference between Q_M and Q_U will cause the tank to completely drain or overflow since it is now an integrator and has lost its self-regulation.
- Even with proportional control, the integrating effect in the process gives zero steady-state error for step commands, but not for disturbances.
- If we substitute integral control to eliminate the Q_U error, the system becomes absolutely unstable.

Optical Encoders

- Any transducer that generates a coded reading of a measurement can be termed an *encoder*.
- *Shaft Encoders* are digital transducers that are used for measuring angular displacements and velocities.
- Relative advantages of digital transducers over their analog counterparts:
  - High resolution (depending on the word size of the encoder output and the number of pulses per revolution of the encoder)
  - High accuracy (particularly due to noise immunity of digital signals and superior construction)

- Relative ease of adaptation in digital control systems (because transducer output is digital) with associated reduction in system cost and improvement of system reliability

- Shaft Encoders can be classified into two categories depending on the nature and method of interpretation of the output:
  - Incremental Encoders
  - Absolute Encoders
- <u>Incremental Encoders</u>
  - Output is a pulse signal that is generated when the transducer disk rotates as a result of the motion that is being measured.

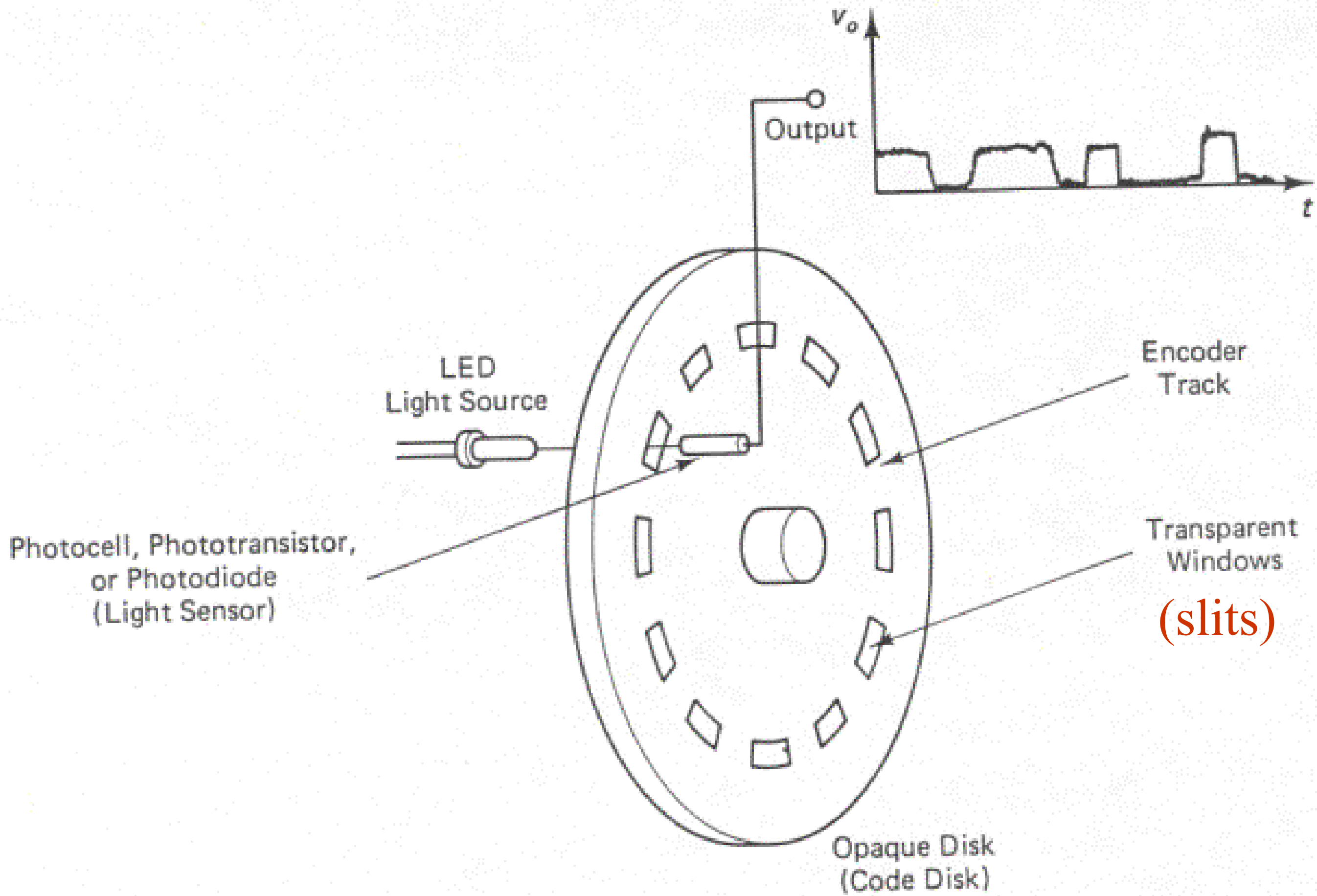
- By counting pulses or by timing the pulse width using a clock signal, both angular displacement and angular velocity can be determined.
- Displacement, however, is obtained with respect to some reference point on the disk, as indicated by a reference pulse (index pulse) generated at that location on the disk. The index pulse count determines the number of full revolutions.

- Absolute Encoders
  - An absolute encoder has many pulse tracks on its transducer disk.
  - When the disk of an absolute encoder rotates, several pulse trains – equal in number to the tracks on the disk – are generated simultaneously.

- At a given instant, the magnitude of each pulse signal will have one of two signal levels (i.e., a binary state) as determined by a level detector. This signal level corresponds to a binary digit (0 or 1). Hence, the set of pulse trains gives an encoded binary number at any instant.
- The pulse windows on the tracks can be organized into some pattern (code) so that each of these binary numbers corresponds to the angular position of the encoder disk at the time when the particular binary number is detected.
- Pulse voltage can be made compatible with some form of digital logic (e.g., TTL)
- Direct digital readout of an angular position is possible.

- Absolute encoders are commonly used to measure fractions of a revolution. However, complete revolutions can be measured using an additional track that generates an index pulse, as in the case of an incremental encoder.

- Signal Generation can be accomplished using any one of four techniques:
  - Optical (photosensor) method
  - Sliding contact (electrical conducting) method
  - Magnetic saturation (reluctance) method
  - Proximity sensor method
- Method of signal interpretation and processing is the same for all four types of signal generation.



Schematic Representation of an Optical Encoder
One Track and One Pick-Off Sensor Shown

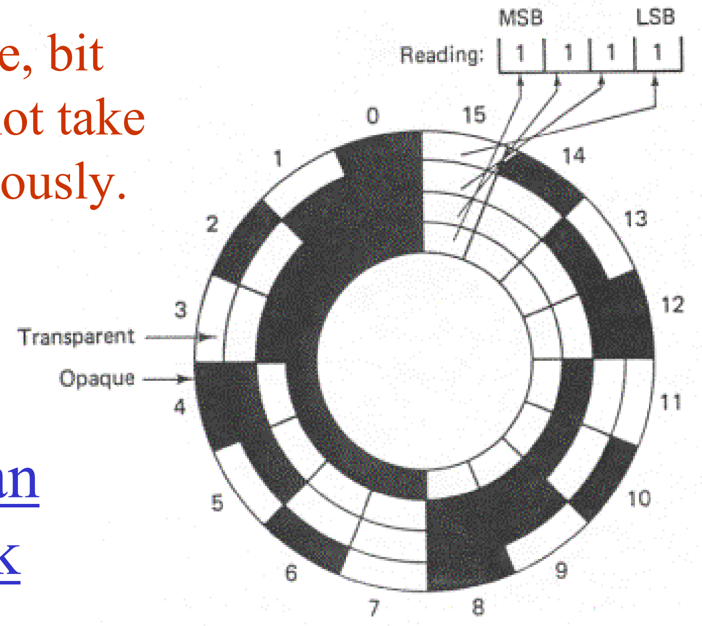
In Binary Code, bit switching may not take place simultaneously.

Schematic Diagram of an Absolute Encoder Disk Pattern

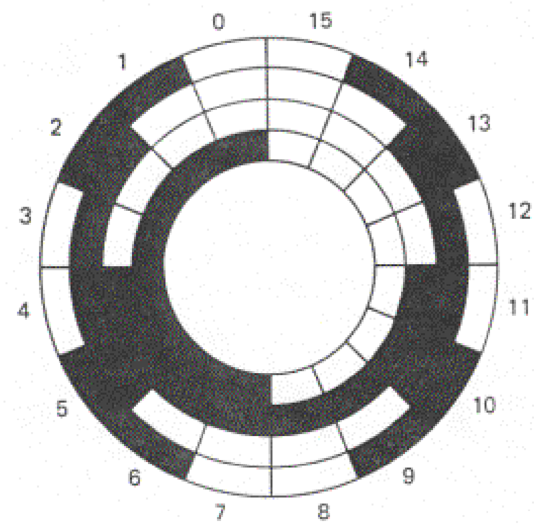
(a) Binary code

(b) Gray code

Ambiguities in bit switching can be avoided by using gray code. However, additional logic is needed to covert the gray-coded number to a corresponding binary number.

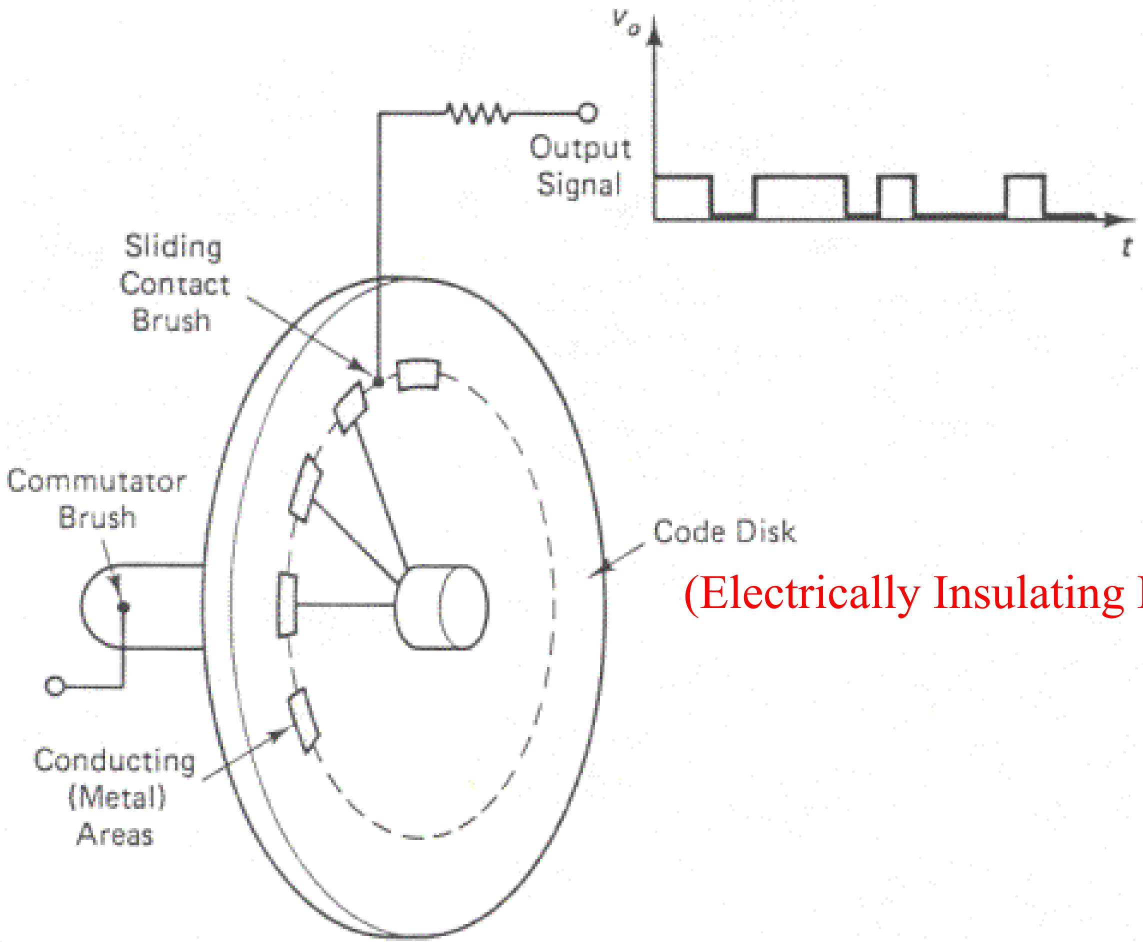


(a)



(b)

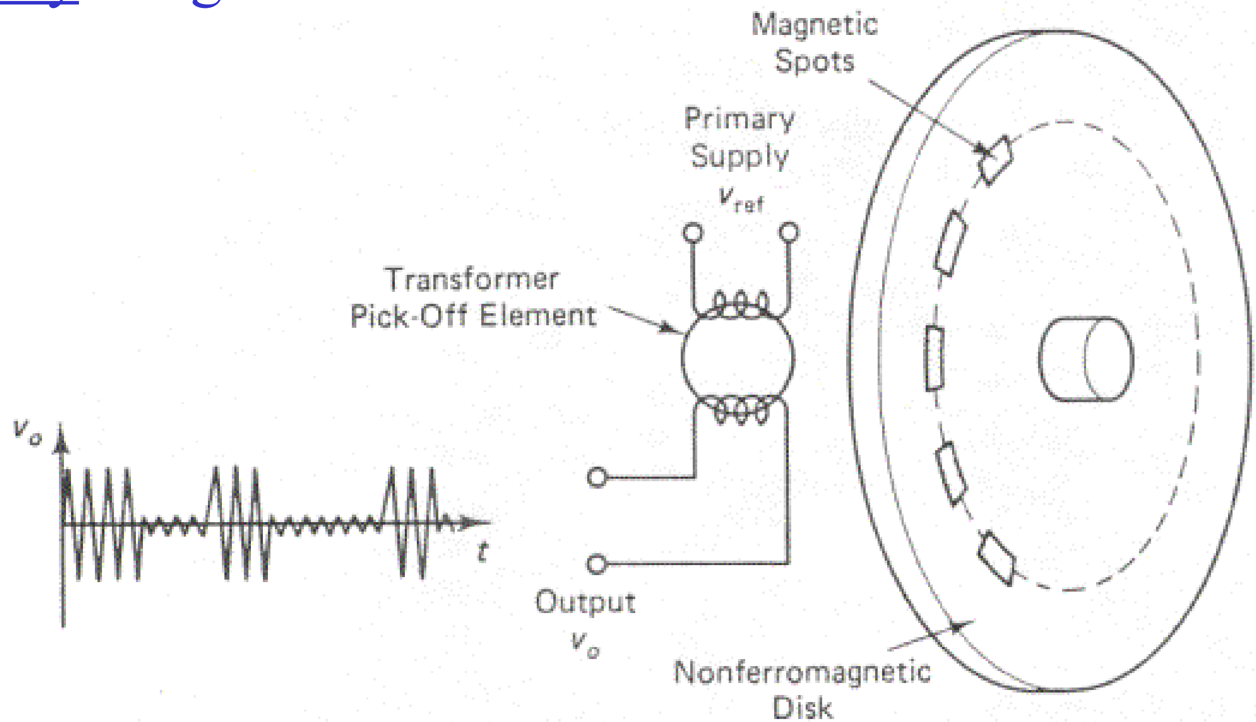
Absolute Encoders must be powered and monitored <u>only</u> when a reading is taken. Also, if a reading is missed, it will not affect the next reading.



(Electrically Insulating Material)

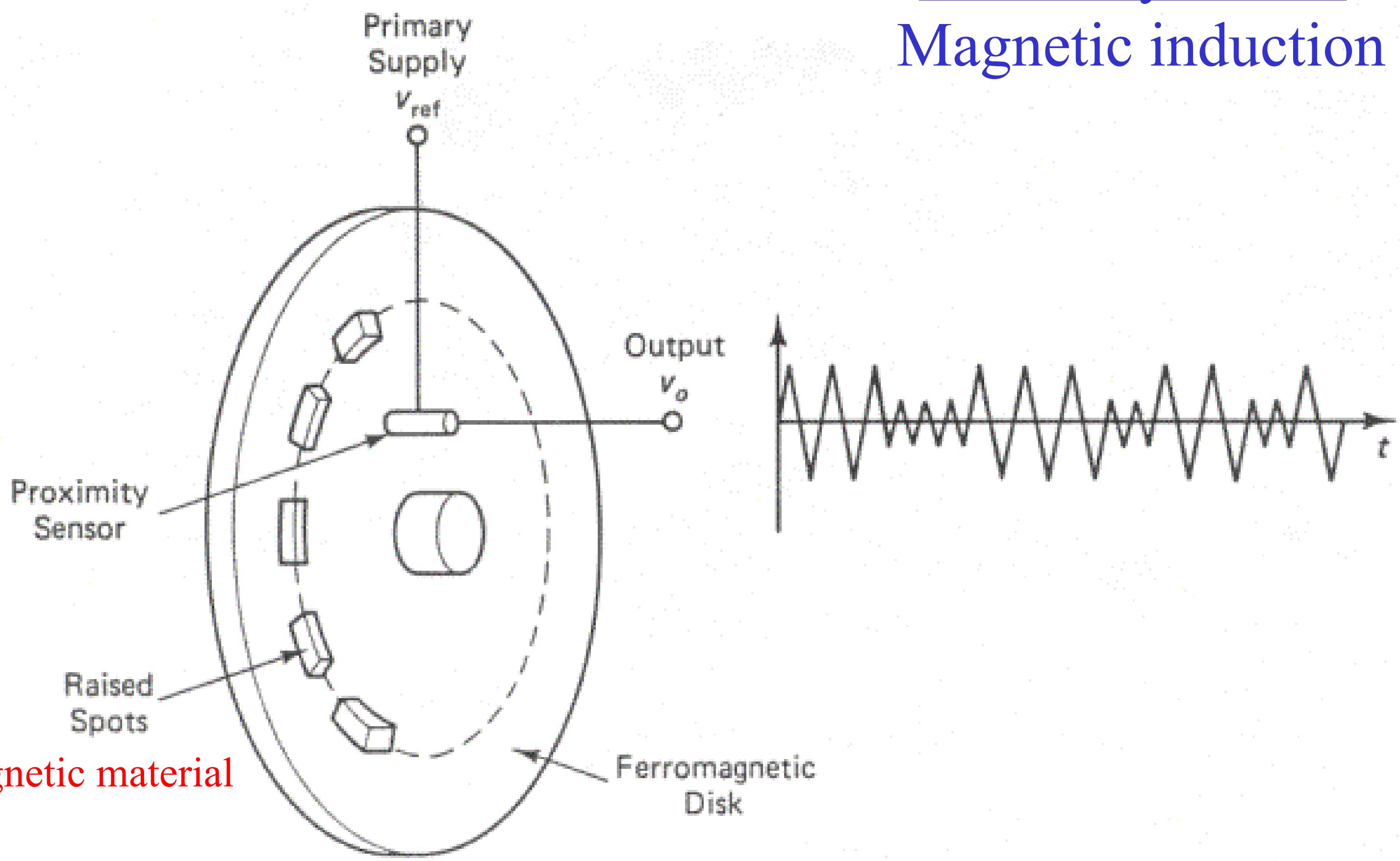
Schematic Representation of a Sliding Contact Encoder

Pulse peak: nonmagnetic are
Pulse valley: magnetic area



Schematic Representation of a Magnetic Encoder

Proximity sensor: Magnetic induction



ferromagnetic material

Schematic Representation of a Proximity Probe Encoder

- Elements of the Optical Encoder
  - The optical encoder uses an opaque disk (code disk) that has one or more circular tracks, with some arrangement of identical transparent windows (slits) in each track.
  - A parallel beam of light (e.g., from a set of light-emitting diodes) is projected to all tracks from one side of the disk.
  - The transmitted light is picked off using a bank of photosensors on the other side of the disk that typically has one sensor for each track.
  - The light sensor could be a silicon photodiode, a phototransistor, or a photovoltaic cell.

– Since the light from the source is interrupted by the opaque areas of the track, the output signal from the probe is a series of voltage pulses. This signal can be interpreted to obtain the angular position and angular velocity of the disk.

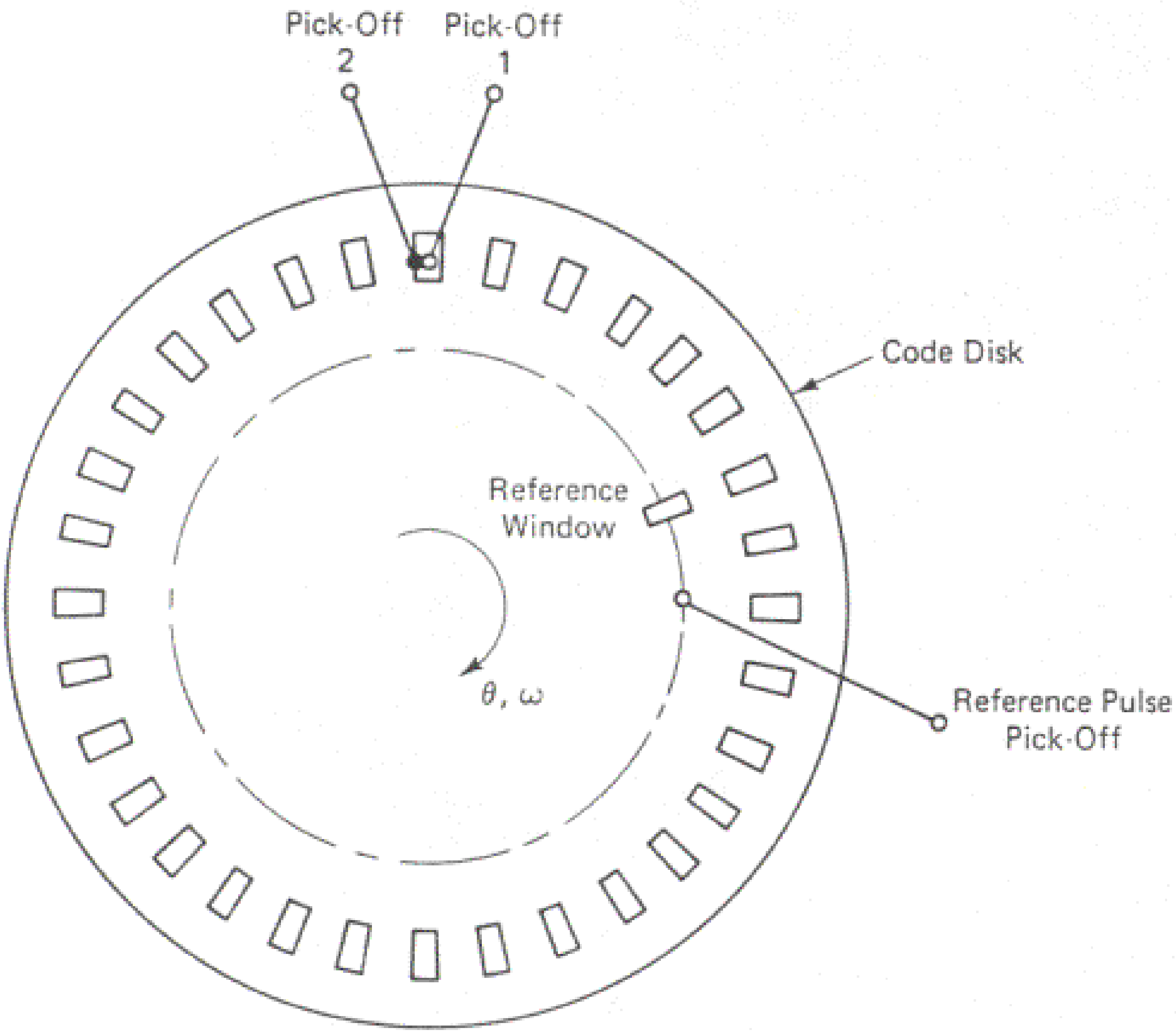
– Note that an incremental encoder disk requires only one primary track that has equally spaced and identical window (pick-off) areas. The window area is equal to the area of the inter-window gap. Usually, a reference track that has just one window is also present in order to generate a pulse (known as the index pulse) to initiate pulse counting for angular position measurement and to detect complete revolutions.

– In contrast, absolute encoder disks have several rows of tracks, equal in number to the bit size of the output data word. Furthermore, the track windows are not equally spaced but are arranged in a specific pattern on each track so as to obtain a binary code (or gray code) for the output data from the transducer.
– It follows that absolute encoders need as least as many signal pick-off sensors as there are tracks, whereas incremental encoders need one pick-off sensor to detect the magnitude of rotation and an additional sensor at a quarter-pitch separation (pitch = center-to-center distance between adjacent windows) to identify the direction of rotation, i.e., the *offset sensor configuration*.

- Some designs of incremental encoders have two identical tracks, one a quarter-pitch offset from the other, and the two pick-off sensors are placed radially without any circumferential offset, i.e., the *offset track configuration*.
- A pick-off sensor for a reference pulse is also used.

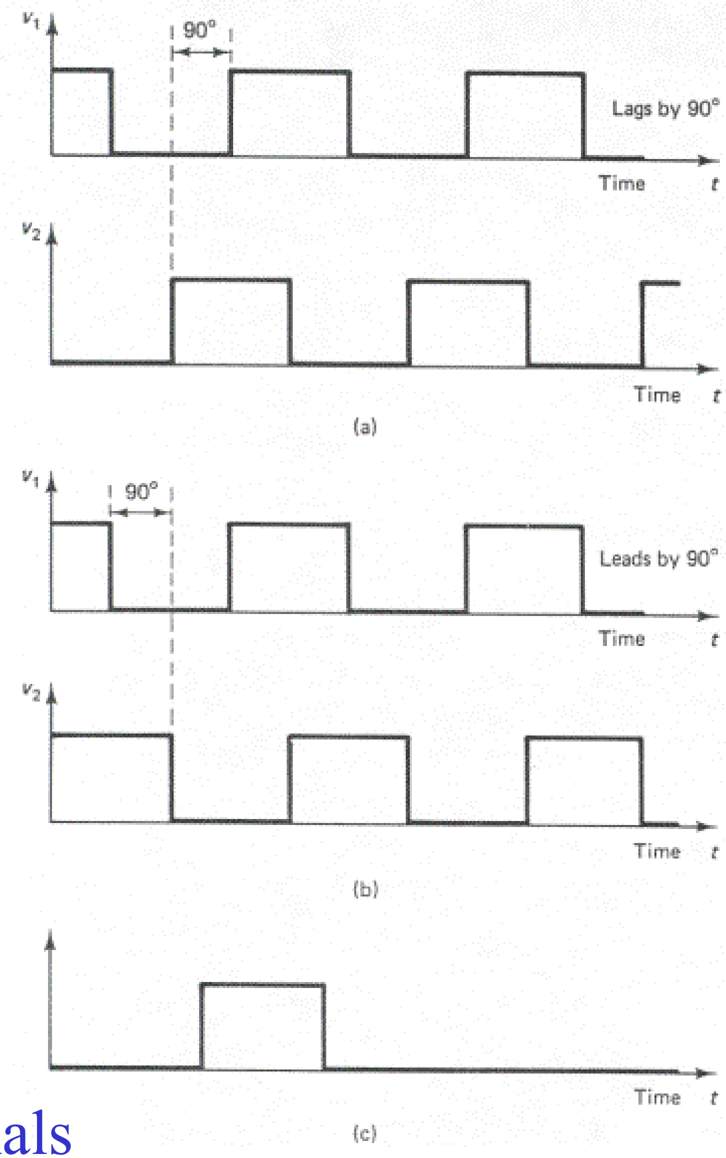
- *Signal interpretation* depends on whether the particular optical encoder is an incremental device or an absolute device.
  - We will focus on the incremental optical encoder.
  - The output signals from either the offset sensor configuration or the offset track configuration are the same.

- Note that the pulse width and pulse-to-pulse period (encoder cycle) are constant in each sensor output when the disk rotates at constant angular velocity. When the disk accelerates, the pulse width decreases continuously; when the disk decelerates, the pulse width increases continuously.
- The quarter-pitch offset in sensor location or track position is used to determine the direction of rotation of the disk. It is obtained by determining the phase difference of the two output signals, using phase-detection circuitry. One method for determining the phase difference is to time the pulses using a high-frequency clock signal.



Incremental Optical Encoder Disk Offset-Sensor Configuration

Clockwise (CW) rotation:
V_1 lags V_2 by a quarter of a cycle
(i.e., a phase lag of 90°)
Counterclockwise (CCW) rotation:
V_1 leads V_2 by a quarter of a cycle



Incremental Encoder Pulse Signals

(a) CW rotation (b) CCW rotation (c) reference

- <u>Displacement Computation</u>
  - Maximum count possible: M pulses
  - Range of the encoder: $\pm\,\theta_{max}$
  - If the data size is *r* bits, allowing for a sign bit, $M = 2^{r-1}$, where zero count is also included.
  - If zero count is not included, $M = 2^{r-1} - 1$
  - If θ_{max} is 2π and θ_{min} is zero, then θ_{max} and θ_{min} will correspond to the same position of the code disk. To avoid this ambiguity, we use
  - The conventional definition for digital resolution is:

$$\theta = \frac{n \text{ pulses}}{M}\theta_{max}$$

$$\theta_{min} = \frac{\theta_{max}}{2^{r-1}}$$

$$\frac{\left(\theta_{max} - \theta_{min}\right)}{\left(2^{r-1} - 1\right)}$$

- Two methods are available for determining velocities using an incremental encoder:
  - *pulse-counting method*
  - *pulse-timing method*
- <u>Pulse-Counting Method</u>
  - The pulse count over the sampling period of the digital processor is measured and is used to calculate the angular velocity. For a given sampling period, there is a lower speed limit below which this method is not very accurate.

- To compute the angular velocity ω, suppose that the count during a sample period *T* is *n* pulses. Hence, the average time for one pulse is *T/n*. If there are *N* windows on the disk, the average time for one revolution is *NT/n*. Hence $\omega\ (rad/s) = 2\pi n/NT$.

- Pulse-Timing Method
  - The time for one encoder cycle is measured using a high-frequency clock signal. This method is particularly suitable for measuring low speeds accurately.
  - Suppose that the clock frequency is *f* Hz. If *m* cycles of the clock signal are counted during an encoder period (interval between two adjacent windows), the time for that encoder cycle (i.e., the time to rotate through one encoder pitch) is given by *m/f*.

– With a total of N windows on the track, the average time for one revolution of the disk is Nm/f. Hence $\omega = 2\pi f/Nm$.

- <u>Resolution of an Encoder</u>
  – The resolution of an encoder represents the smallest change in measurement that can be measured realistically. Since an encoder can be used to measure both displacement and velocity, a resolution can be identified for each case.
  – <u>*Displacement Resolution*</u>
    - *Displacement resolution* is governed by the number of windows N in the code disk and the digital size (number of bits) r of the buffer (counter output). The *physical resolution* is determined by N. If only one pulse signal is used (i.e., no direction sensing), and if the rising edges of the pulses are detected (i.e., full cycles of the encoder are counted), the physical resolution is given by $(360/N)°$.

- But if both pulse signals (quadrature signals) are available and the capability to detect rising and falling edges of a pulse is also present, four counts can be made per encoder cycle, thereby improving the resolution by a factor of four.
- Hence, the physical resolution is given by $(360/4N)^\circ$. When the two signals V_1 and V_2 are added, the resulting signal has a transition at every quarter of the encoder cycle. By detecting each transition (through edge detection or level detection), four pulses can be counted within every main cycle.
- Assuming that the maximum angle measured is 360° (or ±180°), the *digital resolution* is given by: $$\Delta\theta_d = \frac{180^\circ}{2^{r-1}} = \frac{360^\circ}{2^r}$$
- A digital word containing *r* bits can represent 2^r different values (unsigned). Note that 0° and 360° represent the same position of the code disk. An ambiguity does not arise if we take the minimum value of θ to be $360^\circ/2^r$, not zero.

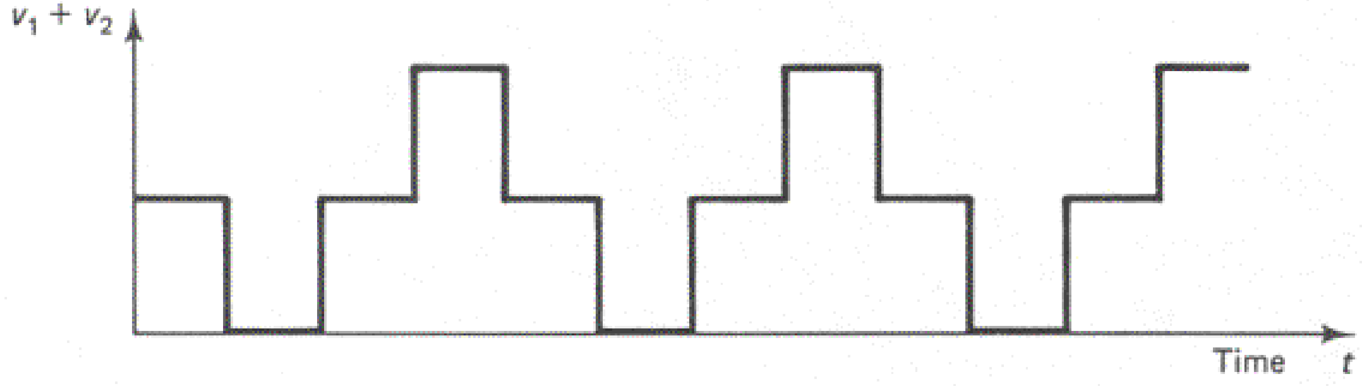
- *The larger of the two resolutions governs the displacement resolution of the encoder*:

$$\Delta\theta_p = \frac{360^\circ}{4N} \qquad \Delta\theta_d = \frac{360^\circ}{2^r}$$

- The physical resolution of an encoder can be improved by using step-up gearing so that one rotation of the moving object that is being monitored corresponds to several rotations of the code disk of the encoder. This improvement is directly proportional to the gear ratio.

– In summary, the *displacement resolution* of the incremental encoder depends on the following factors:

  - Number of windows on the code track
  - Gear ratio
  - Word size of the measurement buffer



Quadrature Signal Addition to Improve Physical Resolution

– *Velocity Resolution*

- An incremental encoder is also a velocity-measuring device. The *velocity resolution* of an incremental encoder depends on the method that is employed to determine velocity. Since the pulse-counting method and the pulse-timing method are both based on counting, the resolution corresponds to the change in angular velocity that results from changing (incrementing or decrementing) the count by one. If the pulse-counting method is employed, a unity change in the count n corresponds to a speed change of

$$\Delta\omega_c = \frac{2\pi}{NT}$$

- N is the number of windows in the code track and T is the sampling period.
- This velocity resolution is independent of the angular velocity itself. The resolution improves, however, with the number of windows and the sampling period.

- But under transient conditions, the accuracy of the velocity reading decreases with increasing T (the sampling frequency has to be at least double the highest frequency of interest in the velocity signal). Hence, the sampling period should not be increased indiscriminately.
- If the pulse-timing method is employed, the velocity resolution is given by:

$$\Delta\omega_t = \frac{2\pi f}{Nm} - \frac{2\pi f}{N(m+1)} = \frac{2\pi f}{Nm(m+1)}$$

- *f* is the clock frequency. For large *m*, (*m*+*1*) can be approximated by *m*. Then

$$\Delta\omega_t = \frac{N\omega^2}{2\pi f}$$

- Note that in this case, the resolution degrades quadratically with speed. For a given speed, the resolution degrades with increasing *N*. The resolution can be improved, however, by increasing the clock frequency. Gearing up has a detrimental effect on the speed resolution in the pulse-timing method, but it has a favorable effect in the pulse-counting method.

- In summary, the *speed resolution* of an incremental encoder depends on the following factors:
  - number of windows N
  - sampling period T
  - clock frequency f
  - speed ω
  - gear ratio
- Errors in shaft encoder readings can come from several factors:
  - Quantization error (due to digital word size limitations)
  - Assembly error (eccentricity, etc.)
  - Coupling error (gear backlash, belt slippage, loose fit, etc.)

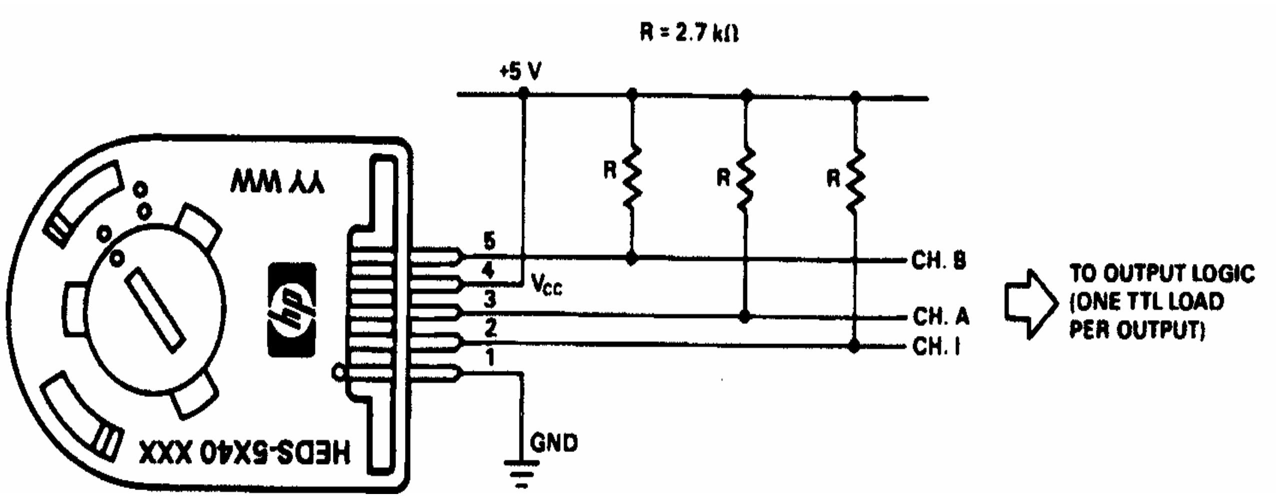
- Structural limitations (disk deformation and shaft deformation due to loading)
- Manufacturing tolerances (errors from inaccurately imprinted code patterns, inexact positioning of the pick-off sensors, limitations and irregularities in signal generation and sensing components, etc.)
- Ambient effects (vibration, temperature, light noise, humidity, dirt, smoke, etc.)

- These factors can result in erroneous displacement and velocity readings and inexact direction detection.
- One form of error in an encoder reading is the hysteresis.

- For a given position of the moving object, if the encoder reading depends on the direction of motion, the measurement has a hysteresis error.
- In that case, if the object rotates from position A to position B and back to position A, for example, the initial and final readings of the encoder will not match.
- The causes of hysteresis include backlash in gear couplings, loose fits, mechanical deformation in the code disk and shaft, delays in electronic circuitry (electrical time constants), and noisy pulse signals that make the detection of pulses (by level detection or edge detection) less accurate.

Properties of the HP HEDS-5505-A14 Incremental Optical Encoder

Specification	Values
Supply Voltage	5 Volts
Supply Current	40 mA
Resolution	500 pulses per revolution
Resolution with quadrature	2000 pulses per revolution
Number of Channels	2
Index	none
Maximum Velocity	30000 RPM
Maximum Acceleration	250000 rad/sec^2



- <u>Problem # 1</u>
  - Explain how resolution of a shaft encoder could be improved by pulse interpolation.
  - Suppose that a pulse generated from an incremental encoder can be approximated by:

$$v = v_o \left(1 + \sin \frac{2\pi\theta}{\Delta\theta} \right)$$

  - Here θ denotes the angular position of the encoder window with respect to the photosensor position.

- Let us consider rotations of a half pitch or smaller, i.e., $0 \leq \theta \leq \Delta\theta/2$, where $\Delta\theta$ is the window pitch angle.
- By using this sinusoidal approximation for a pulse, show that we can improve the resolution of an encoder indefinitely simply by measuring the shape of each pulse at clock cycle intervals using a high-frequency clock signal.

- Problem # 2
  - Consider the two quadrature pulse signals (say, A and B) from an incremental encoder. Using sketches of these signals, show that in one direction of rotation, signal B is at a high level during the up-transition of signal A, and in the opposite direction of rotation, signal B is at a low level during the up-transition of signal A. Note that the direction of motion can be determined in this manner by using level detection of one signal during the up-transition of the other signal.

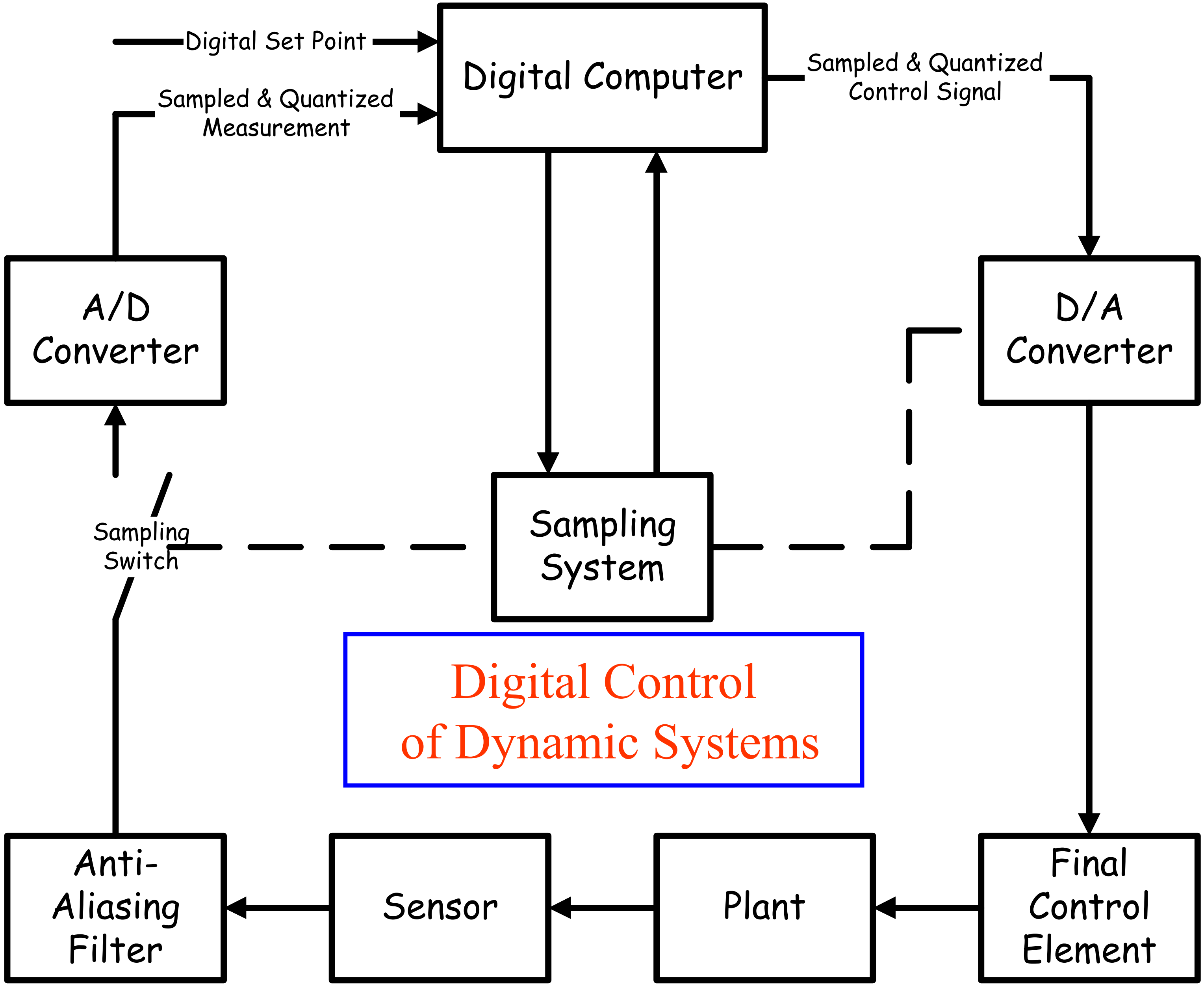
- Problem # 3
  - A shaft encoder that has N windows per track is connected to a shaft through a gear system with gear ratio p. Derive formulas for calculating angular velocity of the shaft by (a) the pulse-counting method, and (b) the pulse-timing method. What is the speed resolution in each case? What effect does the step-up gearing have on the speed resolution?

Sensors & Actuators in Mechatronics

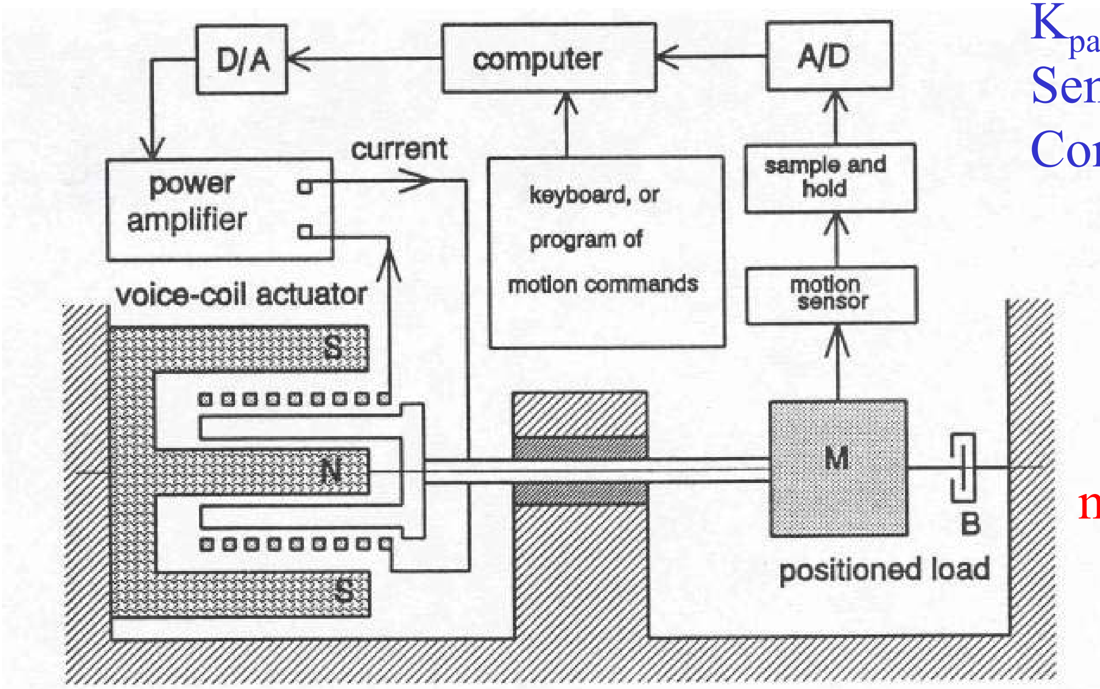
MEAE 6960
Summer 2002

Assignment # 6

Digital Control of Dynamic Systems



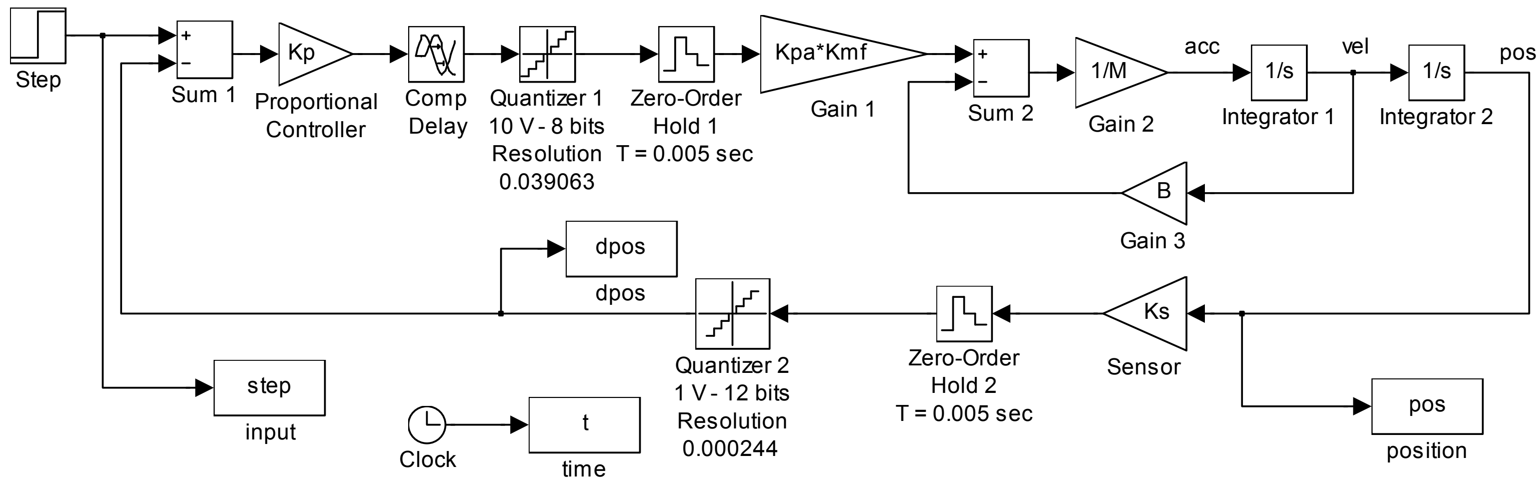
- <u>Problem # 1</u>
  - Shown is a computer-controlled motion control system.



$M = 0.001295$ lbf-s^2/in
$B = 0.259$ lbf-s/in
$K_{mf} = 0.5$ lbf/A
$K_{pa} = 2.0$ A/V
Sensor Gain = 1.0 V/in
Computational Delay = 0.008 s

pure/ideal mass and damper

- A MatLab/Simulink block diagram of this digital control system is shown below.
- Explain all the elements in this block diagram.
- Simulate this system in MatLab/Simulink for various proportional control gains (e.g., 10, 20, 30) and computational delays (e.g., 0, .001, .002). State your observations.



Analog Sensors for Motion

- Introduction
  - Importance of Accurate Measurement in Control Systems
  - Sensitivity Analysis
- Motion Transducers
  - General Discussion
- Types of Motion Transducers
  - Potentiometers (resistively-coupled transducer)
  - Variable-Inductance Transducers
  - Eddy-Current Transducers
  - Variable-Capacitance Transducers
  - Piezoelectric Transducers

Introduction

- Measurement of plant outputs is essential for feedback and feedforward control, as well as for performance evaluation of a process.
- The measurement subsystem in a control system contains sensors and transducers that detect measurands and convert them into acceptable signals – typically voltages. These voltages are then appropriately modified using signal-conditioning hardware such as filters, amplifiers, demodulators, and A/D converters.

- Impedance matching might be necessary to connect sensors and transducers to signal-conditioning hardware.
- Accuracy of sensors, transducers, and associated signal-conditioning devices is important in control system applications for two main reasons:
  – The measurement system in a feedback control system is situated in the feedback path of the control system. Even though measurements are used to compensate for the poor performance in the open-loop system, any errors in measurements themselves will enter directly into the system and cannot be corrected if they are unknown.

- Furthermore, it can be shown that sensitivity of a control system to parameter changes in the measurement system is direct. This sensitivity cannot be reduced by increasing the loop gain, unlike the case of sensitivity to the open-loop components.

- Accordingly, the design strategy for closed-loop control is to make the measurements very accurate and to employ a suitable controller to reduce other types of errors.
- Most sensor-transducer devices used in feedback control applications are analog components that generate analog output signals, that then require A/D conversion to obtain a digital representation of the measured signal.

Sensitivity Analysis

- Consider the function y = f(x). If the parameter *x* changes by an amount Δx, then *y* changes by the amount Δy. If Δx is small, Δy can be estimated from the slope dy/dx as follows:

$$\Delta y = \frac{dy}{dx}\Delta x$$

- The relative or percent change in *y* is $\Delta y/y$. It is related to the relative change in *x* as follows:

$$\frac{\Delta y}{y} = \frac{dy}{dx}\frac{\Delta x}{y} = \left(\frac{x}{y}\frac{dy}{dx}\right)\frac{\Delta x}{x}$$

- The sensitivity of *y* with respect to changes in *x* is given by:

$$S_x^y = \frac{x}{y}\frac{dy}{dx} = \frac{dy/y}{dx/x} = \frac{d(\ln y)}{d(\ln x)}$$

- Thus

$$\frac{\Delta y}{y} = S_x^y \frac{\Delta x}{x}$$

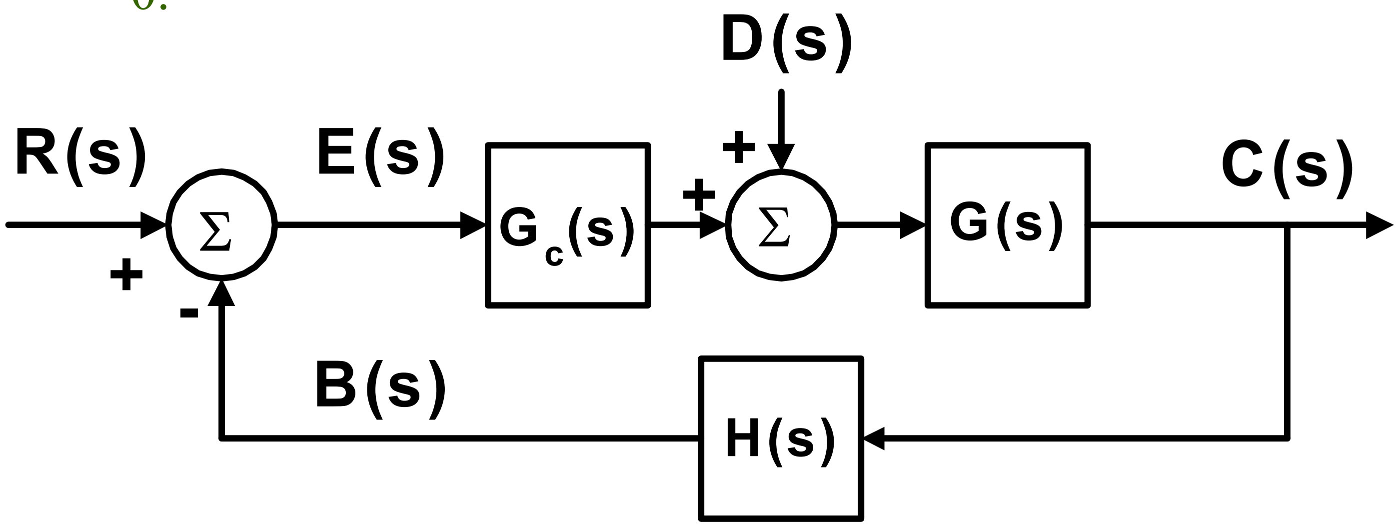
- Usually the sensitivity is not constant. For example, the function y = sin(x) has the sensitivity function:

$$S_x^y = \frac{x}{y}\frac{dy}{dx} = \frac{x}{y}\cos(x) = \frac{x\cos(x)}{\sin(x)} = \frac{x}{\tan(x)}$$

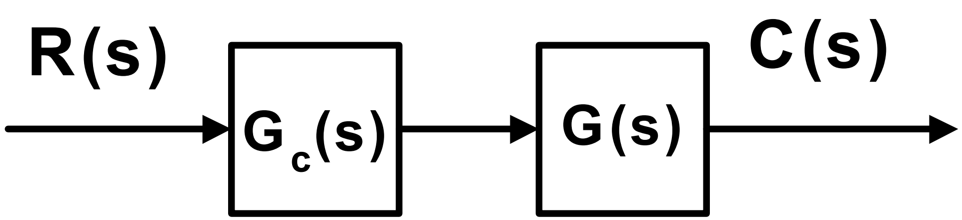
- Sensitivity of Control Systems to Parameter Variation and Parameter Uncertainty
  - A process, represented by the transfer function G(s), is subject to a changing environment, aging, ignorance of the exact values of the process parameters, and other natural factors that affect a control process.
  - In the open-loop system, all these errors and changes result in a changing and inaccurate output.
  - However, a closed-loop system senses the change in the output due to the process changes and attempts to correct the output.
  - The sensitivity of a control system to parameter variations is of prime importance.
  - Accuracy of a measurement system is affected by parameter changes in the control system components and by the influence of external disturbances.

– A primary advantage of a closed-loop feedback control system is its ability to reduce the system's sensitivity.

– Consider the closed-loop system shown. Let the disturbance D(s) = 0.



– An open-loop system's block diagram is given by:



- The system sensitivity is defined as the ratio of the percentage change in the system transfer function T(s) to the percentage change in the process transfer function G(s) (or parameter) for a small incremental change:

$$T(s) = \frac{C(s)}{R(s)}$$

$$S_G^T = \frac{\partial T / T}{\partial G / G} = \frac{\partial T}{\partial G}\frac{G}{T}$$

- For the open-loop system

$$T(s) = \frac{C(s)}{R(s)} = G_c(s)G(s)$$

$$S_G^T = \frac{\partial T / T}{\partial G / G} = \frac{\partial T}{\partial G}\frac{G}{T} = G_c(s)\frac{G(s)}{G_c(s)G(s)} = 1$$

– For the closed-loop system

$$T(s) = \frac{C(s)}{R(s)} = \frac{G_c(s)G(s)}{1 + G_c(s)G(s)H(s)}$$

$$S_G^T = \frac{\partial T / T}{\partial G / G} = \frac{\partial T}{\partial G}\frac{G}{T}$$

$$= \frac{G_c}{(1 + G_c GH)^2}\frac{G}{\frac{G_c G}{1 + G_c GH}} = \frac{1}{\left(1 + G_c GH\right)}$$

– The sensitivity of the system may be reduced below that of the open-loop system by increasing $G_c GH(s)$ over the frequency range of interest.

– Similarly $$S_{G_c}^T = \frac{\partial T / T}{\partial G_c / G_c} = \frac{\partial T}{\partial G_c}\frac{G_c}{T} = \frac{1}{\left(1 + G_c GH\right)}$$

– The sensitivity of the closed-loop system to changes in the feedback element H(s) is:

$$T(s) = \frac{C(s)}{R(s)} = \frac{G_c(s)G(s)}{1 + G_c(s)G(s)H(s)}$$

$$S_H^T = \frac{\partial T / T}{\partial H / H} = \frac{\partial T}{\partial H}\frac{H}{T}$$

$$= \frac{-(G_cG)^2}{(1 + G_cGH)^2}\frac{H}{\dfrac{G_cG}{1 + G_cGH}} = \frac{-G_cGH}{\left(1 + G_cGH\right)}$$

– When G_cGH is large, the sensitivity approaches unity and the changes in H(s) directly affect the output response. Use feedback components that will not vary with environmental changes or can be maintained constant.

- As the gain of the loop (G_cGH) is increased, the sensitivity of the control system to changes in the plant and controller decreases, but the sensitivity to changes in the feedback system (measurement system) becomes -1.
- Also the effect of the disturbance input can be reduced by increasing the gain G_cH since:

$$C(s) = \frac{G(s)}{1 + G_c(s)G(s)H(s)} D(s)$$

- Therefore:
  - Make the measurement system very accurate and stable.
  - Increase the loop gain to reduce sensitivity of the control system to changes in plant and controller.
  - Increase gain G_cH to reduce the influence of external disturbances.

- In practice:
  - G is usually fixed and cannot be altered.
  - H is essentially fixed once an accurate measurement system is chosen.
  - Most of the design freedom is available with respect to G_c only.
- It is virtually impossible to achieve all the design requirements simply by increasing the gain of G_c. The dynamics of G_c also have to be properly designed in order to obtain the desired performance of the control system.

Motion Transducers

- By motion we mean the four kinematic variables:
  - Displacement (including position, distance, proximity, and size or gage)
  - Velocity
  - Acceleration
  - Jerk
- Each variable is the time derivative of the preceding one.
- Motion measurements are extremely useful in controlling mechanical responses and interactions in dynamic systems.

- A one-to-one relationship may not always exist between a measuring device and a measured variable.
  - For example, although strain gages are devices that measure strains (and, hence, stresses and forces), they can be adapted to measure displacements by using a suitable front-end auxiliary sensor element, such as a cantilever or spring.
- Furthermore, the same measuring device may be used to measure different variables through appropriate data-interpretation techniques. For example:

- Resolver signals that provide angular displacements are differentiated to get angular velocities.
- Optical encoders can serve as both displacement and velocity transducers depending on whether the number of pulses generated is counted or the pulse rate is measured (either by counting the number of pulses during a unit interval of time or by gating a high-frequency clock signal through the pulse width).
- In principle, any force sensor can be used as an acceleration sensor, velocity sensor, or displacement sensor, depending on whether an inertia element (converting acceleration into force), a damping element (converting velocity into force), or a spring element (converting displacement into force), respectively, is used as the front-end auxiliary sensor.

- Why do we need separate transducers to measure the four kinematic variables because any one is related to any other through integration or differentiation? It should be possible, in theory, to measure only one of these four variables and use either analog or digital processing to obtain any of the remaining motion variables.
- The feasibility of this approach is highly limited and depends crucially on several factors:
  - The nature of the measured signal (e.g., steady, highly transient, periodic, narrow-band, broad-band)
  - The required frequency content of the processed signal (or the frequency range of interest)

- The signal-to-noise ratio (SNR) of the measurement
- Available processing capabilities (e.g., analog or digital processing, limitations of the digital processor, and interface, such as the speed of processing, sampling rate, and buffer size)
- Controller requirements and the nature of the plant (e.g., time constants, delays, hardware limitations)
- Required accuracy in the end objective (on which processing requirements and hardware costs will depend)

• For example, differentiation of a signal (in the time domain) is often unacceptable for noisy and high-frequency, narrow-band signals.

- Rules of Thumb:
  - In low-frequency applications (on the order of 1 Hz), displacement measurements generally provide good accuracies.
  - In intermediate-frequency applications (less than 1 kHz), velocity measurement is usually favored.
  - In measuring high-frequency motions with high noise levels, acceleration measurement is preferred.
  - Jerk is particularly useful in ground transit (ride quality), manufacturing (forging, rolling, and similar impact-type operations), and shock isolation (delicate and sensitive equipment) applications.

Typical Specifications for Analog Motion Transducers

Transducer	Measurand	Measurand frequency Max	Measurand frequency Min	Output impedance	Typical resolution	Accuracy	Sensitivity
Potentiometer	Displacement	5 Hz	DC	Low	0.1 mm	0.1%	200 mV/mm
LVDT	Displacement	2,500 Hz	DC	Low	0.001 mm or less	0.3%	50 mV/mm
Resolver	Angular displacement	500 Hz (limited by ref. frequency)	DC	Low	2 min	0.2%	10 mV/deg
Tachometer	Velocity	700 Hz	DC	Moderate	0.2 mm/s	0.5%	5 mV/mm/s 100 mV/rad/s
Eddy current proximity sensor	Displacement	100 kHz	DC	Moderate	0.001 mm, 0.05% full scale	0.5%	5 V/mm
Piezoelectric accelerometer	Acceleration (and velocity, etc.)	25 kHz	1 Hz	High	1 mm/s^2	1%	0.5 mV/m/s^2
Semiconductor strain gage	Strain (displacement, acceleration, etc.)	1 kHz (limited by fatigue)	DC	200 Ω	$1 - 10\ \mu\epsilon$ ($1\ \mu\epsilon = 10^{-6}$ unity strain)	1%	1 V/ϵ, 2,000 $\mu\epsilon$ max

Potentiometers

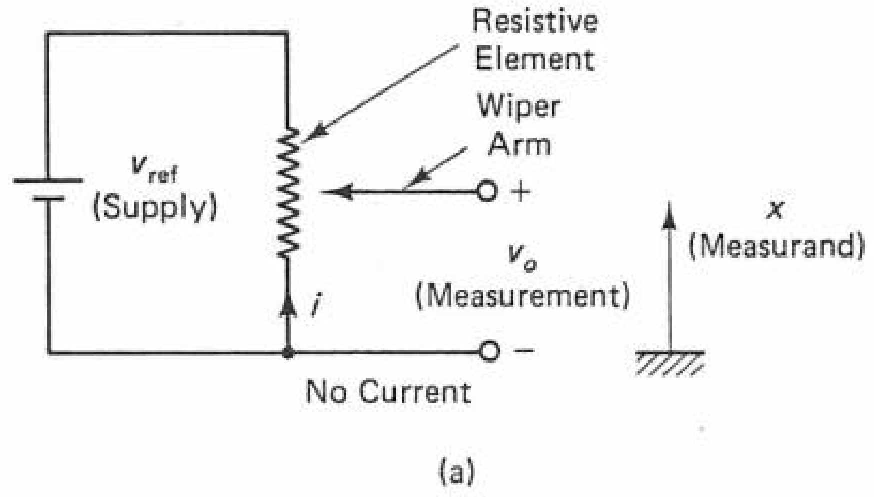
- The potentiometer is a displacement transducer.
  - It is an *active* transducer that consists of a uniform coil of wire or a film of high-resistive material (e.g., carbon, platinum, conductive plastic) whose resistance is proportional to its length.
  - A fixed voltage v_{ref} is applied across the coil or film using an external, constant DC voltage supply.
  - The transducer output signal v_o is the DC voltage between the moving contact (wiper arm) sliding on the coil and one terminal of the coil.
  - Slider displacement is proportional to the output voltage, which assumes that the output terminals are open-circuit: $\mathrm{v_o = kx}$

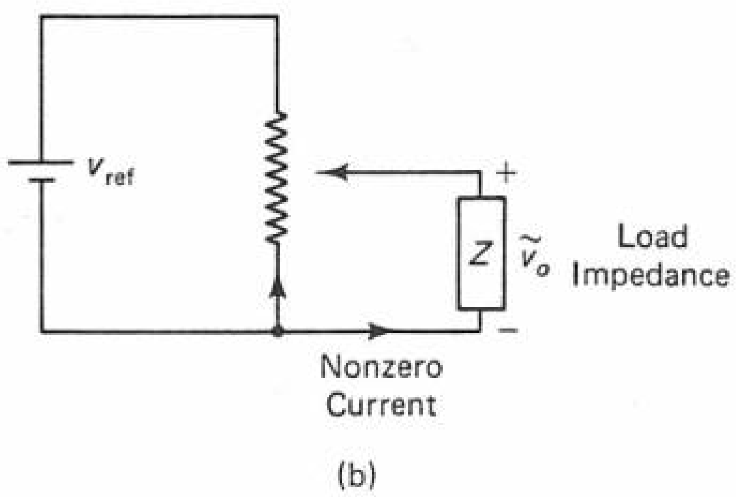
- When we assume that the output terminals are open-circuit, we are assuming an infinite-impedance load (or resistance in the present DC case) present at the output terminals, so that the output current is zero.
- In actual practice, the load (the circuitry into which the potentiometer is fed) has a finite impedance and so the output current (through the load) is nonzero.
- The output voltage thus drops, even if the reference voltage v_{ref} is assumed to remain constant under load variations (i.e., the voltage source has zero output impedance).
- This consequence is known as the *loading effect* of the transducer and the linear relationship is no longer valid. An error in the displacement reading results.

- Loading can effect the transducer reading in two ways:
  - By changing the reference voltage, i.e., loading the voltage source
  - By loading the transducer
- To reduce these effects, one needs:
  - A voltage source that is not seriously affected by load variations (i.e., a power supply with a low output impedance)
  - Data-acquisition and signal-conditioning circuitry that has a high input impedance
- <u>Remember</u>: A perfect measuring device should have the following dynamic characteristics:
  - Output instantly reaches the measured value (fast response)
  - Transducer output is sufficiently large (high gain or low output impedance)

- Output remains at the measured value (without drifting or being affected by environmental effects and other undesirable disturbances and noise) unless the measurand itself changes (stability)
- The output signal level of the transducer varies in proportion to the signal level of the measurand (static linearity)
- Connection of the measuring device does not distort the measurand itself (loading effects are absent and impedances are matched)
- Power consumption is small (high input impedance)

- Choose resistance of a potentiometer with care.
  - High resistance is preferred as this results in reduced power dissipation for a given voltage which also results in reduced thermal effects.
  - However, increased resistance increases the output impedance of the potentiometer and results in loading nonlinearity error unless the load resistance is also increased proportionately.
  - Low-resistance potentiometers have resistances less than 10 Ω.
  - High-resistance potentiometers can have resistances on the order of 100 kΩ.
  - Conductive plastics can provide high resistances (e.g., 100 Ω per mm) and have reduced friction, reduced wear, reduced weight, and increased resolution.
- Potentiometers that measure angular displacements are more common and convenient than rectilinear potentiometers.

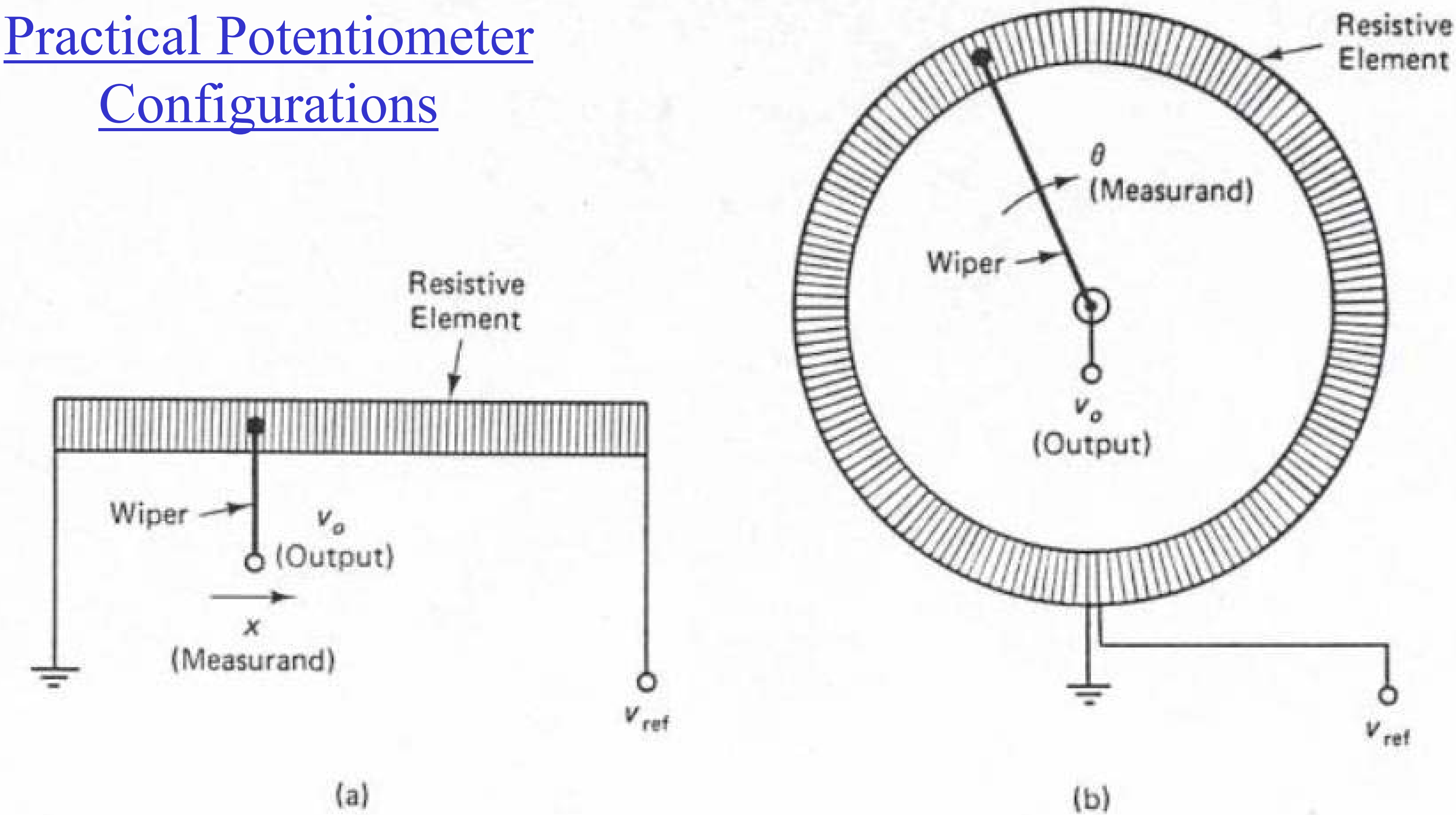




Schematic Diagram of a Potentiometer

Potentiometer Loading

Practical Potentiometer Configurations



Rectilinear Motions

Angular Motions

– Consider a rotary potentiometer and let's discuss the significance of the loading nonlinearity error caused by a purely resistive load connected to the potentiometer.

- For a general position θ of the potentiometer slider arm, suppose that the output segment of the coil is R_θ. Assuming a uniform coil:

$$R_\theta = \frac{\theta}{\theta_{max}} R_C$$

- R_C is the total resistance of the potentiometer coil.
- Current balance at the sliding contact point gives:

$$\frac{V_{ref} - V_o}{R_C - R_\theta} = \frac{V_o}{R_\theta} + \frac{V_o}{R_L}$$

- R_L is the load resistance.

- Combining equations results in:

$$\frac{V_{ref} - V_o}{1-\left(\theta/\theta_{max}\right)} = \frac{V_o}{\theta/\theta_{max}} + \frac{V_o}{R_L/R_C}$$

$$\frac{V_o}{V_{ref}} = \frac{\left(\theta/\theta_{max}\right)\left(R_L/R_C\right)}{\left(R_L/R_C\right)+\left(\theta/\theta_{max}\right)-\left(\theta/\theta_{max}\right)^2}$$

- Loading error appears to be high for low values of the R_L/R_C ratio. Good accuracy is possible for $R_L/R_C > 10$, particularly for small values of θ/θ_{max}.
- Hence to reduce loading error in potentiometers: (1) Increase R_L/R_C (increase load impedance, reduce coil impedance); and (2) Use potentiometers to measure small values of θ/θ_{max} (or calibrate only a small segment of the element for linear reading).

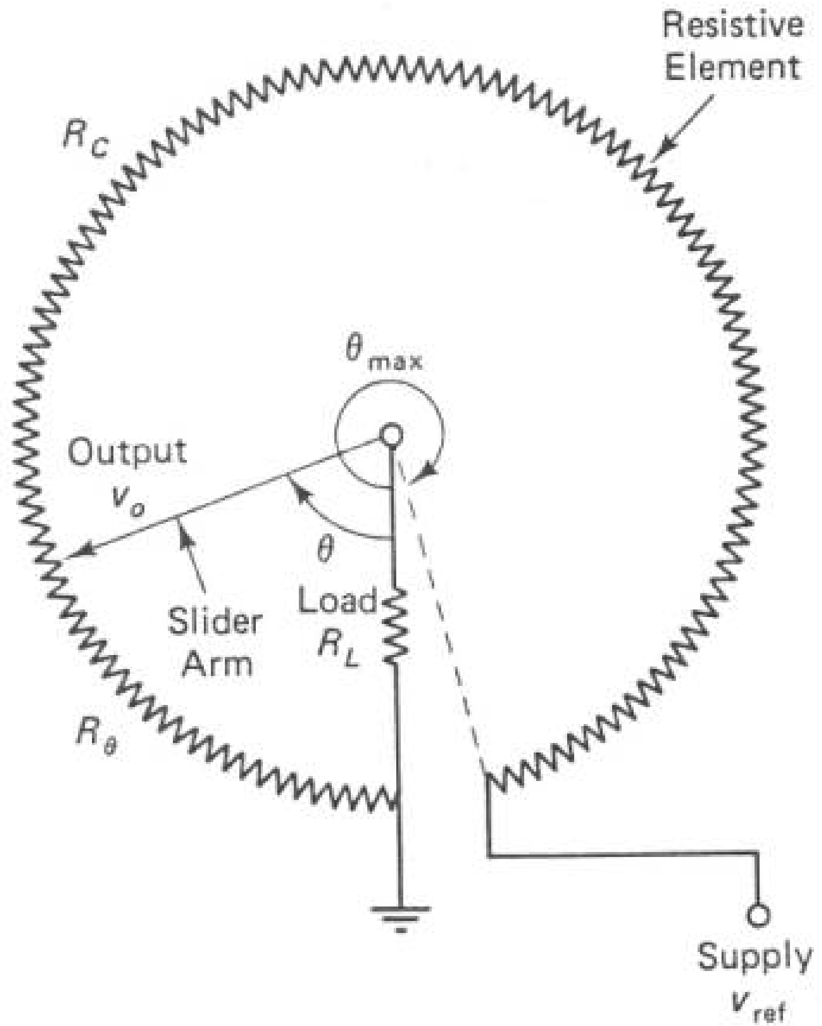
- The loading nonlinearity error is defined by:

$$e = \frac{\left(v_o / v_{ref}\right) - \left(\theta / \theta_{max}\right)}{\theta / \theta_{max}} 100\%$$

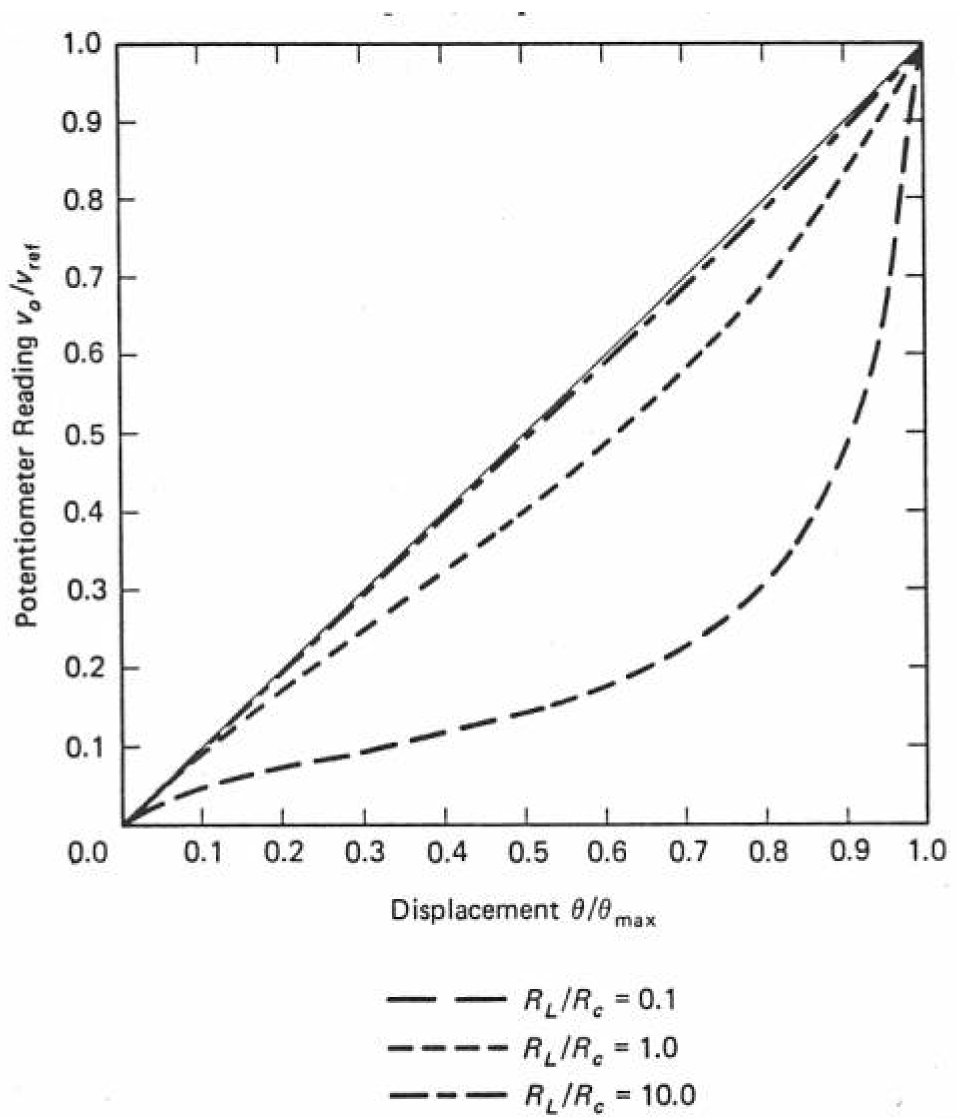
  - Using only a segment of the resistance element as the range of the potentiometer is similar to adding two end resistors to the elements. It is known that this tends to linearize the potentiometer.
  - If the load resistance is known to be small, a voltage follower may be used at the potentiometer output to virtually eliminate loading error. This arrangement provides a high load impedance to the potentiometer and a low impedance at the output of the amplifier.

– Three obvious disadvantages of this resistively-coupled transducer:
  - The force required to move the slider arm comes from the motion source, and the resulting energy is dissipated through friction. This energy conversion involves relatively high forces and the energy is wasted rather than being converted into the output signal of the transducer.
  - The electrical energy from the reference source is dissipated through the resistor coil (or film) resulting in an undesirable temperature rise.
  - Finite resolution in coil-type potentiometers, where resolution is determined by the number of turns in the coil. Infinitesimal resolutions are now possible with high-quality resistive-film potentiometers that use conductive plastics. In this case, resolution is limited by other factors (e.g., mechanical limitations and signal-to-noise ratio).

A Rotary Potentiometer with a Resistive Load



Loading Nonlinearity in a Potentiometer



Loading Nonlinearity in a Potentiometer

Loading Nonlinearity Error in Potentiometer
For $\theta / \theta_{max} = 0.5$

Load Resistance Ratio R_L / R_C	Loading Nonlinearity Error e
0.1	-71.4%
1.0	-20%
10.0	-2.4%

– Limitations of Potentiometers as Displacement-Measuring Devices:

- The force needed to move the slider (against friction and arm inertia) is provided by the displacement source. The mechanical loading distorts the measured signal itself.
- High-frequency (or highly transient) measurements are not feasible because of such factors as slider bounce, friction and inertia resistance, and induced voltages in the wiper arm and primary coil.
- Variations in the supply voltage cause error.
- Electrical loading error can be significant when the load resistance is low.
- Resolution is limited by the number of turns in the coil and by the coil uniformity. This limits small displacement measurements.

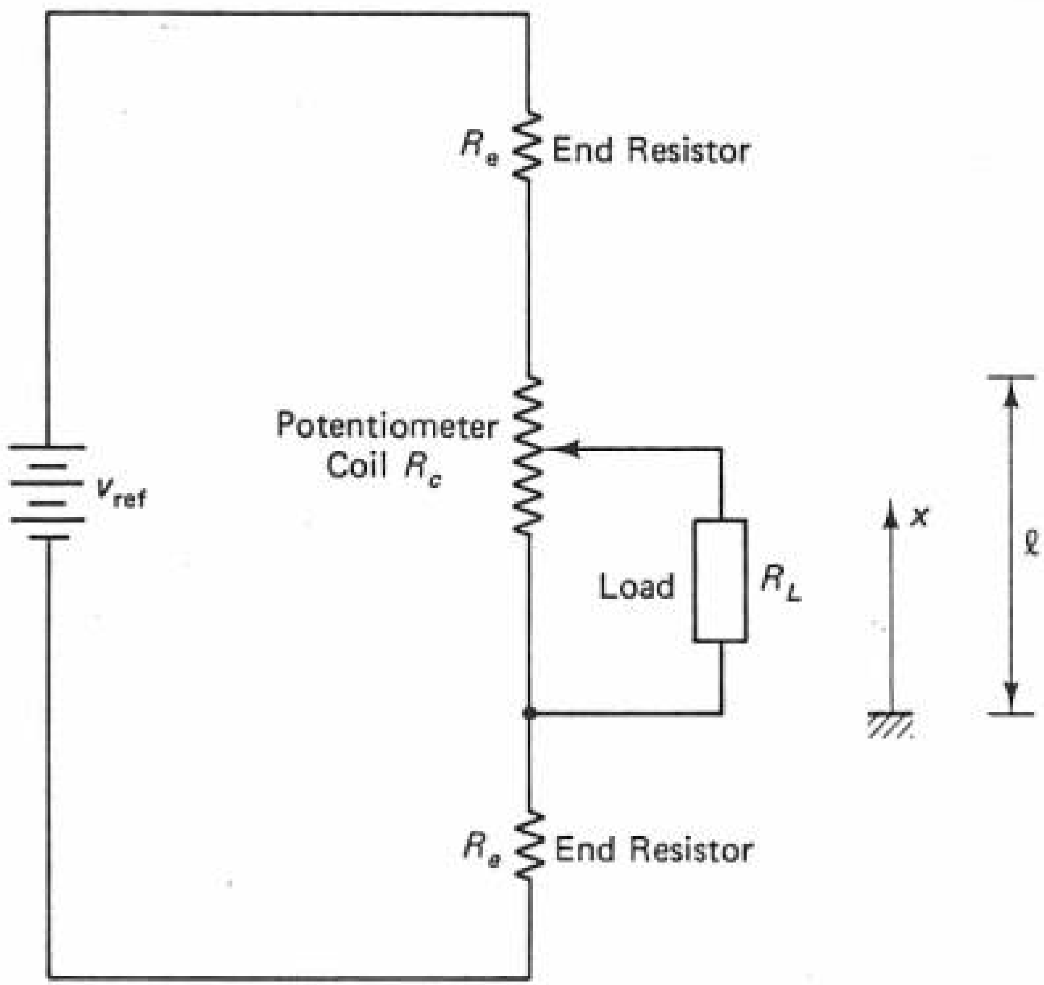
- Wearout and heating up (with associated oxidation) in the coil (film) and slider contact cause accelerated degradation.

– Advantages Associated with Potentiometer Devices:
  - They are relatively less costly.
  - Potentiometers provide high-voltage (low impedance) output signals, requiring no amplification in most applications. Transducer impedance can be varied simply by changing the coil resistance and supply voltage.

– Although potentiometers are primarily displacement transducers, they can be adapted to measure other types of signals, such as pressure and force, using appropriate auxiliary sensor (front-end) elements.

- <u>Problem</u>
  - A potentiometer circuit with element resistance R_C and equal end resistors R_e is shown. Derive the necessary input/output relations. Show that the end resistors can produce a linearizing effect in the potentiometer. At half the maximum reading of the potentiometer, calculate the percentage loading error for the three values of the resistance ratio $R_C/R_e = 0.1, 1.0, 10.0$, assuming that the load resistance R_L is equal to the element resistance. Compare the results with the corresponding value for $R_e = 0$. Finally, choose a suitable value for R_C/R_e and plot the curve of percentage loading error versus fractional displacement x/x_{max}. From the graph, estimate the maximum loading error.

Potentiometer Circuit with End Resistors



Variable-Inductance Transducers

- Motion transducers that employ the principle of electromagnetic induction are termed variable-inductance transducers.
  - When the flux linkage through an electrical conductor changes, a voltage is induced in the conductor. This, in turn, generates a magnetic field that opposes the primary field. Hence, a mechanical force is necessary to sustain the change of flux linkage. If the change in flux linkage is brought about by a relative motion, the mechanical energy is directly converted into electrical energy. This is the basis of electromagnetic induction and the principle of operation of variable-inductance transducers.

- In these devices, the change of flux linkage is caused by a mechanical motion and the mechanical-to-electrical energy transfer takes place under near-ideal conditions.
- The induced voltage or change in inductance may be used as a measure of the motion. Variable-inductance transducers are generally electromechanical devices coupled by a magnetic field.
- There are many different types of variable-inductance transducers. Three primary types can be identified:
  - Mutual-Induction transducers
  - Self-induction transducers
  - Permanent-magnet transducers

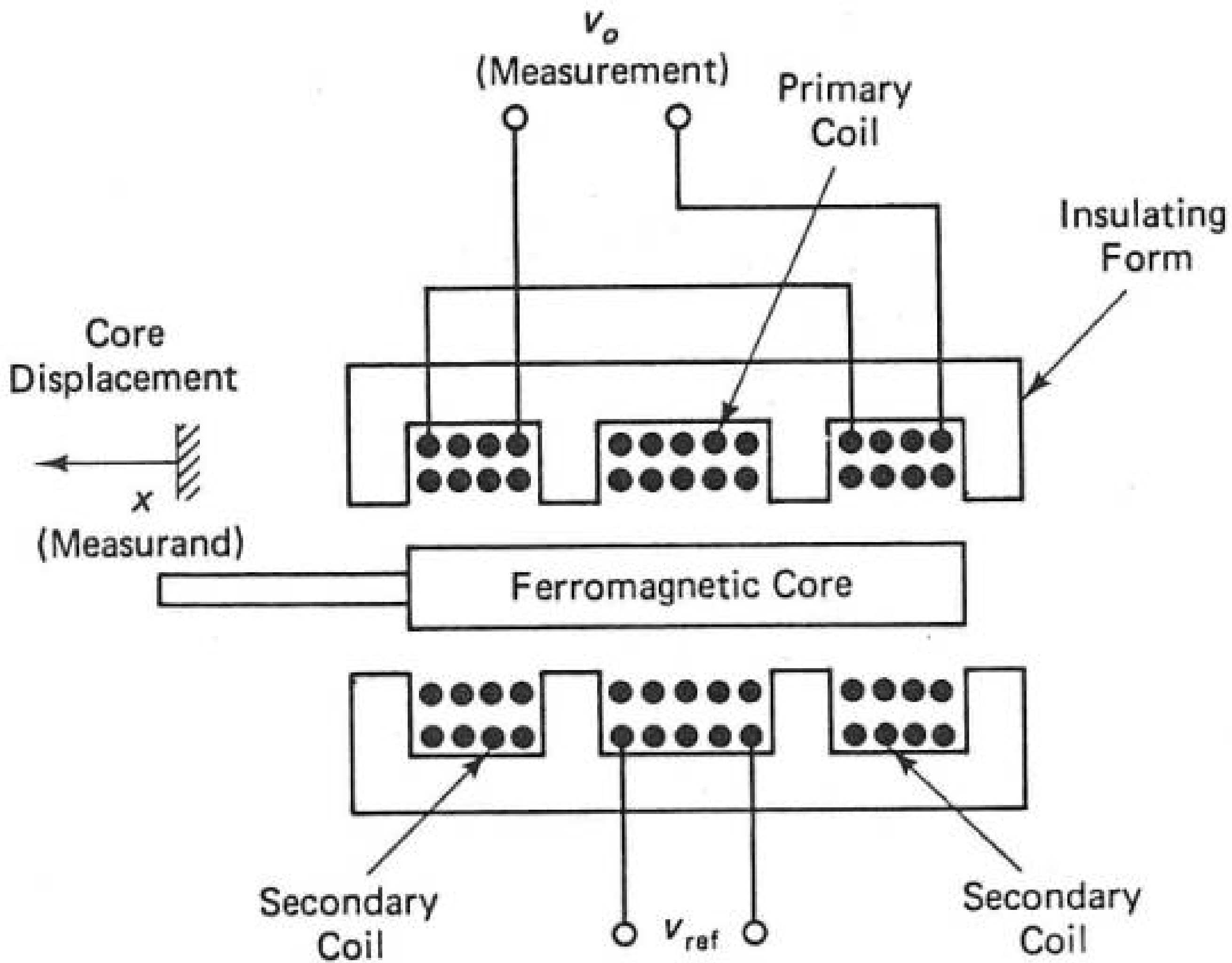
– Variable-inductance transducers that use a non-magnetized ferromagnetic medium to alter the reluctance (magnetic resistance) of the flux path are known as *variable-reluctance transducers*. Some of the mutual-inductance transducers and most of the self-inductance transducers are of this type. Permanent-magnet transducers are not considered variable-reluctance transducers.

- Mutual-Induction Transducers and Differential Transformers
  - The basic arrangement of a mutual-induction transducer constitutes two coils, the *primary winding* and the *secondary winding*.
  - The primary winding carries an AC excitation that induces a steady AC voltage in the secondary winding.
  - The level of the induced voltage depends on the flux linkage between the coils.
  - In mutual-induction transducers, a change in the flux linkage is effected by one of two common techniques.
    - One technique is to move an object made of ferromagnetic material within the flux path. This changes the reluctance of the flux path, with an associated change of the flux linkage in the secondary coil.

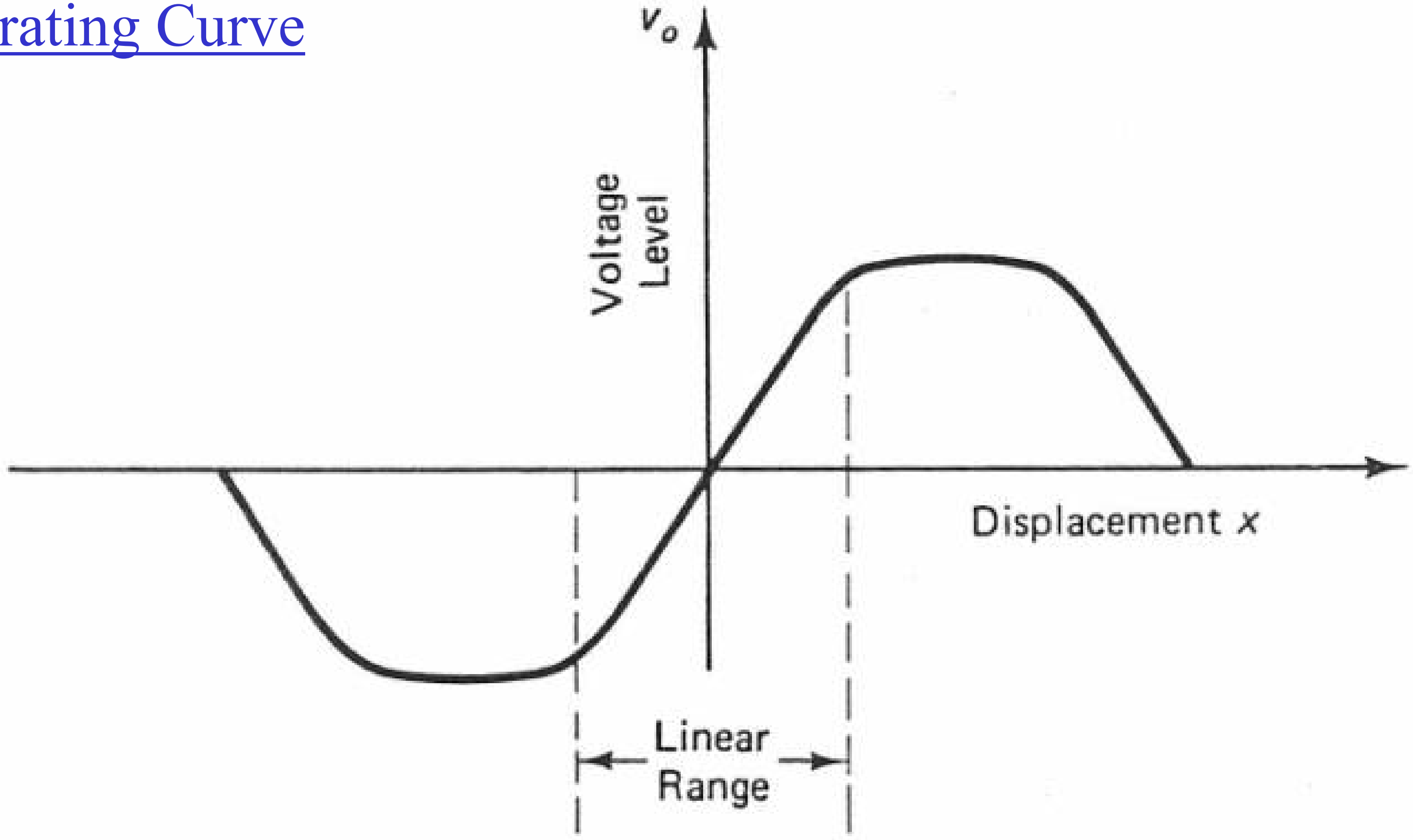
- This is the operating principle of the linear variable differential transformer (LVDT), the rotary variable differential transformer (RVDT), and the mutual-induction proximity probe. All of these are in fact variable-reluctance transducers.
- The other common way to change the flux linkage is to move one coil with respect to the other. This is the operating principle of the resolver. This is not a variable-reluctance transducer.
- The motion can be measured by using the secondary signal in several ways: (1) The AC signal in the secondary winding may be demodulated by rejecting the carrier frequency (primary-winding excitation frequency) and directly measuring the resulting signal, which represents the motion. This method is particularly suitable for measuring transient motions; (2) the amplitude of the secondary (induced) voltage may be measured; (3) measure the change of inductance in the secondary circuit directly, by using a device such as an inductance bridge circuit.

- Linear Variable Differential Transformer (LVDT)
  - The LVDT is a displacement-measuring device that overcomes most of the shortcomings of the potentiometer.
  - It is considered a passive transducer because the measured displacement provides energy for “changing” the induced voltage, even though an external power supply is used to energize the primary coil which in turn induces a steady carrier voltage in the secondary coil.
  - The LVDT is a variable-reluctance transducer of the mutual induction type.

Schematic Diagram of a LVDT

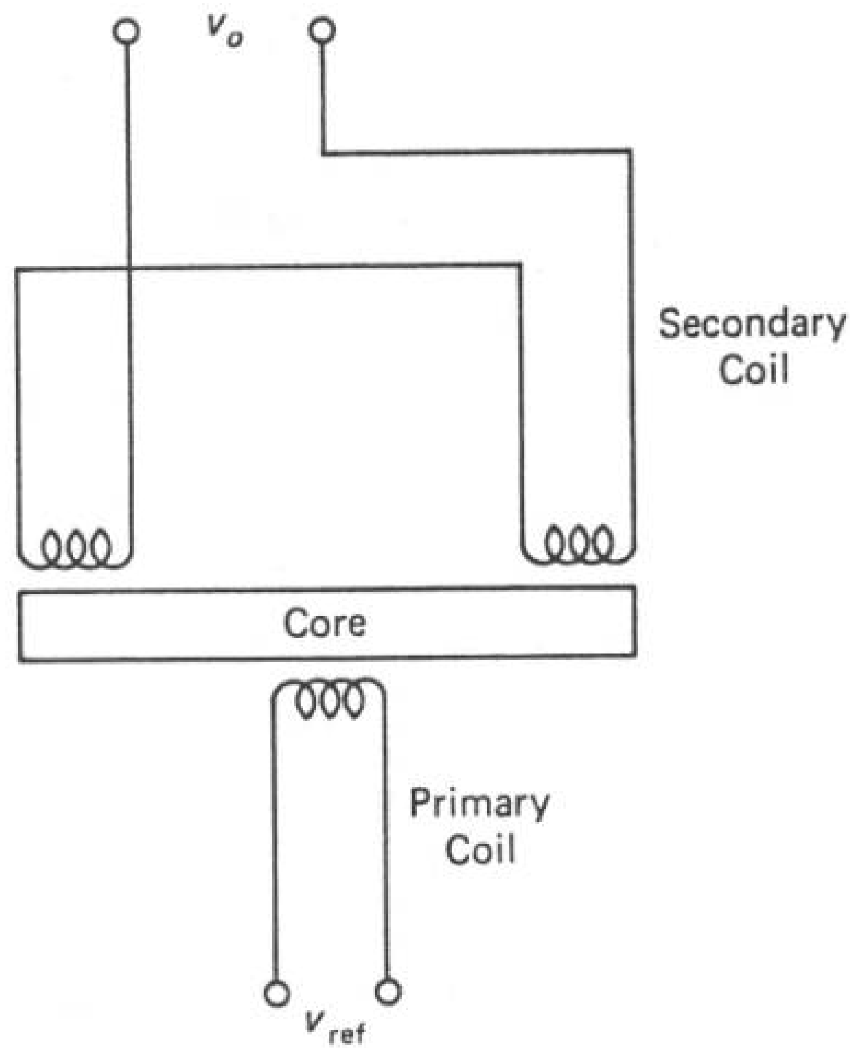


Typical LVDT Operating Curve



- In its simplest form, the LVDT consists of a cylindrical, insulating, nonmagnetic form that has a primary coil in the mid-segment and a secondary coil symmetrically wound in the two end segments.
- The primary coil is energized by an AC supply of voltage v_{ref}. This will generate by mutual induction an AC of the same frequency in the secondary winding.
- A core of ferromagnetic material is inserted coaxially into the cylindrical form without actually touching it.
- As the core moves the reluctance of the flux path changes. Hence, the degree of flux linkage depends on the axial position of the core.

Series Opposition Connection of Secondary Windings



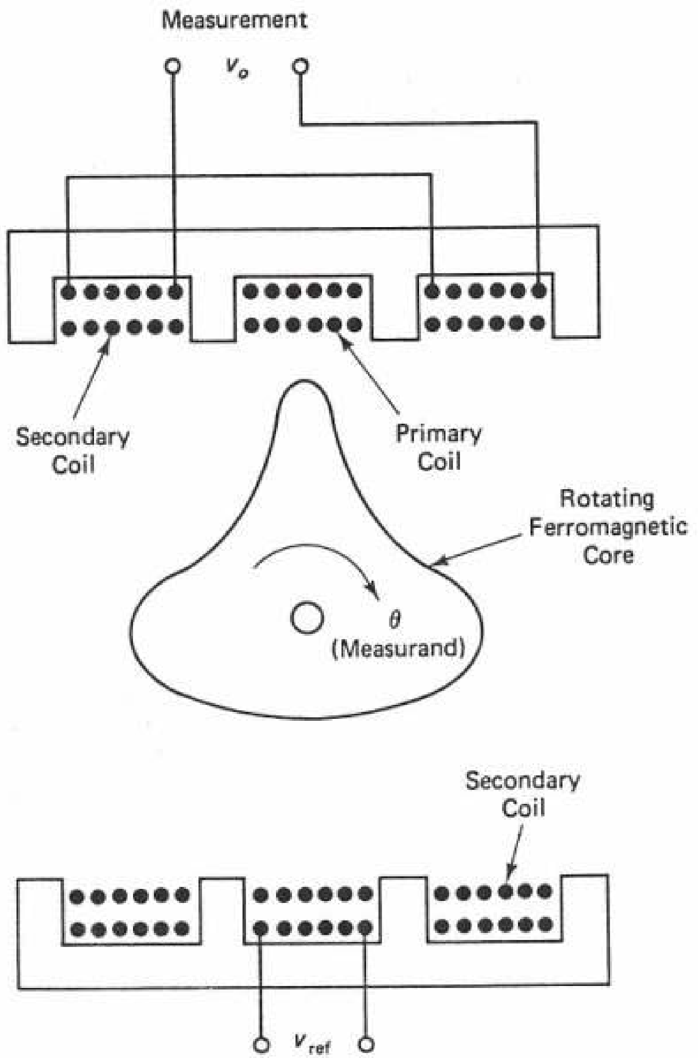
- Since the two secondary coils are connected in series opposition, so that the potentials induced in the two secondary coil segments oppose each other, the net induced voltage is zero when the core is centered between the two secondary winding segments. This is known as the *null position*.
- When the core is displaced from this position, a nonzero induced voltage will be generated. At steady-state, the amplitude v_o of this induced voltage is proportional, in the linear (operating) region, to the core displacement x. Consequently, v_o may be used as a measure of the displacement.
- Because of opposed secondary windings, the LVDT provides direction as well as the magnitude of the displacement.

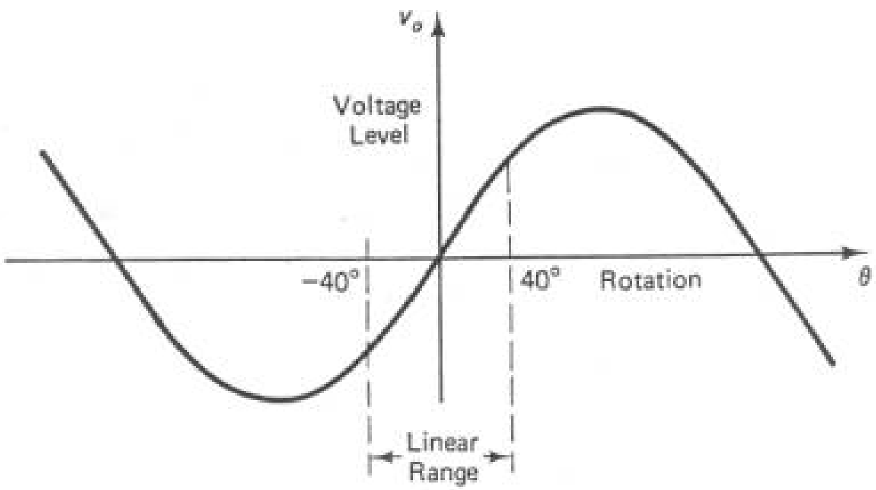
- For an LVDT to measure transient motions accurately, the frequency of the reference voltage (the carrier frequency) has to be at least 10 times larger than the largest significant frequency component in the measured motion. For quasi-dynamic displacements and slow transients on the order of a few hertz, a standard AC supply (at 60 Hz line frequency) is adequate. The performance (particularly sensitivity and accuracy) is known to improve with the excitation frequency, however.

- Advantages of the LVDT include:
  - It is essentially a non-contacting device with no frictional resistance. Near-ideal electromechanical energy conversion and light-weight core result in very small resistive forces. Hysteresis (both magnetic hysteresis and mechanical backlash) is negligible.
  - It has low output impedance, typically on the order of 100 Ω. Signal amplification is usually not needed.
  - Directional measurements (positive/negative) are obtained.
  - It is available is small size, e.g., 1 cm long with maximum travel of 2 mm.
  - It has a simple and robust construction (inexpensive and durable).
  - Fine resolutions are possible (theoretically, infinitesimal resolution; practically, much better than a coil potentiometer).

- Rotary Variable Differential Transformer (RVDT)
  - The RVDT operates using the same principle as the LVDT, except that in an RVDT, a rotating ferromagnetic core is used.
  - The RVDT is used for measuring angular displacements.
  - The rotating core is shaped so that a reasonably wide linear operating region is obtained.
  - Advantages of the RVDT are essentially the same as those cited for the LVDT.
  - The RVDT measures angular motions directly, without requiring nonlinear transformations (as is the case for resolvers). The linear range is typically ± 40°, with a nonlinearity error less than 1 percent.

Schematic Diagram of the RVDT





Operating Curve

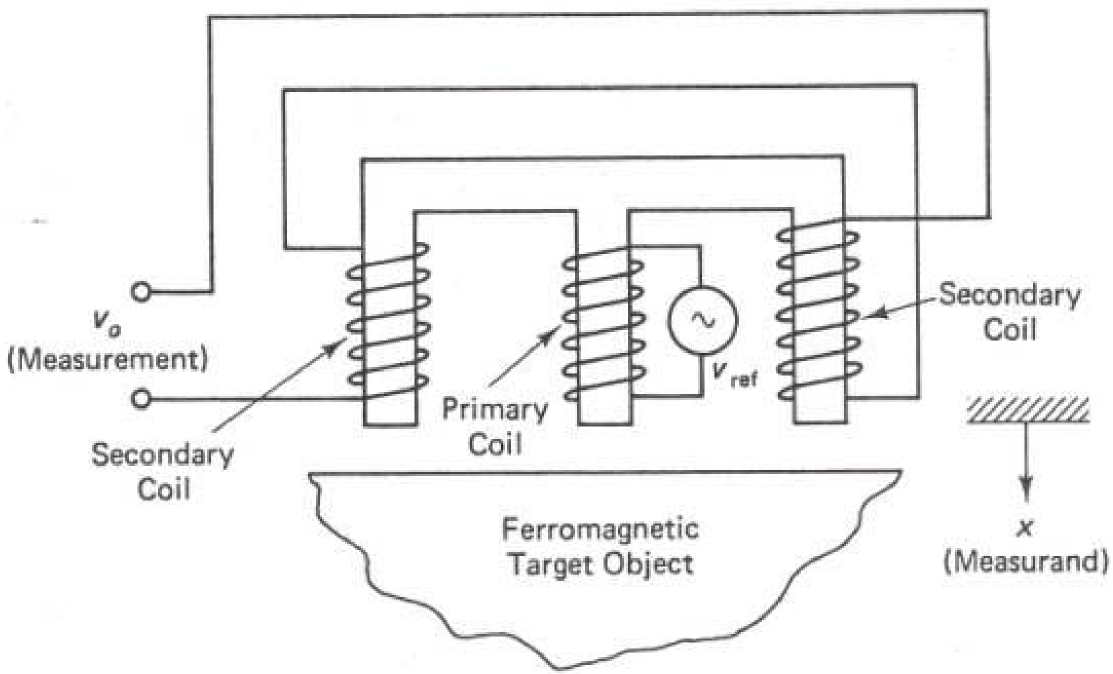
- In variable-inductance devices, the induced voltage is generated through the rate of change of the magnetic flux linkage. Therefore, displacement readings are distorted by velocity; similarly, velocity readings are affected by acceleration.
- For the same displacement value, the transducer reading will depend on the velocity at that displacement. This error is known to increase with the ratio (cyclic velocity of the core / carrier frequency). Hence, these rate errors can be reduced by increasing carrier frequency.

- Mutual-Induction Proximity Sensor
  - This displacement operates also on the mutual-induction principle.
  - The insulating core carries the primary winding in its middle limb. The two end limbs carry the secondary windings that are connected in series.
  - Unlike the LVDT and the RVDT, the two voltages induced in the secondary winding segments are additive in this case.
  - The region of the moving surface (target object) that faces the coils has to be made of ferromagnetic material so that as it moves, the magnetic reluctance and the flux linkage will change.

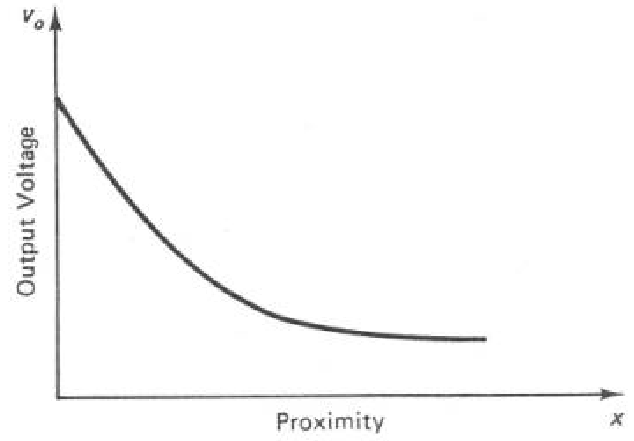
- This, in turn, changes the induced voltage in the secondary windings, and this change is a measure of the displacement.
- Unlike the LVDT, which has an axial displacement configuration, the proximity probe has a transverse displacement configuration. It measures transverse displacements or proximities of moving objects.
- The displacement-voltage relation of a proximity probe is nonlinear. Hence, these proximity sensors should only be used for measuring very small displacements, unless accurate nonlinear calibration curves are available.
- The proximity sensor is a non-contact device and so mechanical loading is negligible.

- Because a ferromagnetic object is used to alter the reluctance of the flux path, the mutual-inductance proximity sensor is a variable-reluctance device.
- Proximity sensors are used in a wide variety of applications pertaining to non-contacting displacement sensing and dimensional gaging, e.g., level detection, angular speed measurement at steady state, detecting surface irregularities in machined parts, measurement and control of the gap between a robotic welding torch head and the work surface, gaging the thickness of metal plates in manufacturing operations.

Schematic Diagram of the Mutual-Induction Proximity Sensor



Operating Curve



- Resolver
  - This mutual-induction displacement transducer depends on relative motion between the primary coil and the secondary coil to produce a change in flux linkage.
  - It is not a variable-reluctance transducer because it does not employ a ferromagnetic moving element.
  - It is widely used for measuring angular displacements.
  - The rotor contains the primary coil. It consists of a single two-pole winding element energized by an AC supply voltage v_{ref}. The rotor is directly attached to the object whose rotation is being measured.
  - The stator consists of two sets of windings placed 90° apart.

- If the angular position of the rotor with respect to one pair of stator windings is denoted by θ, the induced voltage in this pair of windings is given by:

$$v_{o1} = av_{ref}\cos\theta$$

- The induced voltage in the other pair of windings is given by:

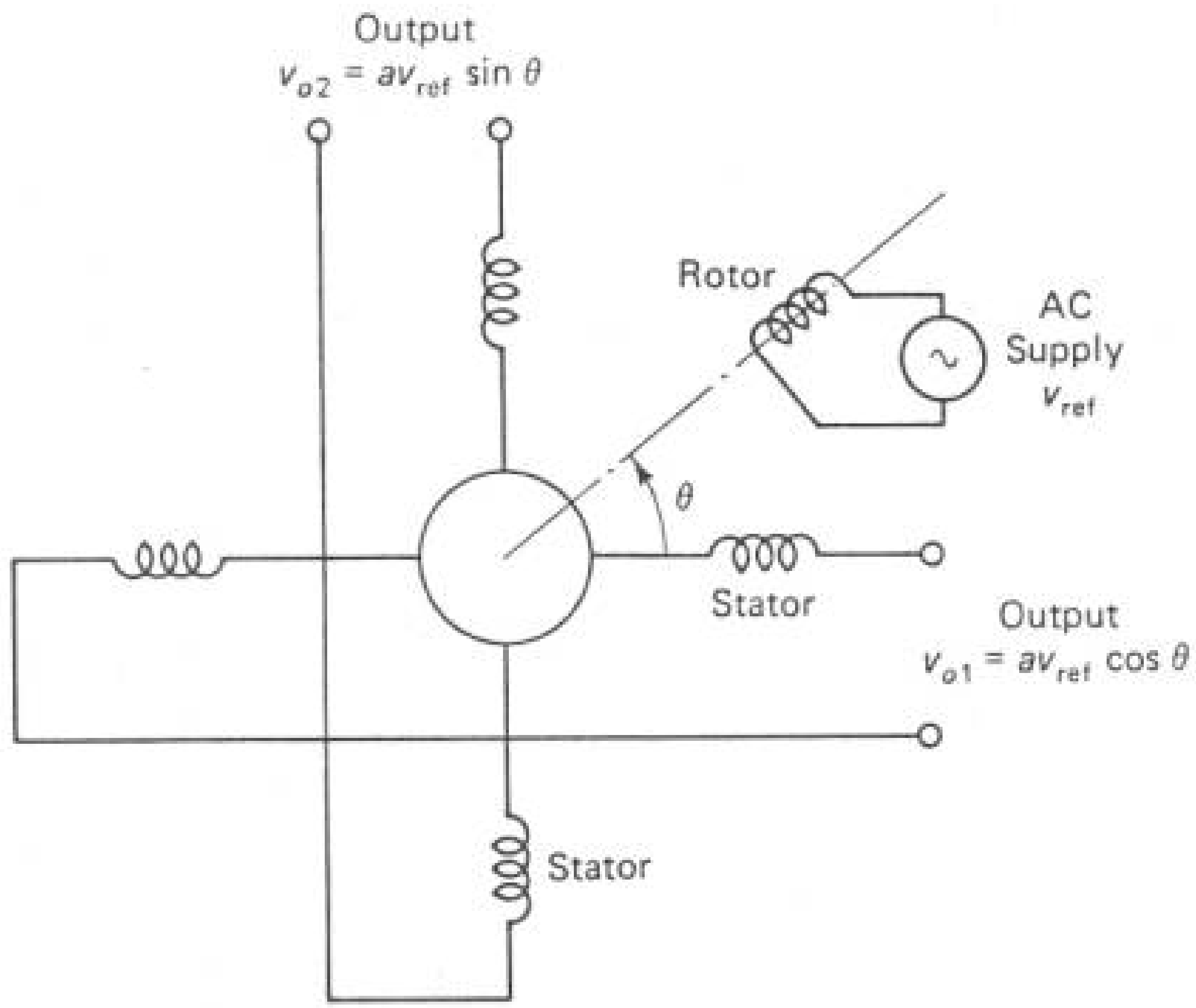
$$v_{o2} = av_{ref}\sin\theta$$

- Note that these are amplitude-modulated signals; the carrier signal v_{ref} is modulated by the motion θ. The constant parameter a depends primarily on geometric and material characteristics of the device.
- Either of the two output signals may be used to determine the angular position in the first quadrant ($0 \leq \theta \leq 90°$).

- Both signals are needed, however, to determine the displacement (direction as well as magnitude) in all four quadrants ($0 \leq \theta \leq 360°$) without causing any ambiguity.
- As for differential transformers, transient displacement signals can be extracted by demodulating the modulated outputs. This is accomplished by filtering out the carrier signal, thereby extracting the modulating signal.
- The output signals of a resolver are nonlinear (trigonometric) functions of the angle of rotation.

- The primary advantages of the resolver include:
  - Fine resolution and high accuracy
  - Low output impedance (high signal levels)
  - Small size
  - Simple and robust operation
- Its main limitations are:
  - Nonlinear output signals (an advantage in some applications where trigonometric functions of rotations are needed)
  - Bandwidth limited by supply frequency
  - Slip rings and brushes needed (which adds mechanical loading and also creates wearout, oxidation, and thermal and noise problems).

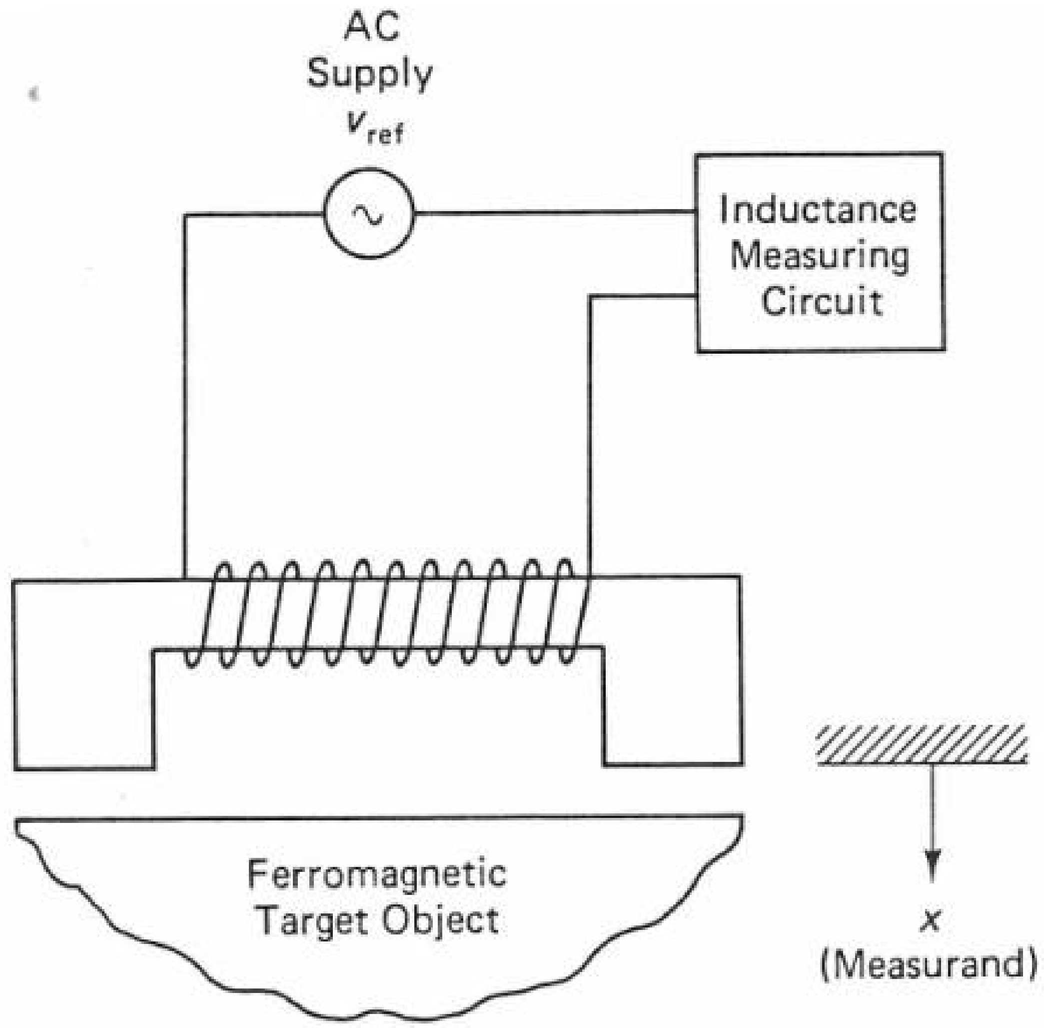
Schematic of a Resolver



- Self-Induction Transducers
  - These transducers are based on the principle of self-induction. Unlike mutual-induction transducers, only a single coil is employed. This coil is activated by an AC supply voltage v_{ref}.
  - The current produces a magnetic flux, which is linked with the coil. The level of flux linkage (or self-inductance) can be varied by moving a ferromagnetic object within the magnetic field. This changes the reluctance of the flux path and the inductance of the coil. This change is a measure of the displacement of the ferromagnetic object. The change in inductance is measured using an inductance measuring circuit.

- Note that self-inductance transducers are usually variable-reluctance devices.
- A self-induction proximity sensor can be used as a displacement sensor for transverse displacements, e.g., the distance between the sensor tip and the ferromagnetic surface of a moving object can be measured.
- High-speed displacement measurements can result in velocity error (rate error) when the variable-inductance displacement sensors (including self-induction transducers) are used. This effect may be reduced, as in other AC-powered variable-inductance sensors, by increasing the carrier frequency.

Diagram of a Self-Inductance Proximity Sensor



- <u>Permanent-Magnet Transducers</u>
  - A distinctive feature of permanent-magnet transducers is that they have a permanent magnet to generate a uniform and steady magnetic field.
  - A relative motion between the magnetic field and an electric conductor induces a voltage that is proportional to the speed at which the conductor crosses the magnetic field.
  - In some designs, a unidirectional magnetic field generated by a DC supply, i.e., an electromagnet, is used in place of a permanent magnet. Nevertheless, this is generally termed a permanent-magnet transducer.

- Permanent-Magnet Speed Transducers
  - The principle of electromagnetic induction between a permanent magnet and a conducting coil is used in speed measurement by permanent-magnet transducers.
  - Depending on the configuration, either rectilinear speeds or angular speeds can be measured.
  - Note that these are passive transducers, because the energy for the output signal v_o is derived from the motion (measured signal) itself.
  - The entire device is usually enclosed in a steel casing to isolate it from ambient magnetic fields.

- In the rectilinear velocity transducer, the conductor coil is wrapped on a core and placed centrally between two magnetic poles, which produce a cross-magnetic field.
- The core is attached to the moving object whose velocity must be measured. The velocity v is proportional to the induced voltage v_o.
- A moving-magnet and fixed-coil arrangement can also be used, thus eliminating the need for any sliding contacts (slip rings and brushes) for the object leads, thereby reducing mechanical loading error, wearout, and related problems.
- The tachometer is a very common permanent-magnet device. Here the rotor is directly connected to the rotating object.

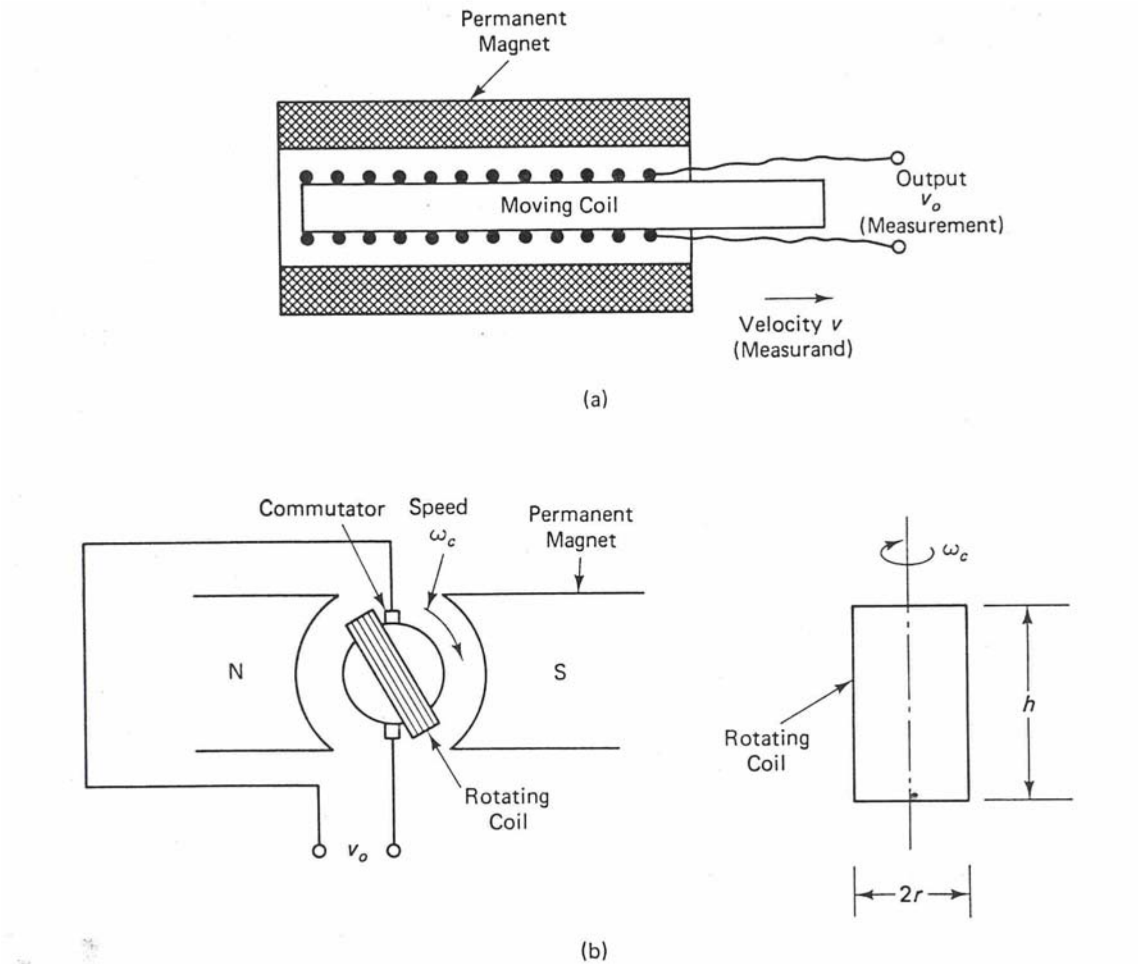
- The output signal that is induced in the rotating coil is picked up as a DC voltage v_o using a suitable commutator device – typically consisting of a pair of low-resistance carbon brushes – that is stationary but makes contact with the rotating coil through slip rings so as to maintain the positive direction of induced voltage throughout each revolution.
- The induced voltage is given by: $\mathrm{v_o = \left(2nhr\beta\right)\omega_c}$
- *h* is the coil height, *2r* is the coil width, *n* is the number of turns in the coil, β is the flux density of the uniform magnetic field, and ω_c is the angular speed.

- When tachometers are used to measure transient velocities, some error will result from the rate (acceleration) effect. This error generally increases with the maximum significant frequency that must be retained in the transient velocity signal.
- Output distortion can also result because of reactive (inductive and capacitive) loading of the tachometer. Both types of error can be reduced by increasing the load impedance.

Permanent Magnet Transducers:

(a) Rectilinear velocity transducer

(b) DC tachometer-generator



- Eddy-Current Transducers
  - If a conducting (i.e., low-resistivity) medium is subjected to a fluctuating magnetic field, eddy currents are generated in the medium. The strength of eddy currents increases with the strength of the magnetic field and the frequency of the magnetic flux.
  - This principle is used in eddy-current proximity sensors. Eddy-current sensors may be used as either dimensional gaging devices or displacement sensors.
  - Unlike variable-inductance proximity sensors, the target object of the eddy-current sensor does not have to be made of a ferromagnetic material. A conducting target object is needed, but a thin film conducting material is adequate.

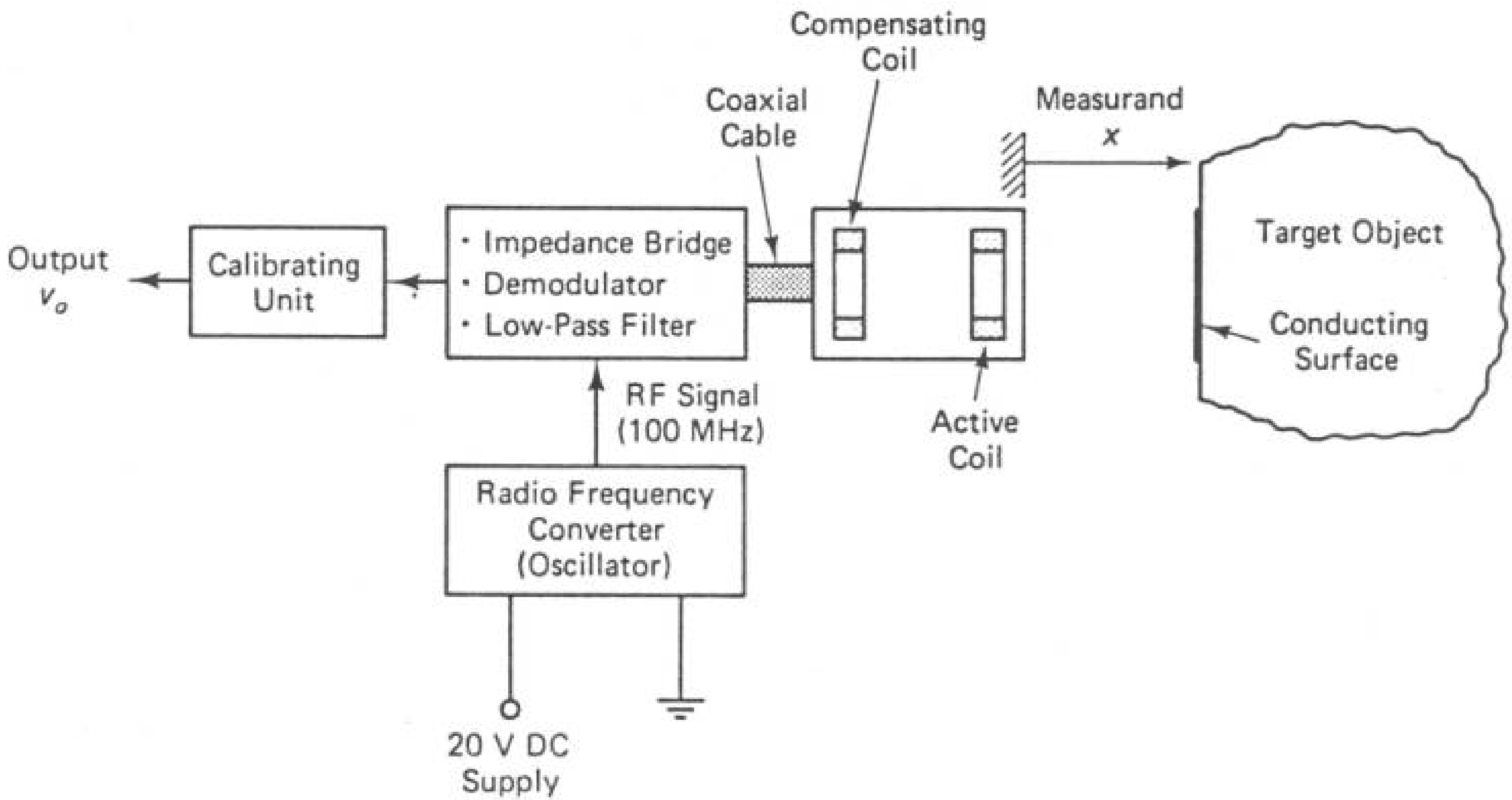
- The probe head has two identical coils, which form two arms of an impedance bridge. The coil closer to the probe face is the active coil. The other coil is the compensating coil. It compensates for ambient changes, particularly thermal effects.
- The other two arms of the bridge consist of purely resistive elements.
- The bridge is excited by a radio-frequency voltage supply, the frequency ranging from 1 MHz to 100 MHz. This signal is generated from a radio-frequency converter (an oscillator) that is typically powered by a 20-volt DC supply.
- In the absence of the target object, the output of the impedance bridge is zero, which corresponds to the balanced condition.

- When the target object is moved close to the sensor, eddy currents are generated in the conducting medium because of the radio-frequency magnetic flux from the active coil. The magnetic field of the eddy currents opposes the primary field that generates these currents. Hence, the inductance of the active coil increases, creating an imbalance in the bridge. The resulting output from the bridge is an amplitude-modulated signal containing the radio-frequency carrier. This signal is demodulated by removing the carrier. The resulting signal (modulating signal) measures the transient displacement of the target object.
- Low-pass filtering is used to remove high-frequency leftover noise in the output signal once the carrier is removed.

- For large displacements, the output is not linearly related to the displacement. Furthermore, the sensitivity of the eddy-current probe depends nonlinearly on the nature of the conducting medium, particularly the resistivity (for low resistivities, sensitivity increases with resistivity; for high resistivities, sensitivity decreases with resistivity).
- The facial area of the conducting medium on the target has to be slightly larger than the frontal area of the eddy-current probe head. If the target area has a curved surface, its radius of curvature has to be at least four times the diameter of the probe.
- Eddy-current sensors are medium-impedance devices; 1000 Ω output impedance is typical.

- Sensitivity is on the order of 5 V/mm.
- Since the carrier frequency is very high, eddy-current devices are suitable for highly transient displacement measurements (e.g., bandwidths up to 100 kHz).
- Another advantage of the eddy-current sensor is that it is a non-contacting device; there is no mechanical loading on the moving (target) object.

Eddy-Current Proximity Sensor



Impedance Bridge