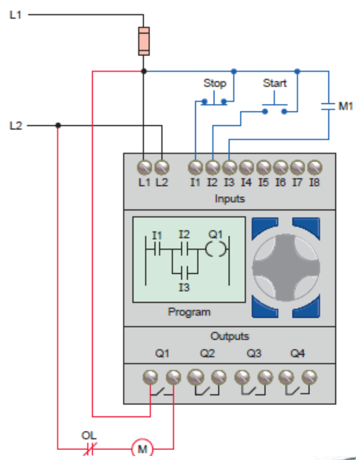
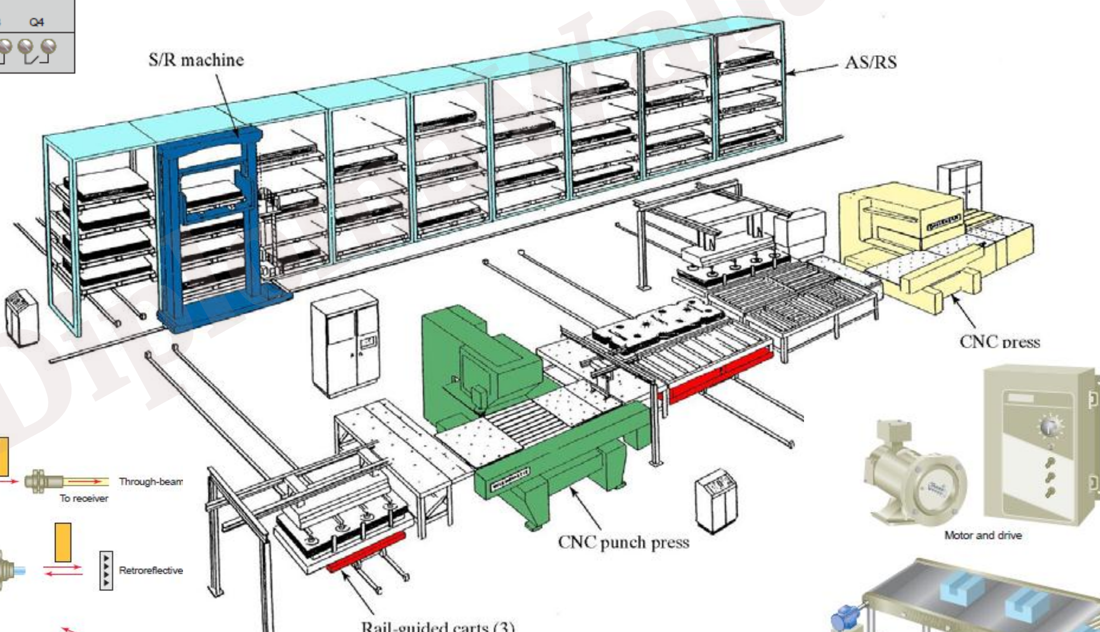
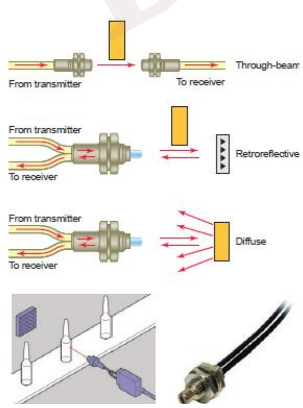
Jabatan Pengajian Politeknik

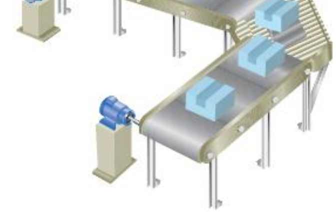
JM 608

INDUSTRIAL AUTOMATION









Industrial Automation: An Engineering Approach

Politeknik Port Dickson, 2013

CONTENTS

CHAPTER 3
MECHANICAL SYSTEM: COMPONENTS, DYNAMICS AND MODELING

CHAPTER 4
CONTROL OF ACTUATORS IN AUTOMATION MECHANISMS

CHAPTER 1
INTRODUCTION AND BASIC CONCEPT OF INDUSTRIAL AUTOMATION

Upon completion of this course, students should be able to:-

- Describes the definition and classification of automation in industry
- Explain the basic concept of automation terminology
- Explain the positioning concept of automation

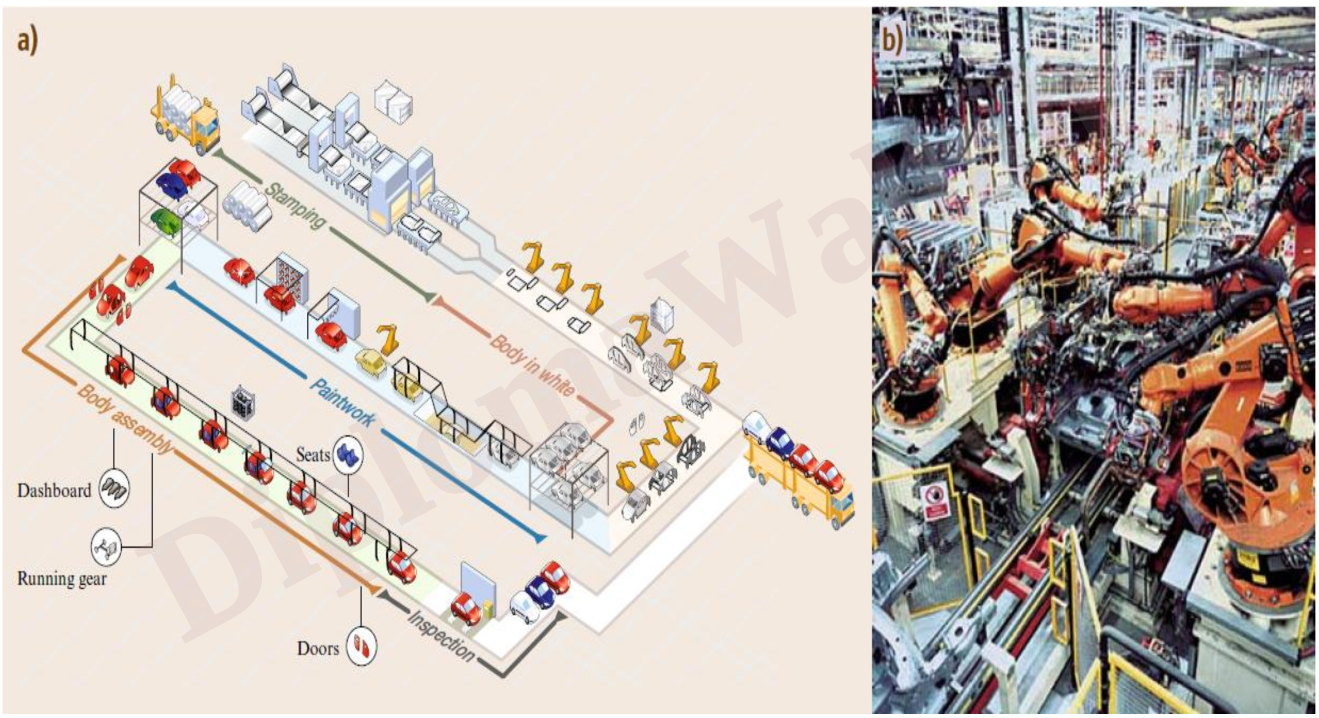


Figure 1.1 Car body assembly. (a) A car assembly usually follows the illustrated steps: Stamping of the metal sheet into plates, fixing and alignment of the plates on trays, spot welding, painting the car body and finally assembly of the car body (doors, dashboard, windscreens, power-train seats and tires). Car factories can host well over 1000 robots working two to three shifts per day. (b) The Mercedes A class assembly in Rastatt Germany is highly automated. The picture shows spot welding robots along the body in white transfer line.

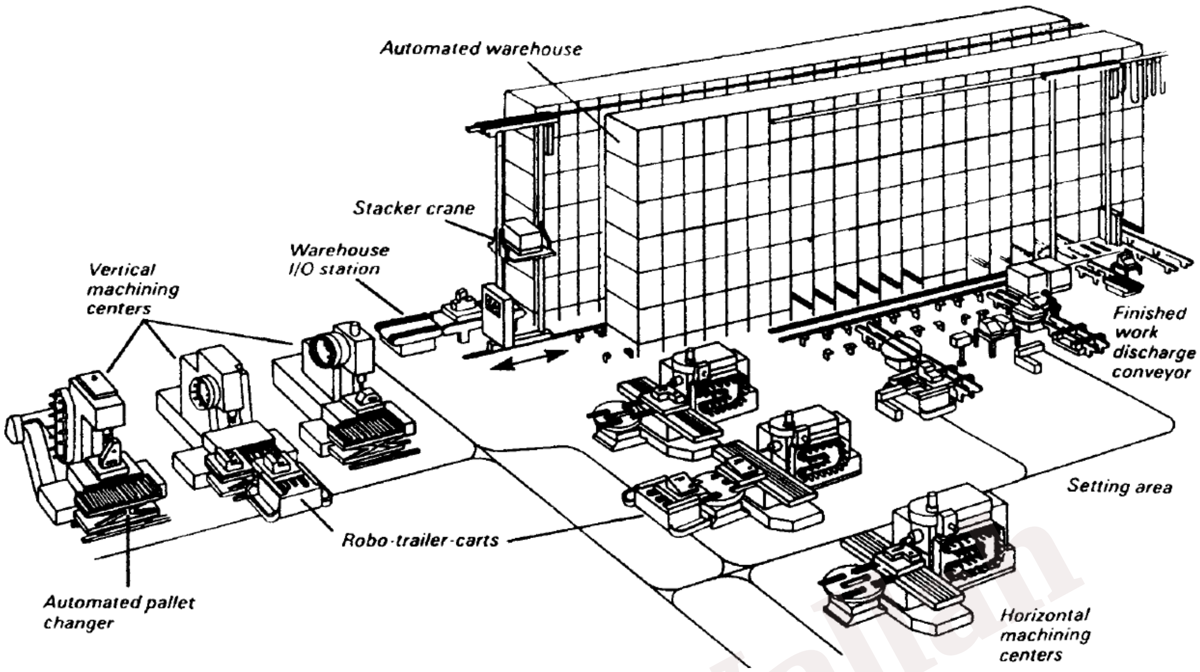


Figure 1.2 Conversion process in manufacturing Cell

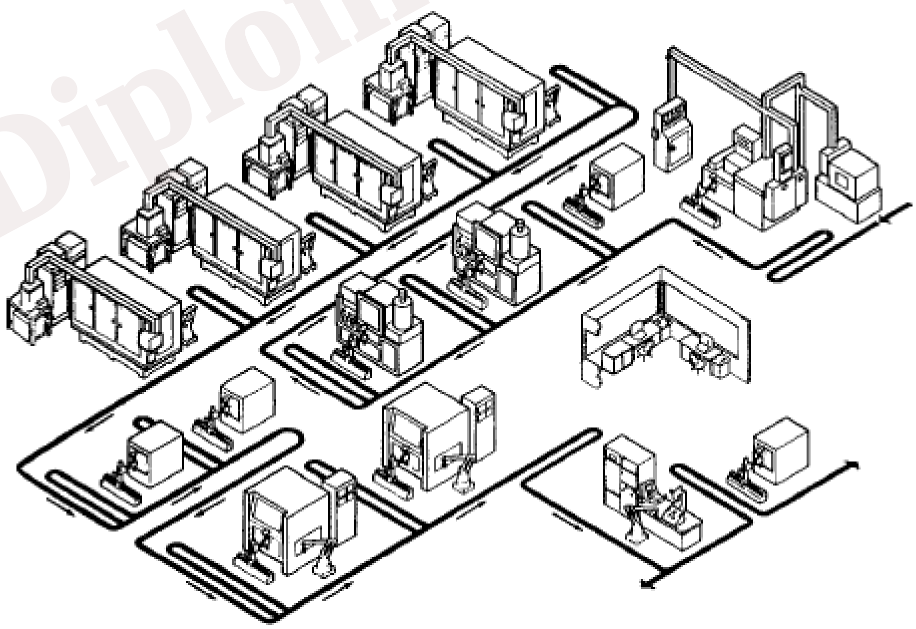


Figure 1.3 Workstation, work cell and work center

1.1 Introduction to Industrial Automation

As the global marketplace demands higher quality goods and lower costs, factory floor automation has been changing from separate machines with simple hardware-based controls, if any, to an integrated manufacturing enterprise with linked and sophisticated control and data systems. For many organizations the transformation has been gradual, starting with the introduction of programmable logic controllers and personal computers to machines and processes. However, for others the change has been rapid and is still accelerating.

There are two ways to achieve high yields in manufacturing. The simplest, yet most expensive way is to increase the number of production lines. An alternative and more desirable way is to increase the rate of production in the existing production lines. It is possible to increase the production rate by reducing the cycle time needed to produce a single part or product. There are also two ways to reduce cycle time. The first approach is to improve the manufacturing process. The second approach is to automate the manufacturing process by using re-programmable and automatically controlled equipment. This chapter discusses the type of automation and reason that make up industrial automation.

1.1.1 Definition of an Industrial Automation

Automation refers to a technology which based on the usage of mechanical, electronic and computer system in handling process and manufacturing process control. The usage of automation technology started when work done by labor / worker was started replace by machine. Technology development process continuous improve until human started introduce the usage of robotic, CAD/CAM, Flexible manufacturing system and others technology to increase human quality of life and increase productivity in the industrial.

1.1.1.1 Industrial

In a general sense the term "Industry" is defined as follows.

> ***Definition: Systematic Economic Activity that could be related to <u>Manufacture</u>/Service/ Trade.***

In this course, we shall be concerned with Manufacturing Industries only.

1.1.1.2 Automation

The word 'Automation' is derived from greek words "Auto"(self) and "Matos" (moving). Automation therefore is the mechanism for systems that "move by itself". However, apart from this original sense of the word, automated systems also achieve significantly superior performance than what is possible with manual systems, in terms of power, precision and speed of operation.

> **Definition:** ***Automation is a set of technologies that results in operation of machines and systems without significant human intervention and achieves performance superior to manual operation***

A Definition from Encyclopedia Britannica

The application of machines to tasks once performed by human beings or, increasingly, to tasks that would otherwise be impossible. Although the term mechanization is often used to refer to the simple replacement of human labor by machines, automation generally implies the integration of machines into a self-governing system.

1.1.2 Advantages for Automation

Companies undertake projects in manufacturing automation and computer integrated manufacturing for a variety of good reasons. Some of the reasons used to justify automated are the following:

1. *To increase labor productivity.* Automating a manufacturing operation usually increases production rate and labor productivity. This means greater output per hour of labor input.

2. To reduce labor cost. Ever-increasing tabor cost has been and continues to be the trend in the world's industrialized societies. Consequently, higher investment in automation has become economically justifiable to replace manual operations. Machines are increasingly being substituted for human lahar to reduce unit product cost.

3. *To migrate the effects of labor shortages.* There is a general shortage of labor in many advanced nations and this has stimulated the development of automated operations as a substitute tor labor.

4. To reduce or eliminate routine manual and clerical tasks. An argument can be put forth that there is social value in automating operations that are routine, boring, fatiguing, and possibly irksome. Automating such tasks serves a purpose of improving the general level of working conditions.

5. To improve worker safety. By automating a given operation and transferring the worker from active participation in the process to a supervisory role, the work is made safer. The safety and physical well-being of the worker has become a national objective with the' enactment of the Occupational Safety and Health Act (OSHA) in 1970. This has provided an impetus for automation.

6. To improve product quality. Automation not only results in higher production rates than manual operations; it also performs the manufacturing process with greater uniform and conformity to quality specifications. Reduction attraction defect rate is one of the chief benefits of automation.

7. To reduce manufacturing lead lime. Automation helps to reduce the elapsed time between customer order and product delivery, providing a competitive advantage 10 the manufacturer for future orders. By reducing manufacturing lead time, the manufacturer also reduces work-in-process inventory.

8. To accomplish processes that cannot be done manually. Certain operations cannot be accomplished without the aid of a machine. These processes have requirements for precision, miniaturization or complexity of geometry that cannot be achieved manually.
Examples include certain integrated circuit fabrication operations, rapid prototyping processes based on computer graphics (CAD) models, and the machining of complex, mathematically defined surfaces using computer numerical control. These processes can only be realized by computer controlled systems.

9. To avoid the high cost of not automating. There is a significant competitive advantage gained in automating a manufacturing plant. The advantage cannot easily be demonstrated on a company's project authorization form. The benefits of automation often show up in unexpected and intangible ways, such as in improved quality, higher sales, better labor relations, and better company image. Companies that do not automate are likely to find themselves at a competitive disadvantage with their customers, their employees, and the general public

1.1.3 Disadvantages of Automation

Aside from these advantages, it is also important for us to discuss about the disadvantages of using and implementing automation in the industrial.

1. Higher Start-up cost and the cost of operation. Automated equipment includes the high capital expenditure required to invest in automation. An automated system can cost millions of dollars to design, fabricate, and install.

2. *Higher Cost of Maintenance*. A higher level of maintenance needed than with a manually operated machine. These include buying electromechanical devices such as electromechanically valve, sensory devices, and smart devices. Cost of spare parts for automation system may consider higher compare to the manual operate.

3. *Obsolescence/Depreciation Cost.* Obsolescence and depreciation is a gradual reduction in the value of physical assets. This phenomenon is characteristic of all physical assets in the form of equipment and machinery. It was something that was inevitable due to technology development. Obsolescence or depreciation can be classified into two parts, namely: -

i. Physical Depreciation - occurred as a result of physical damage of equipment or robots. It describes a form that can be seen clearly as damage, wear and corrosion.
ii. Depreciation of the functions - it existed from changes in demand for services may be provided. Depreciation caused by changes in the need for an equipment service discovery of new equipment or a robot system inability to meet demand

4. *Unemployment.* A disadvantage often associated with automation, is worker displacement. Due to the fact that manual laborers are being replaced by robots or other automated machineries, this results to mass lay-off. A lot of people are losing their jobs especially those who work in the manufacturing industry such as a car factory.

5. Not economically justifiable for small scale production.

1.1.4 Types of Automation System

Automated manufacturing systems can be classified into three basic types:

i. Fixed automation.
ii. Programmable automation, and
iii. Flexible automation.

1.1.4.1 Fixed Automation

Fixed automation is a system in which the sequence of processing (or assembly) operations is fixed by the equipment configuration. Each of the operations is the sequence is usually simple, involving perhaps a plain linear or rotational motion or an uncomplicated combination of the two; for example, the feeding of a rotating spindle. It is the integration and coordination of many such operations into one piece of equipment that makes the system complex. Typical features of fixed automation are:

i. high initial investment for custom-engineered equipment
ii. high production rates
iii. relatively inflexible in accommodating product variety

The economic justification for fixed automation is found in products *that* are produced in very large quantities and at high production rates. The high initial cost of the equipment can be spread over a very large number of units, thus making the unit cost attractive compared with alternative methods of production. Examples of fixed automation include machining transfer lines, automated assembly machines, distillation process, conveyors and paint shops.

1.1.4.2 Programmable Automation

In *programmable automation,* the production equipment is designed with the capability to change the sequence of operations to accommodate different product configuration. The operation sequence is controlled by a *program,* which is a set of instructions coded so that they can be read and interpreted by the system. New programs can be prepared and entered into the equipment to produce new products. Some of the features that characterize programmable automation include:

i. high investment in general purpose equipment
ii. lower production rates than fixed automation
iii. flexibility to deal with variations and changes in product configuration
iv. most suitable for batch production

Programmable automated production systems are used in low- and medium-volume production. The parts or products are typically made in batches. To produce each new batch of a different product, the system must be reprogrammed with the set of machine instructions that correspond to the new product. The physical setup of the machine must also be changed. Tools must be loaded, fixtures must be attached to the machine table and the required machine setting must be entered. This changeover procedure takes time. Consequently, the typical cycle for a given product includes a period during which the setup and reprogramming takes place, followed by a period in which the batch is produced. Examples of programmable automation include numerically controlled (NC) machine tools, industrial robots, programmable logic controller, steel rolling Mills and paper mills.

1.1.4.3 Flexible Automation

Flexible automation is an extension of programmable automation. A flexible automated system is capable of producing a variety of parts (or products) with virtually no time lost for changeovers from one part style to the next. There is no lost production time while reprogramming the system and altering the physical setup (tooling, fixtures, machine settings). Consequently, the system can produce various combinations and schedules of parts or products instead of requiring that they be made in batches. What makes flexible automation possible is that the differences between parts processed by the system are not significant. It is a case of soft variety.so that the amount of changeover required between styles is minimal. The features of flexible automation can be summarized as follows:

i. high investment for a custom-engineered system
ii. continuous production of variable mixtures of products
iii. medium production rate
iv. flexibility to deal with product design variations

Examples of flexible automation are the flexible manufacturing systems for performing machining operations that date back to the late 1960s. The relative positions of the three types of automation for different production volumes and product varieties are depicted in Figure 1.4. For low production quantities and new product introduction manual production is competitive with programmable automation, as we indicate in the figure.

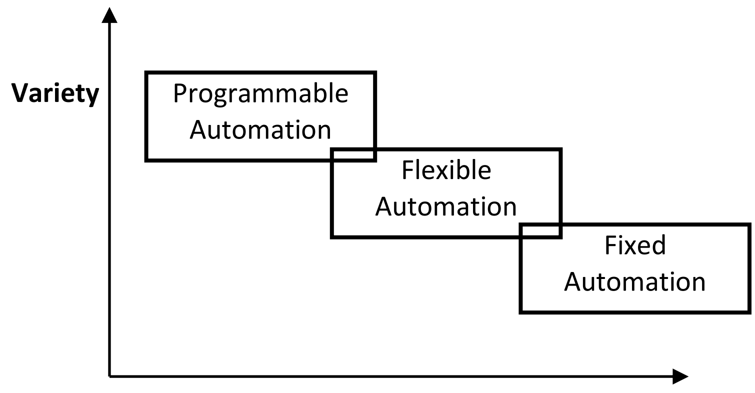


Figure 1.4: Three types of automation relative to production quantity and product variety.

1.1.5 Describe the Automation in Production System

A production system is a collection of people, equipment, and procedures organized to perform the manufacturing operations of an organization. A production system consists of facilities and manufacturing support systems (Figure 1.5):

i. Facilities—the factory, the equipment in the factory, and the way the equipment is organized around the shop floor.
ii. Manufacturing support systems—the set of procedures used to manage production and to solve technical and logistics problems met in manufacturing processes. These systems include product design, planning and control, logistics and other business functions.

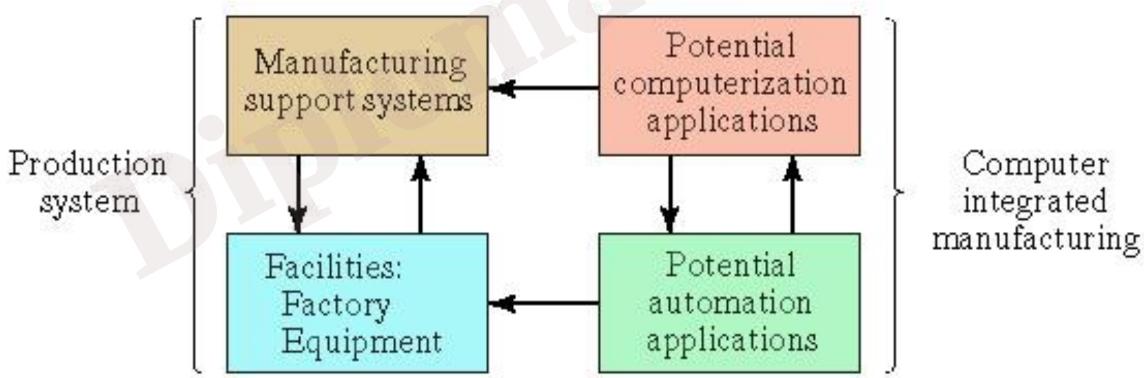


Figure 1.5: Production System consists of facilities and manufacturing support systems

A manufacturing system is a logical grouping of equipment in the factory and the workers who operate it. Examples include worker-machine systems, production lines, and machine cells. A production system is a larger system that includes a collection of manufacturing systems and the support systems used to manage them. A manufacturing system is a subset of the production system. Portions of production systems tend to be automated and/or computerized, while other parts may be operated by manual labor (see Figure 1.6). The overall operation of the production system is controlled by people, including direct labor staff for facility operation, and professional staff with responsibilities over the manufacturing support systems. Facilities include the factory, production machines and tooling, material handling equipment, inspection equipment, and computer systems that control the manufacturing operations. Facilities can also include the plant layout—that is, the physical arrangement of the equipment in the factory, which is usually organized into logical groupings called manufacturing systems.

Figure 1.6: View of manufacturing cells within a production system

Manufacturing systems consist of groups of machines and associated workers. Typically, the manufacturing system comes in direct physical contact with the product or parts to be made. Three types may be identified, as outlined in Table 1.1.

Table 1.1: Three categories of manufacturing systems

Category	Description
Manual Work System	One (or more) workers performing one (or more) tasks without powered tools. Typical example is the material handling task. In production tasks the use of hand tools is pre-dominant, sometimes with optional work-holder. Examples include: filing milled parts; checking quality of parts with micrometer; moving cartons using a dolly; and, assembling machinery using hand tools.
Worker-Machine Systems	A human worker operates powered equipment, in various combinations of one (or more) workers, and one (or more) pieces of equipment. Relative strengths of humans and machines are combined. Examples include: machinist operating engine lathe; a fitter working with an industrial robot; a crew of workers operating a rolling mill; and personnel performing work on a mechanized conveyor.
Automated Systems	Process is performed by machine without the direct participation of a human worker. Automation uses a programmed of instructions and a control system for implementation; there are two sub-categories: semi-automated, and fully automated. Semi-automation implies only part of the work cycle is completely automated, with other work done by a human worker. A fully automated machine, on the other hand, has the capacity to operate for extended periods of time (longer than one work cycle) with no human interaction. However, although fully automated, human monitoring may still be used. Examples include: injection moulding machines; and automated processes in oil refineries and nuclear power plants.

Manufacturing Support Systems are used by a company to manage its production operations. Most support systems do not directly contact the product, but they plan and control its progress through the factory. Manufacturing Support Systems use a four-function information-processing cycle that is explained in Table 1.2. This list of functions is activated by customers' orders, which propels the system into action, and operates by deploying the facilities detailed above.

Table 1.2: Information-processing cycle for Manufacturing Support Systems

Function	Description
Business Function	First and last phase. The principle means of communicating with the customer, includes sales and marketing, sales forecasting, order entry, cost accounting, and customer billing. Product originates from customer order, and after sales and marketing, proceeds to become a production order. Production order is in the form of one of the following: a manufacturing order against customer specifications; a customer order to buy one or more of manufacturer's proprietary products; or, an internal company order based on future-demand Forecasts.
Product Design	Second phase. If product is manufactured by customer design, then design supplied by customer. If there are customer specifications, then manufacturer's design department may be contracted to create a design on this basis, as well as to manufacture the product also. For a proprietary product, the manufacturing firm is responsible for its development and design.
Manufacturing Planning	Third phase. Upon completion of product design, the associated information is given to the manufacturing planning function. Process planning, master scheduling, requirements planning, and capacity planning are performed here. Process planning determines the process and assembly steps, and the order of the steps, needed to produce the product. The master production schedule lists products to be made, when they are to be made, and the quantities of each to be produced. Based upon the master production schedule, requirements planning are performed—that is, the individual components, sub-assemblies, raw materials etc. required are purchased, created, and scheduled to be available when needed. Capacity planning is concerned with planning the manpower and machine resources to carry out the manufacturing function.
Manufacturing Control	Fourth phase. Concerned with managing and controlling the physical operations in the factory to implement the manufacturing plans. Shop floor control, inventory control, and quality control are performed here. Shop floor control monitors the product as it moves about the shop floor; as the product is a work-in-process inventory as it proceeds across the shop floor, shop-floor control is related to inventory control also. Inventory control tries to maintain the correct amount of inventory in the manufacturing system, and avoid overloading or starving the system. Quality control tries to ensure correct product and component quality, as per the specified design. It uses inspection activities on the shop-floor, and at the point of entry of outsourced components, to do this.

Automated manufacturing systems operate in the factory on the physical product. They perform operations such as processing, assembly, inspection, or material handling, in some cases accomplishing more than one of these operations in the same system. They are called automated because they perform their operations with a reduced level of human participation compared with the corresponding manual process. In some highly automated systems, there is virtually no human participation. Examples of automated manufacturing systems include:

i. Automated machine tools that process parts
ii. Transfer lines that perform a series of machining operations
iii. Automated assembly systems
iv. Manufacturing systems that use industrial robots to perform processing or assembly operations
v. Industrial Robots
vi. Automatic material handling and storage systems to integrate manufacturing operations
vii. Automatic inspection systems for quality control

Generally, there are two types of production system automation: automation of the manufacturing systems (Figure 1.7), and computerization of the manufacturing support systems. Since automation of the manufacturing systems consists of some computerization for control and operational purposes, the two types tend to overlap. The term computer-integrated manufacturing is used to indicate this extensive use of computers in production systems (see Figure 1.8).

- Storage Systems
- Handling Systems
- Assembly Lines
  - Assembly Cells
  - Machines
    - Actuators
    - Sensors
- Production Lines
  - Production Cells
  - Machines
    - Actuators
    - Sensors

Figure 1.7: Automation of the manufacturing systems

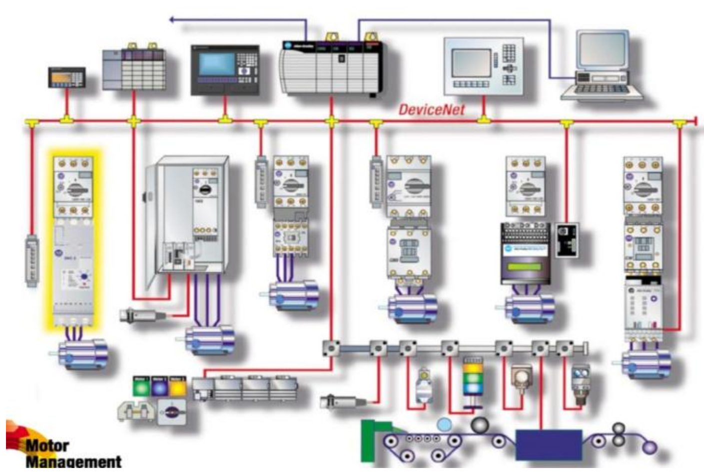


Figure 1.8: Computer equipment, sensors and actuators used in production

1.1.5.1 Industrial Automation and Robotic

Robotic is the main component in the industrial automation. Industrial work situations that promote the substitution of robots for human labor include: hazardous work environments; repetitive work cycles; difficult handling; multi-shift environments; infrequent changeovers; and part position and orientation are established in the work cell i.e. repetitive tasks. Robots are being used mainly in three types of applications: material handling; processing operations; and assembly and inspection. In material handling robots move parts between various locations by means of a gripper type end effector. Two sub-divisions may be noted in material handling: material transfer; and machine loading and/or unloading (Figure 1.9).

Figure 1.9: Robot in material handling

1.2 Basic Concept of Automation Terminology

There is a set of basic terminology and concepts common to all robots. These terms follow with brief explanations of each.

1.2.1 Links and Joints

Links are the solid structural members of a robot, and joints are the movable couplings between them. Joints or axes found in the manipulator (robotic arm). *A joint* of an industrial robot (Figure 1.10) is similar to a joint in the human body. It provides relative motion between two parts of the body. Joints consists of two types, major axis comprising the base, shoulder and elbow and minor axis comprising wrist pitch, wrist roll and wrist yaw.

Connected to each joint are two links (Figure 1.11), an input link and an output link. *Links* are the rigid components of the robot manipulator. The purpose of the joint is to provide controlled relative movement between the input link and the output link.

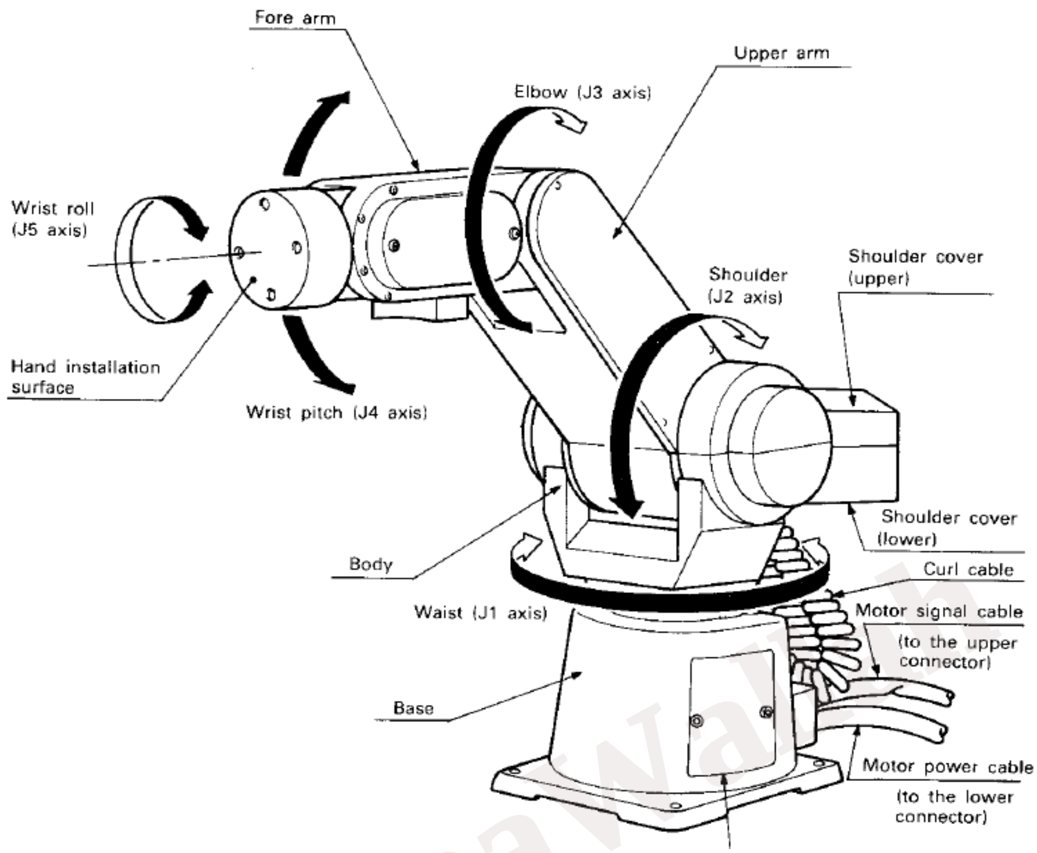


Figure 1.10: A Mitsubishi RV-M1 Industrial Robot

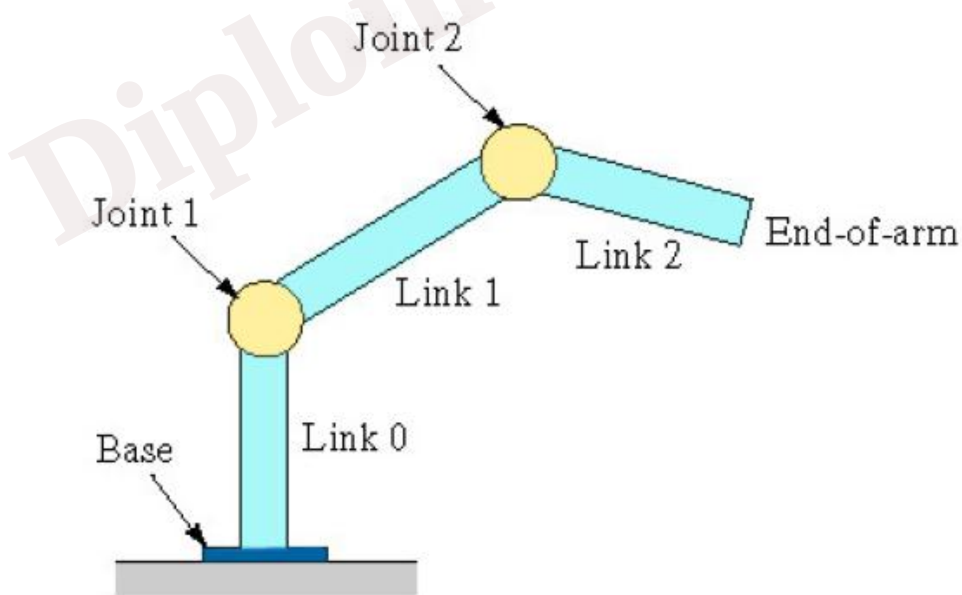
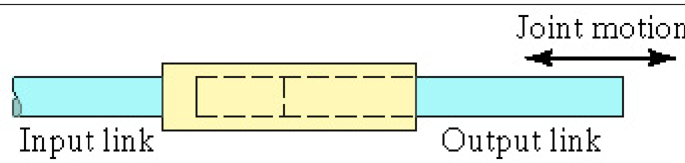
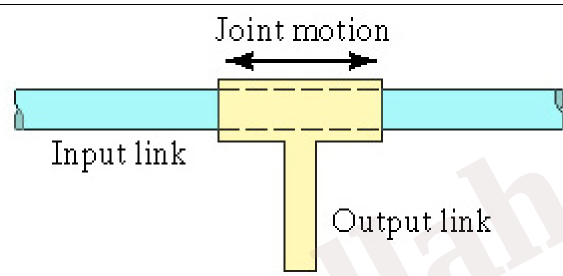
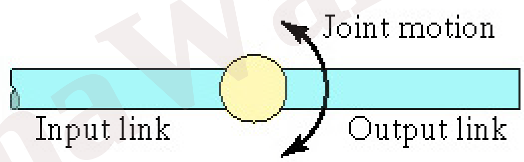
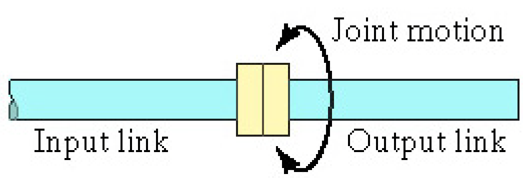
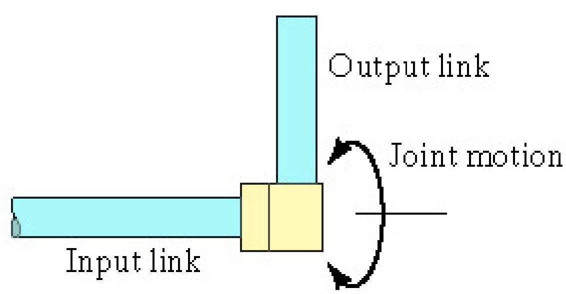


Figure 1.11: Diagram of robot construction showing how a robot is made up of a series of joint-link combinations

Five types of mechanical joints for robots may be classified; these are outlined in Table 1.3.

Table 1.3: Mechanical Joints for Robots

Joint	Description	Schematic
Linear joint	Type L joint; the relative movement between the input link and the output link is a translational sliding motion, with the axes of the two links parallel.	
Orthogonal joint	Type O joint; the relative movement between the input link and the output link is a translational sliding motion, but the output link is perpendicular to the input link.	
Rotational joint	Type R joint; this provides rotational relative motion, with the axis of rotation perpendicular to the axes of the input and output links.	
Twisting joint	Type T joint; this provides rotary motion, but the axis of rotation is parallel to the axes of the two links.	
Revolving joint	Type V joint; the axis of the input link is parallel to the axis of rotation of the joint, and the axis of the output link is perpendicular to the axis of rotation.	

1.2.2 Degree of freedom (DOF)

Each joint on the robot introduces a degree of freedom. Each dof can be a slider, rotary, or other type of actuator. Robots typically have 5 or 6 degrees of freedom. 3 of the degrees of freedom allow positioning in 3D space, while the other 2or 3 are used for orientation of the end effector. 6 degrees of freedom are enough to allow the robot to reach all positions and orientations in 3D space. 5 dof requires a restriction to 2D space, or else it limits orientations. 5 dof robots are commonly used for handling tools such as arc welders.

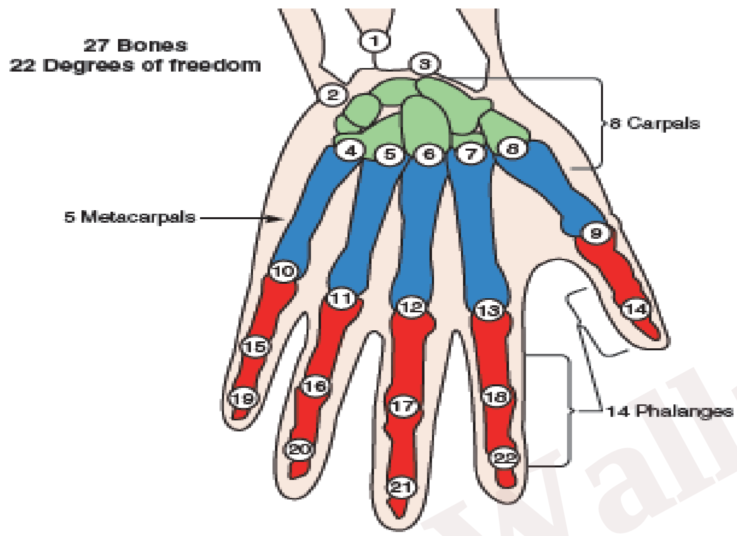


Figure 1.12: Each joint represents a degree of freedom; there are 22 joints and thus 22 degree of freedom in the human hand

From the Figure 1.13 we can see that the robot has five degrees of freedom. That means we can have it move in five independent ways. The five different movements are created in five different joints as described below.

1. Base Joint: This joint allows movement of 350° rotational motion.
2. Shoulder Joint: This joint allows movement of 120° rotational motion.
3. Elbow Joint: This joint allows movement of 135° rotational motion.
4. Wrist Joint: This joint allows movement of 340° rotational motion.
5. Gripper: This joint allows movement of 2 linear motions (open and close actions).

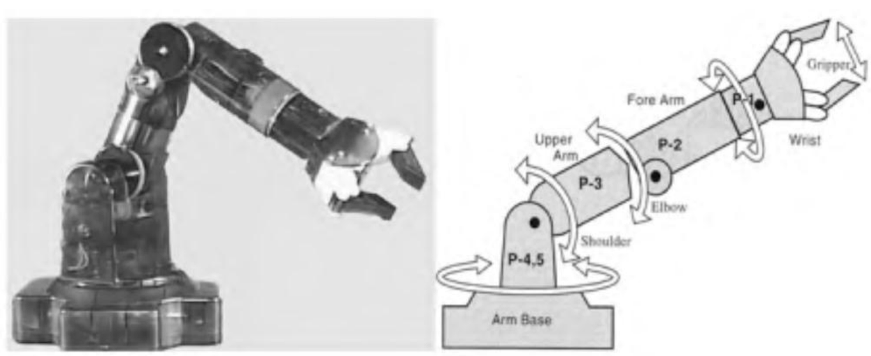


Figure 1.13: OWI-007 robotic arm trainer (Copyright OWI Robots)

1.2.3 Orientation Axes

Basically, if the tool is held at a fixed position, the orientation determines which direction it can be pointed in. Roll, pitch and yaw are the common orientation axes used. Looking at the figure below it will be obvious that the tool can be positioned at any orientation in space. (Imagine sitting in a plane. If the plane rolls you will turn upside down. The pitch changes for takeoff and landing and when flying in a crosswind the plane will yaw.)

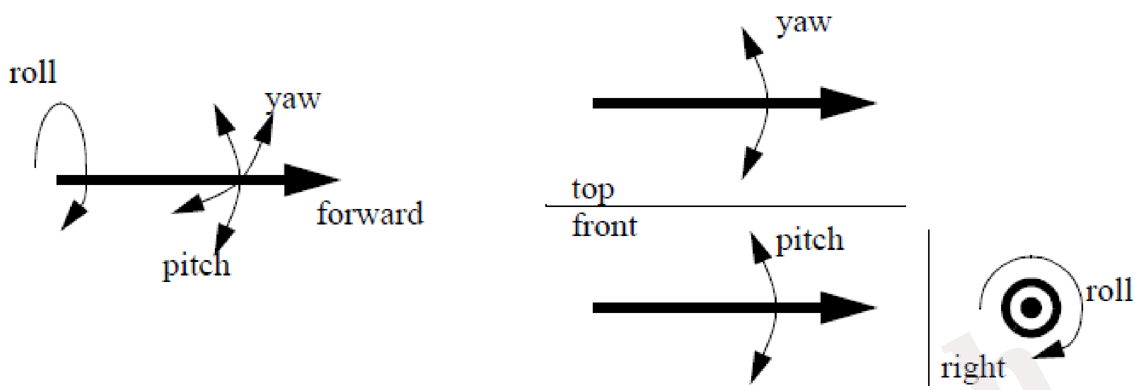


Figure 1.14: Orientation Axes

1.2.4 Position Axes

The tool, regardless of orientation, can be moved to a number of positions in space. Various robot geometries are suited to different work geometries. (more later) The definition of an object's location in 3-D space, usually defined by a 3-D coordinate system using X, Y, and Z coordinates. Part of a robot can move to a spot within its work envelope, using devices that tell it exactly where it is. Translatory degrees of freedom.

1.2.5 Tool Centre Point (TCP)

The tool center point is located either on the robot, or the tool. Typically the TCP is used when referring to the robots position, as well as the focal point of the tool. (e.g. the TCP could be at the tip of a welding torch) The TCP can be specified in Cartesian, cylindrical, spherical, etc. coordinates depending on the robot. As tools are changed we will often reprogram the robot for the TCP.Tool midpoint is the reference point of tools controlled by the robot.

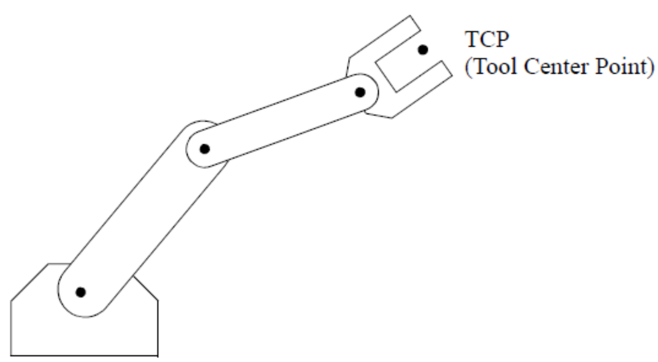


Figure 1.15: Tool Centre Point (TCP)

1.2.6 Work envelope/workspace

Work envelope is the volume/area where the robotic arm can perform task/work (Figure 1.16 & Figure 1.17). The space in which a robot can operate is its work envelope, which encloses its workspace. While the workspace of the robot defines positions and orientations that it can achieve to accomplish a task, the work envelope also includes the volume of space the robot itself occupies as it moves. This envelope is defined by the types of joints, their range of movement and the lengths of the links that connect them. The physical size of this envelope and the loads on the robot within this envelope are of primary consideration in the design of the mechanical structure of a robot.

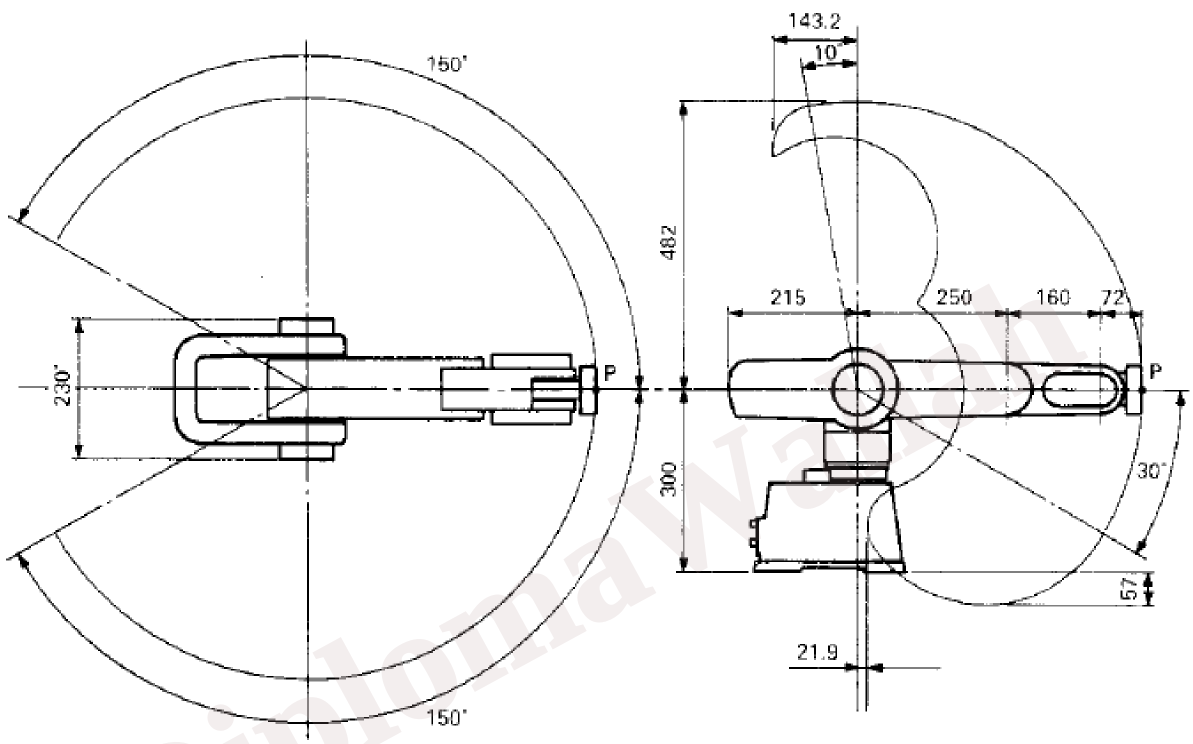


Figure 1.16: Mitsubishi RV-M1's work envelope

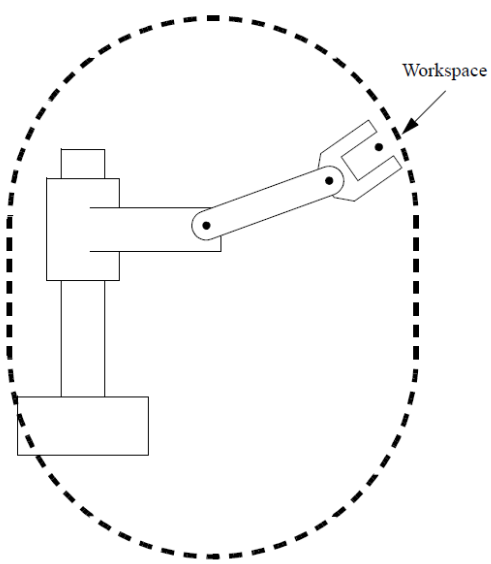


Figure 1.17: Work envelope/workspace

1.2.7 Speed

It's refers either to the maximum velocity that is achievable by the TCP, or by individual joints. This number is not accurate in most robots, and will vary over the workspace as the geometry of the robot changes (and hence the dynamic effects). The number will often reflect the maximum safest speed possible. Some robots allow the maximum rated speed (100%) to be passed, but it should be done with great care. Speed is the rate of movement from point to point done by robots under the control of the program. It is a measure of the speed of the device. Maximum joint velocity (angular or linear) is not an independent value. For longer motions it is often limited by servomotor bus voltage or maximum allowable motor speed. For manipulators with high accelerations, even short point-to-point motions may be velocity limited. For low-acceleration robots, only gross motions will be velocity limited. Typical peak end-effector speeds can range up to 20m/s for large robots.

1.2.8 Payload

The payload indicates the maximum mass the robot can lift before either failure of the robots, or dramatic loss of accuracy. It is possible to exceed the maximum payload, and still have the robot operate, but this is not advised. When the robot is accelerating fast, the payload should be less than the maximum mass. This is affected by the ability to firmly grip the part, as well as the robot structure, and the actuators. The end of arm tooling should be considered part of the payload. Maximum payload is specified in kilograms.

1.2.9 Repeatability

The degree ability of a robotic arm to detect targets has been set correctly and then returns to its original point in the work cell. The robot has a high repeatability will be able to repeat the task with the right repeatedly without error. The robot mechanism will have some natural variance in it. This means that when the robot is repeatedly instructed to return to the same point, it will not always stop at the same position. Repeatability is considered to be +/-3 times the standard deviation of the position, or where 99.5% of all repeatability measurements fall. This figure will vary over the workspace, especially near the boundaries of the workspace, but manufacturers will give a single value in specifications.

1.2.10 Accuracy

The degree of ability that can be made by a robotic arm to move to a certain point in the work cell as we enter the coordinates in the off-line programming station (off-line programming). This is determined by the resolution of the workspace. If the robot is commanded to travel to a point in space, it will often be off by some amount, the maximum distance should be considered the accuracy. This is an effect of a control system that is not necessarily continuous (Figure 1.18).

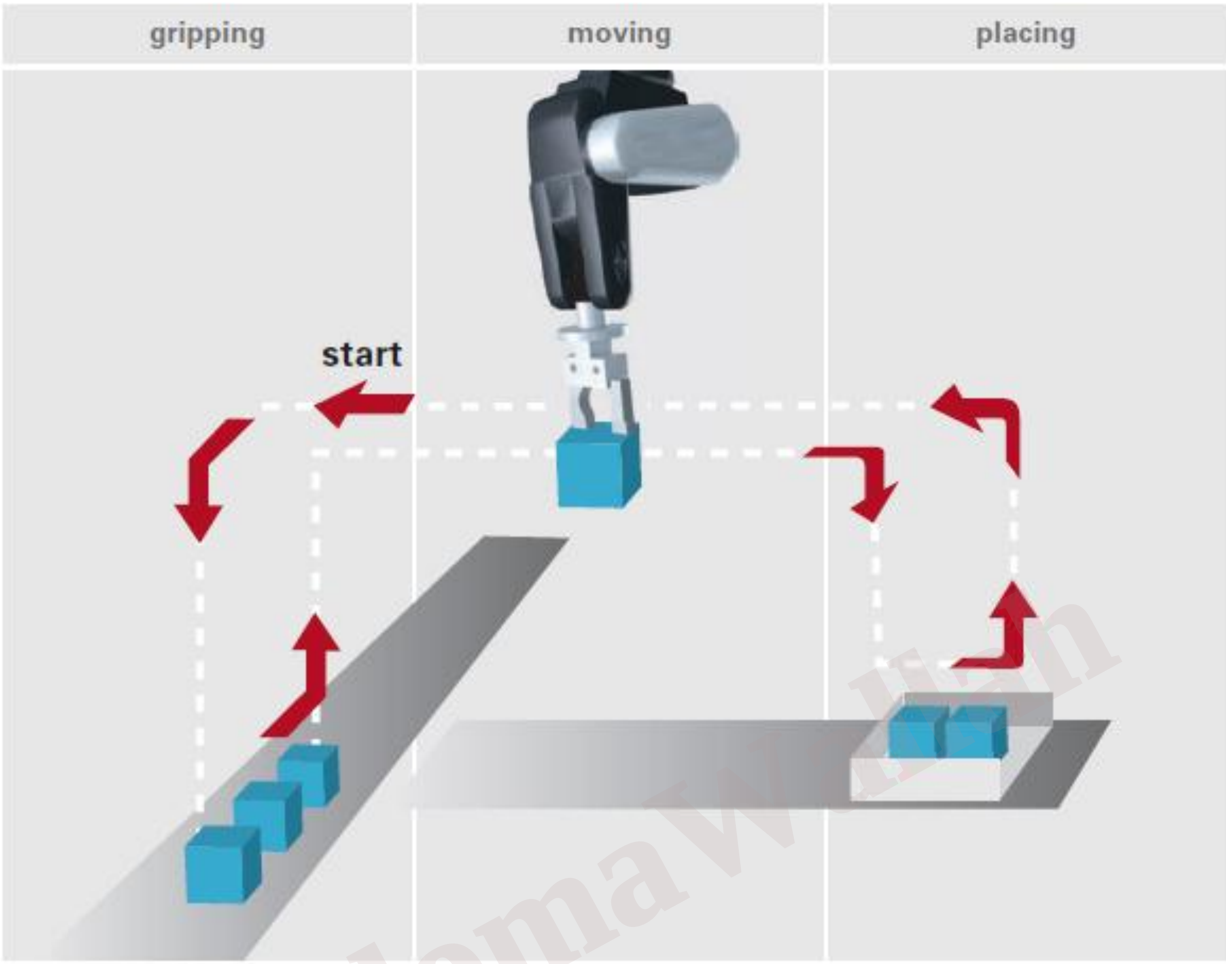


Figure 1.18: Accuracy is determined by the resolution of the workspace

1.2.11 Settling Time

During a movement, the robot moves fast, but as the robot approaches the final position is slows down, and slowly approaches. The settling time is the time required for the robot to be within a given distance from the final position. The time-instant when the actual output converges to the desired output is known as the *settling time.*

1.2.12 Control Resolution

This is the smallest change that can be measured by the feedback sensors, or caused by the actuators, whichever is larger. If a rotary joint has an encoder that measures every 0.01 degree of rotation, and a direct drive servo motor is used to drive the joint, with a resolution of 0.5 degrees, then the control resolution is about 0.5 degrees (the worst case can be 0.5+0.01). Capability of robot's positioning system to divide the motion range of each joint into closely spaced points.

1.2.13 Coordinates

The robot can move, therefore it is necessary to define positions. Note that coordinates are a combination of both the position of the origin and orientation of the axes. Points are programmed in the cells identified job position by using the values of the coordinates x, y and z of the tools midpoint and extension angles at the wrist axis robot arm is pitch, roll and yaw.

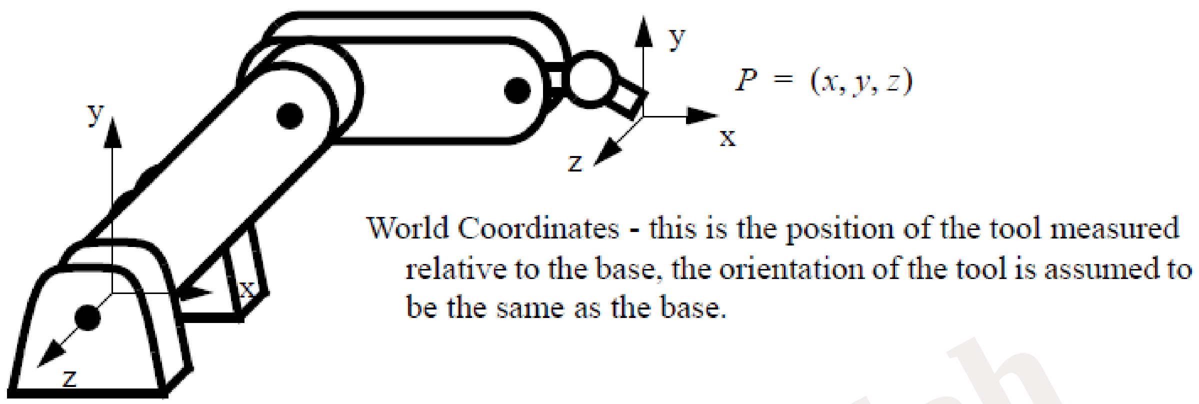


Figure 1.19: World Coordinates – To Locate the TCP

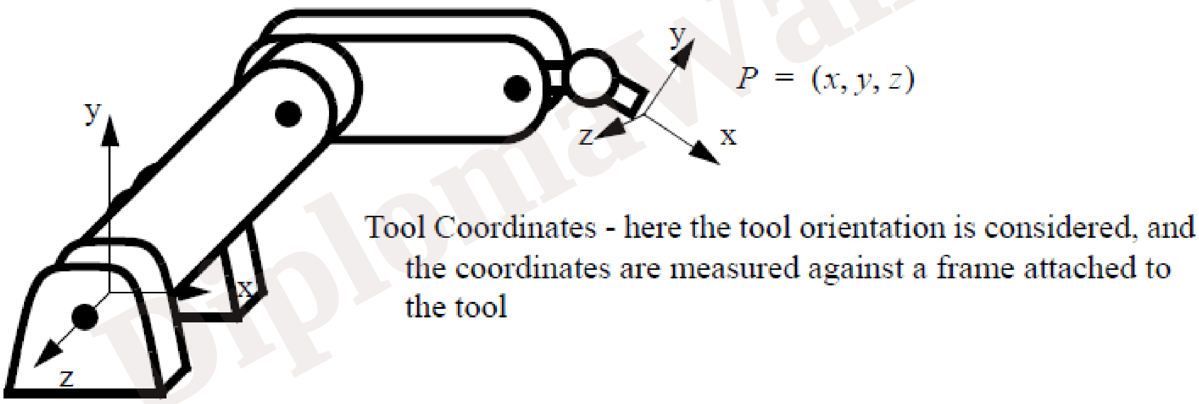


Figure 1.20: Tools Coordinates – describing Positions relative to the Tool

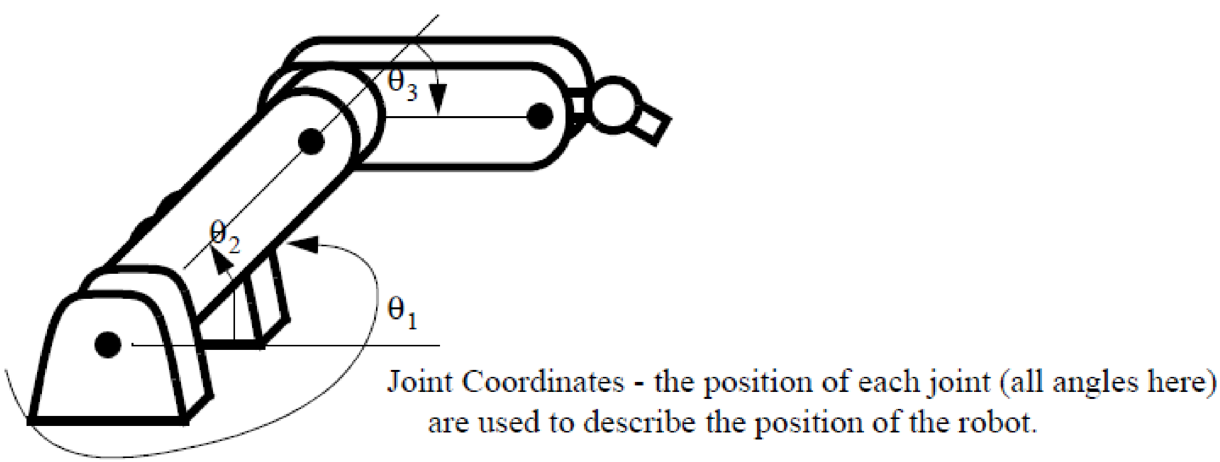


Figure 1.21: Joint Coordinates – the positions of the Actuators

1.3 Positioning Concept of Automation

1.3.1 Accuracy and Repeatability

The accuracy and repeatability are functions of,

i. Resolution- the use of digital systems, and other factors mean that only a limited number of positions are available. Thus user input coordinates are often adjusted to the nearest discrete position.

ii. Kinematic modeling error - the kinematic model of the robot does not exactly match the robot. As a result the calculations of required joint angles contain a small error.

iii. Calibration errors - The position determined during calibration may be off slightly, resulting in an error in calculated position.

iv. Random errors - problems arise as the robot operates. For example, friction, structural bending, thermal expansion, backlash/slip in transmissions, etc. can cause variations in position.

1.3.1.1 Accuracy

"How close does the robot get to the desired point". This measures the distance between the specified position, and the actual position of the robot end effector. Accuracy is more important when performing off-line programming, because absolute coordinates are used.

1.3.1.2 Repeatability

"How close will the robot be to the same position as the same move made before". It is a measure of the error or variability when repeatedly reaching for a single position. This is the result of random errors only. Repeatability is often smaller than accuracy.

Resolution is based on a limited number of points that the robot can be commanded to reach for these are shown here as black dots. These points are typically separated by a millimeter or less, depending on the type of robot. This is further complicated by the fact that the user might ask for a position such as 456.4mm, and the system can only move to the nearest millimeter, 456mm, this is the accuracy error of 0.4mm.

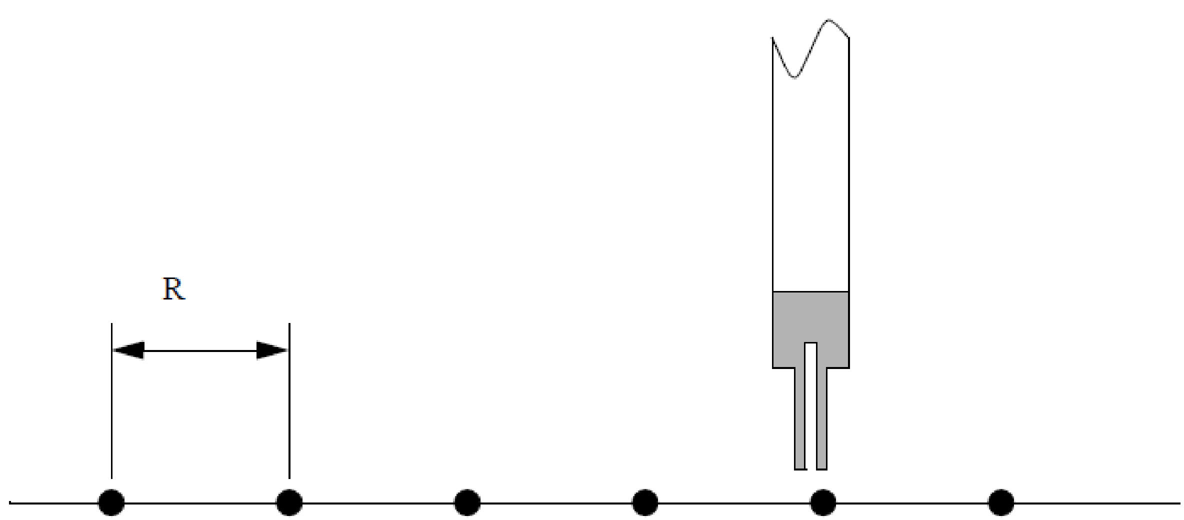


Figure 1.22: In a perfect mechanical situation the accuracy and control resolution would be determined as below,

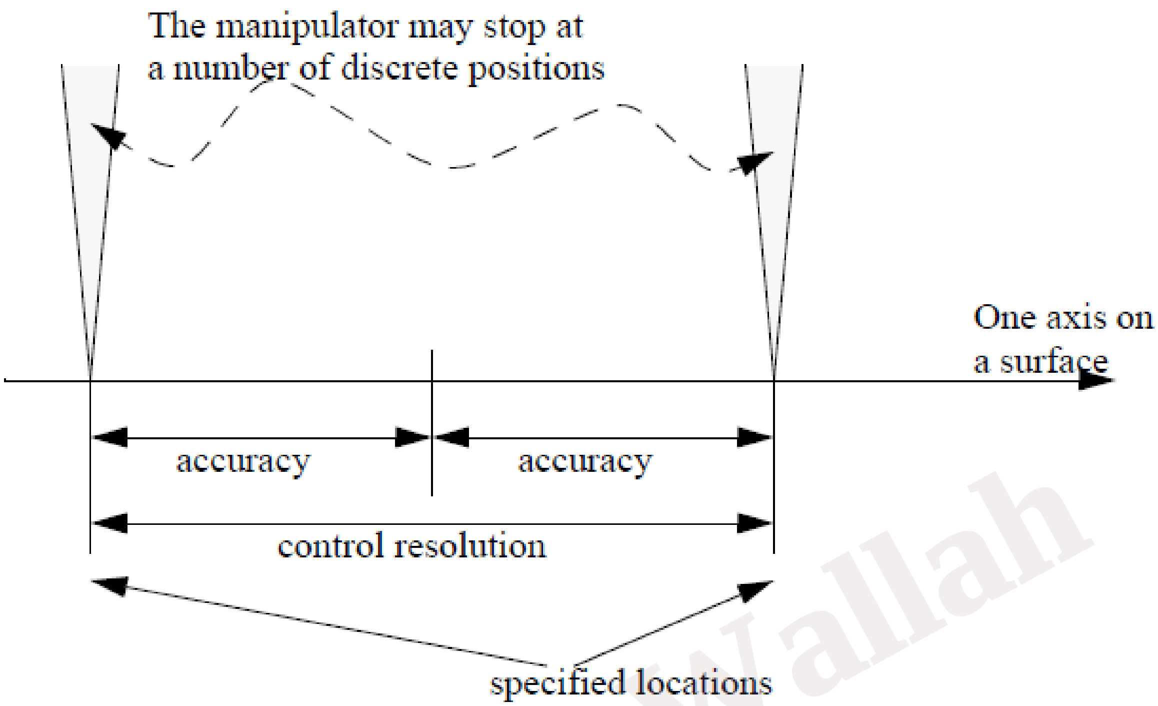


Figure 1.23: Kinematic and calibration errors basically shift the points in the workspace resulting in an error 'e'. Typically vendor specifications assume that calibration and modeling errors are zero.

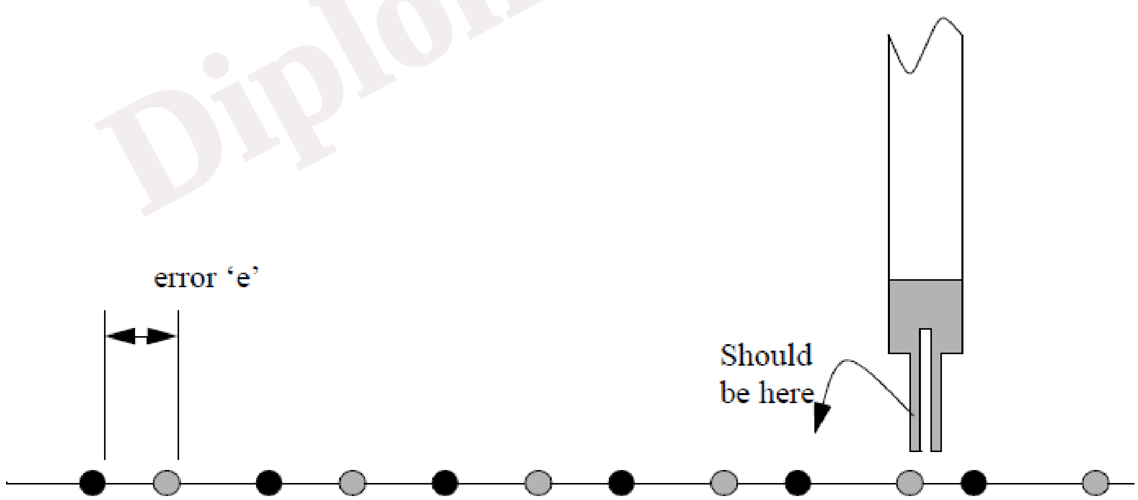
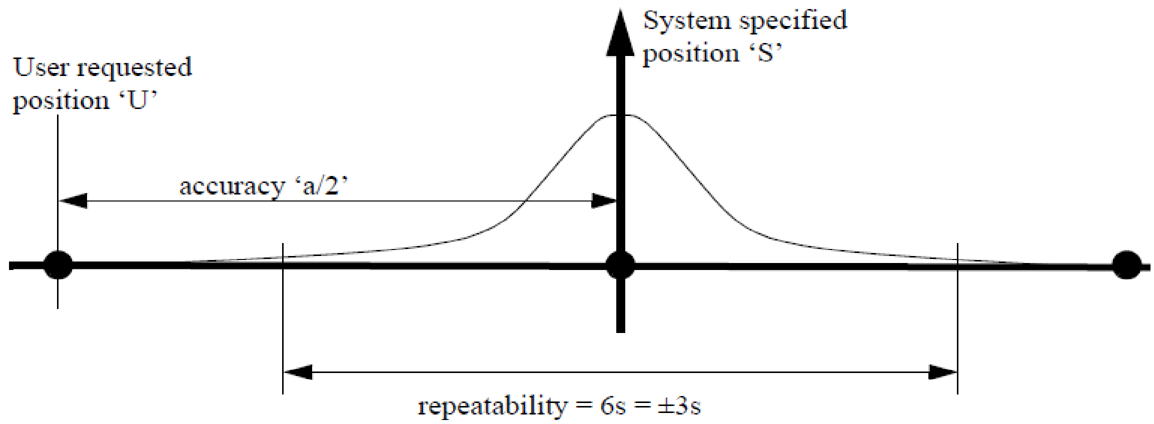
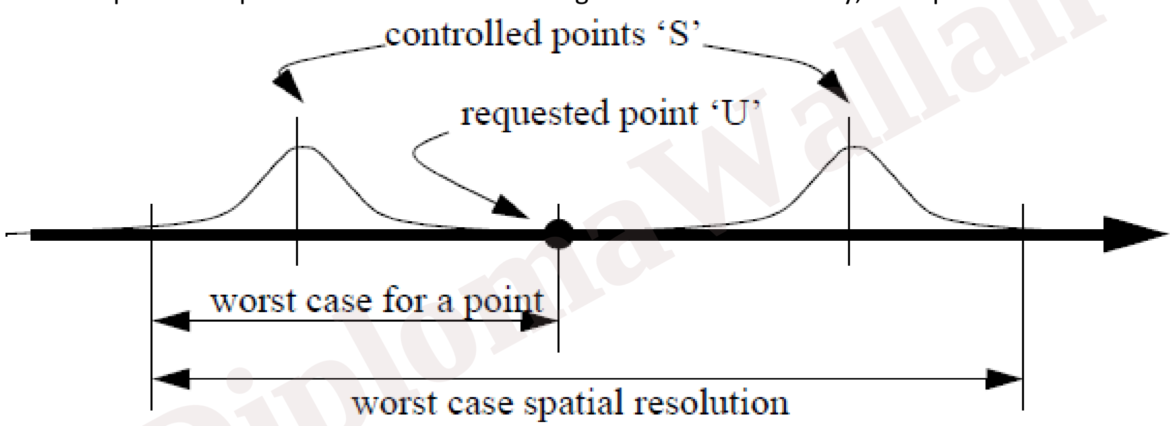


Figure 1.24: Random errors will prevent the robot from returning to the exact same location each time and this can be shown with a probability distribution about each point.

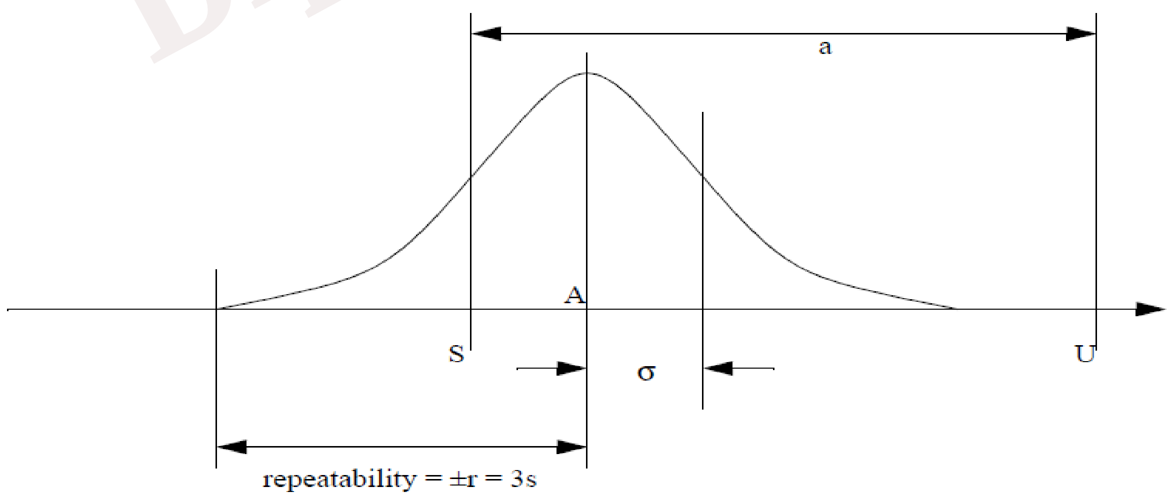


$$a = \frac{\text{control resolution}}{2} \qquad e_{max} = a + \text{modeling error} + 3s$$

Figure 1.25: If the distribution is normal, the limits for repeatability are typically chosen as ±3 standard deviations 's'. We can look at distributions for each specified position for the robot end effector in relationship to other point distributions. This will give us overall accuracy, and spatial resolution.



The fundamental calculations are,



$$accuracy = (S - U) + e$$

1.3.2 Control Resolution

Spatial resolution is the smallest increment of movement into which the robot can divide its work volume. Spatial resolution depends on two factors: the systems control resolution and the robots mechanical inaccuracies. It is easiest to conceptualize these factors in terms of a robot with 1 degree of freedom.

Control resolution - is determined by the robot's position control system and its feedback measurement system. It is the controller's ability to divide the total range of movement for the particular joint into individual increments that can be addressed in the controller. The increments are sometimes referred to as "addressable parts". The ability to divide the joint range into increments depends on the bit storage capacity in the control memory. The number of separate, identifiable increments (addressable points) for a particular axis is,

# of increments 2^n = where n is the number of control bits

A robot with 8 bit control resolution can divide a motion range into 256 discrete positions. The control resolution is about (range of motion)/256. The increments are almost always uniform and equal. If mechanical inaccuracies are negligible, Accuracy = Control Resolution/2

The second limit on control resolution is the bit storage capacity of the controller. If B = the number of bits in the bit storage register devoted to a particular joint, then the number of addressable points in that joint's range of motion is given by 2B•The control resolution is therefore defined as the distance between adjacent addressable points. This electro-mechanical control resolution may be denoted CR. Owing to the wide variety of joints used by robots, and their individual mechanical characteristics, it is not possible to characterize each joint in detail. There is, however, a mechanical limit on the capacity to divide the range of each joint-link system into addressable points, and that limit is given by CR_2 is the bit storage capacity of the controller. This is given by:

$$CR_2 = \frac{R}{2^B - 1}$$

where CR_2 is the control resolution determined by the robot controller; R is the range of the joint-link combination, expressed in linear or angular units; and B is the number of bits in the bit storage register devoted to a particular joint. The maximum of CR_2 gives the control resolution. For repeatability, the mechanical errors that make the robot's end-of-wrist return to slightly different locations than the programmed point are to blame. For a single joint-link mechanism:

$$\mathrm{Re} = \pm 3\sigma$$

where Re is repeatability; and σ is the standard deviation of the error distribution.
For accuracy, we have:

$$Acc = \frac{CR}{2} + 3\sigma$$

where CR is control resolution; and σ is the standard deviation of the error distribution.
Robot precision is determined by three important considerations; these are: control resolution, repeatability, and accuracy.

Example 1.1: Control Resolution, Accuracy and Repeatability in Robotic Arm joint
One of the joints of a certain industrial robot has a type L joint with a range of 0.5 m. The bit storage capacity of the robot controller is 10 bits for this joint. The mechanical errors form a normally distributed random variable about a given taught point. The mean of the distribution is zero and the standard deviation is 0.06 mm in the direction of the output link of the joint. Determine the control resolution *(CR_2)*, accuracy, and repeatability for this robot joint.

Solution: The number of addressable points in the joint range is 2^{10} = 1024. The control resolution is therefore

$\mathrm{CR}_2 = \frac{0.5}{1024-1} = 0.004888\, m = 0.4888\, mm$

Accuracy is given by:

Accuracy = $\frac{0.4888}{2} + 3(0.06) = 0.4244mm$

Repeatability is defines as $\pm 3\ standard\ deviations$

Repeatability = 3 x 0.06 = 0.18 mm

1.3.3 Payload

The payload is always specified as a maximum value, this can be before failure, or more commonly, before serious performance loss.

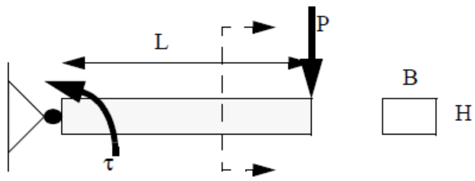
Static considerations,

i. Gravity effects cause downward deflection of the arm and support systems drive gears and belts often have noticeable amounts of slack (backlash) that cause positioning errors
ii. Joint play (windup) - when long rotary members are used in a drive system and twist under load
iii. Thermal effects - temperature changes lead to dimensional changes in the manipulator

Dynamic considerations,

i. Acceleration effects - inertial forces can lead to deflection in structural members. These are normally only problems when a robot is moving very fast, or when a continuous path following is essential. (But, of course, during the design of a robot these factors must be carefully examined)

Example 1.2: Consider a steel cantilever beam of length L, width B and height H, fixed at one end and with a force P, applied at the free end due to the gravitational force on the load.



$$\delta = \frac{PL^3}{3EI}$$

δ = deflection of beamtip caused by point load

E = Youngs modulus = 30×10^6 (psi)

$I = \dfrac{BH^3}{12}$ for rectangular beam

**Note: this deflection does not include the mass of the beam, as might be important in many cases.

1a. Gravity Effects (payload)

Say,
$$P = 100 \text{(lbs)}$$
$$L = 60 \text{(in)}$$
$$B = 4 \text{ (in)}$$
$$H = 6 \text{ (in)}$$

$$\therefore \delta_{\text{payload}} = 0.0033 \text{ (in)}$$

If accuracy = 0.01 then the gravity effects are less
If accuracy = 0.001 then the gravity effects are too large

Aside: Note that the length has a length cubed effect on the tip deflection, so if a second similar link was added to the robot, the deflection would increase 8 times, a third link would increase deflection by 81 times.

1b. Gravity effects (robot link mass)

$$\delta = \frac{\omega L^4}{8EI} \qquad \omega = \frac{\text{weight}}{\text{length}} = 0.91\left(\frac{lb}{in}\right)$$

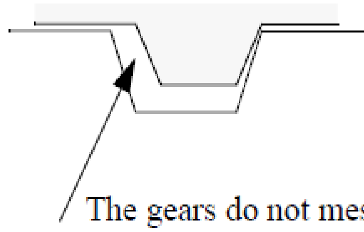
$$\therefore \delta_{\text{link mass}} = 0.00066 \text{ (in)}$$

$$\delta_{\text{total}} = 0.0033 + 0.00066 = 0.00396$$

Aside: If the deflection were too large, then we could use lighter link materials, or larger annular (round tubular) members. Annular members allow actuators, and instrumentation inside.

2. Drive Gear and Belt Drive Play

assume we are using gears, or timing belts, that do not mesh perfectly



The gears do not mesh perfectly, and the resulting space is 'D'

The input drive has to move a distance 'D' before the output engages, and motion begins (this is often after a direction change). This error is multiplied by the gear ratio between input gears and the final position of the robot arm. Similar errors occur for chains, belts, and other types of errors.

> Aside: Some errors can be taken out of the system by using very precise gearing, or anti-backlash gearing that uses springs to hold the input gear against the drive gear. It is also possible to compensate for this in software.

With good gearing, Backlash can be held to less than 0.010 (in), but special design is required when accuracies of 0.001 (in) are desired.

3. Joint Flexibility - (the angular twist of the joints, rotary drives, shafts, under the load)

$$\theta = \frac{32LT}{\pi D^4 G}$$

θ = twist of the cantilevered link in radians

L = distance of the applied moment from the fixed end

T = the applied moment

G = the polar moment of inertia

D = the effective diameter of application of the moment

4. Thermal effects

$$\delta_{thermal} = \alpha \Delta T L \qquad \alpha = \text{coefficient of linear thermal expansion}$$

If for the previous values we consider,

$$\alpha = 6.5 \times 10^{-6} \left(\frac{in}{inF} \right) (\text{for steel})$$

$$\Delta T = T_1(\text{working temp.}) - T_0(\text{calib. temp.}) = 80F - 60F = 20F$$

$$\delta_{thermal} = 0.0078 \text{ (in)}$$

Major errors in accuracy can result from thermal expansion/contraction

5. Acceleration Effects

The robot arm, and payload are exposed to forces generated by acceleration. This applies mainly to the payload mass, but also to the link mass. These forces cause bending moments that must be added to the masses considered before.

$$F_{payload} = M_{payload} r_{payload} \omega' \qquad F_{link} = M_{link} r_{centroid} \omega' (\text{approximate})$$

The robot arm also experiences radial forces due to centripetal forces. These lead to elongation of the arm, but are often negligible.

$$F_{payload} = M_{payload} r_{payload} \omega^2$$

And, if the centre of rotation moves, we must also consider coriolis forces, these could potentially result in a 'whip' effect. This does occur in multi-link robots.

$$\delta = \frac{F_{payload} L^3}{3EI}$$

6. Combine cartesian components of deflection into one vector

$$\delta_{accuracy} = \sqrt{\left(\sum \delta_{x_i}\right)^2 + \left(\sum \delta_{y_i}\right)^2 + \left(\sum \delta_{z_i}\right)^2}$$

*** Remember to compare to control resolution

End of Chapter 1

1. Define Automation.

2. State the main objectives of a modern industry (at least five) and explain the role of automation in helping achieve these.

3. Differentiate between Automation and Mechanization.

4. What are different types of Automation?

5. Can you give an example of an automated system, which contains a control system as a part of it?

6. List some major points why automation is required in industry.

7. What are the advantages and disadvantages of automation?

8. Discuss the concept of low cost automation with the help of suitable example.

9. (a) Discuss the basic elements of an automated system. (b) Describe with neat sketch close loop and open loop control system.

10. How do you classify assembly lines? Explain.

11. Why does an automated system achieve superior performance compared to a manual one? Can you give an example where this happens?

12. During a technical visit to an industry how can you identify the type of automation prevailing there from among the above types?

13. For what kind of a factory would you recommend computer integrated manufacturing and why?

14. What kind of automation would you recommend for manufacturing
    i. Light bulbs
    ii. Garments
    iii. Textile
    iv. Cement
    v. Printing
    vi. Pharmaceuticals
    vii. Toys

15. What is meant by a production system, and what categories of production system are generally specified?

16. Manufacturing systems depend for their operation on the interaction of manual labor and automation. What are the categories of manual labor / automation that can be identified? What mode of automation do these categories usually operate in?

17. When is automation used in a manufacturing system? Describe the three types of automation that can be used in a manufacturing system.

18. Manual labor is used alongside automation in production systems. Name a number of the issues that affect the use of manual labor in production systems.

19. What elements should a strategy for automation implementation consider?

20. Describe the role of Industrial Automation in ensuring overall profitability of a industrial production system. Be specific and answer point wise. Give examples as appropriate.

21. State TWO types of robot control systems and sketch the control system diagram.

22. Name and sketch FIVE types of robots joint with arrow indicate the movement of the joint.

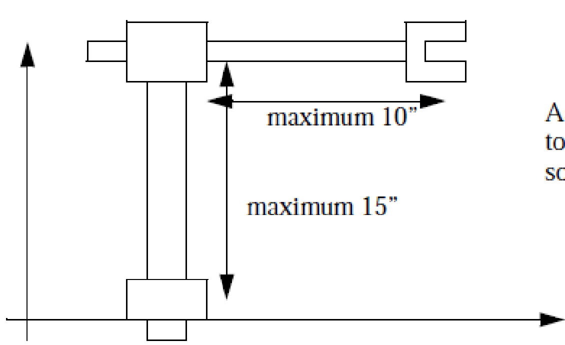
23. Explain the following for the basic concept of Automation
    i. Degree of Freedom (DOF)
    ii. Tool Center Point (TCP)
    iii. Work Envelope
    iv. Pay load
    v. Repeatability
    vi. Accuracy
    vii. Settle time
    viii. Control Resolution

24. Explain with sketches the terms used in automation.
    i. Joint
    ii. Link

25. Explain three (3) types of coordinate that a robot might have.

26. One of the joints of a certain industrial robot has a type L joint with a range of 1.3 m. The bit storage capacity of the robot controller is 25 bits for this joint. The mechanical errors form a normally distributed random variable about a given taught point. The mean of the distribution is zero and the standard deviation is 0.08 mm in the direction of the output link of the joint. Determine the control resolution *(CR_2)*, accuracy, and repeatability for this robot joint.

27. For the robot pictured below, assume that a maximum payload of 10kg is specified. The joints are controlled by stepper motors with 200 steps per revolution. Each of the joints slides, and the gearing is such that 1 revolution of the stepper motor will result in 1" of travel. What is the accuracy of the robot?

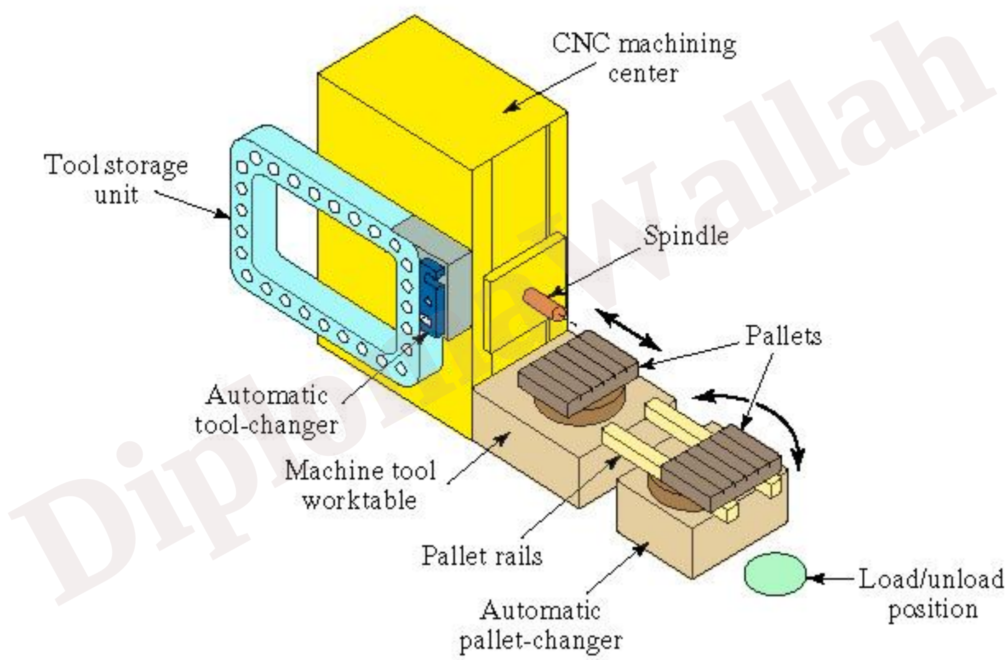


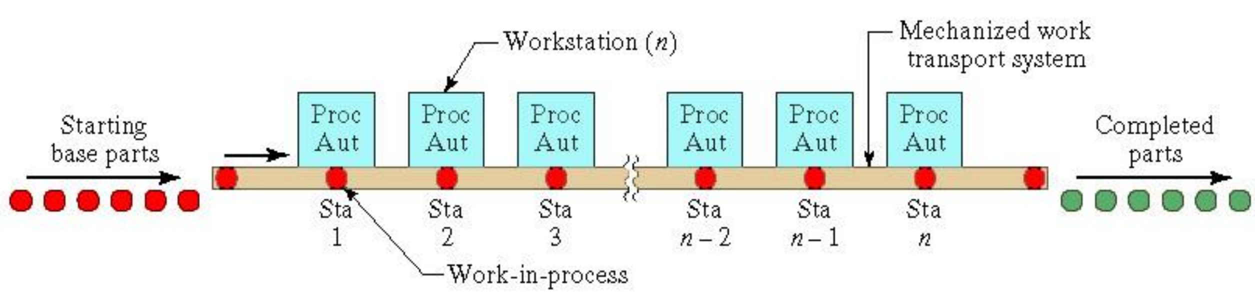
Assume the joints are solid, and to robot links are made from 1" solid aluminum stock.

CHAPTER 2
COMPONENTS AND APPLICATION OF AUTOMATION SYSTEM

Upon completion of this course, students should be able to:-

- Describes the basic component of an automation system
- Describe the automation system in an application
- Define function of an automation systems





2.1 Basic Component of an Automation System

2.1.1 Introduction to Basic Component of an Automation System

Automation is the technology by which a process or procedure is accomplished without human assistance. It is implemented using a program of instructions combined with a control system that executes the instructions. To automate a process, power is required, both to drive the process itself and to operate the program and control system. Although automation can be applied in a wide variety of areas, it is most closely associated with the manufacturing industries. It was in the context of manufacturing that the term was originally coined by an engineering manager at Ford Motor Company in 1946 to describe the variety of automatic transfer devices and feed mechanisms that had been installed in Ford's production plants. It is ironic that nearly all modern applications of automation are controlled by computer technologies that were not available in 1946.

In this part of the book, we examine technologies that have been developed to automate manufacturing operations. The position of automation and control technologies in the larger production system is shown in Figure 2.1. In the present chapter, we provide an overview of automation: What are the elements of an automated system? and What are some of the advanced features beyond the basic elements? In the following two chapters, we discuss industrial control systems and the hardware components of these systems. These two chapters serve as a foundation for the remaining chapters in our coverage of automation and control technologies. These technologies are: (1) numerical control, (2) industrial robotics, and (3) programmable logic controllers.

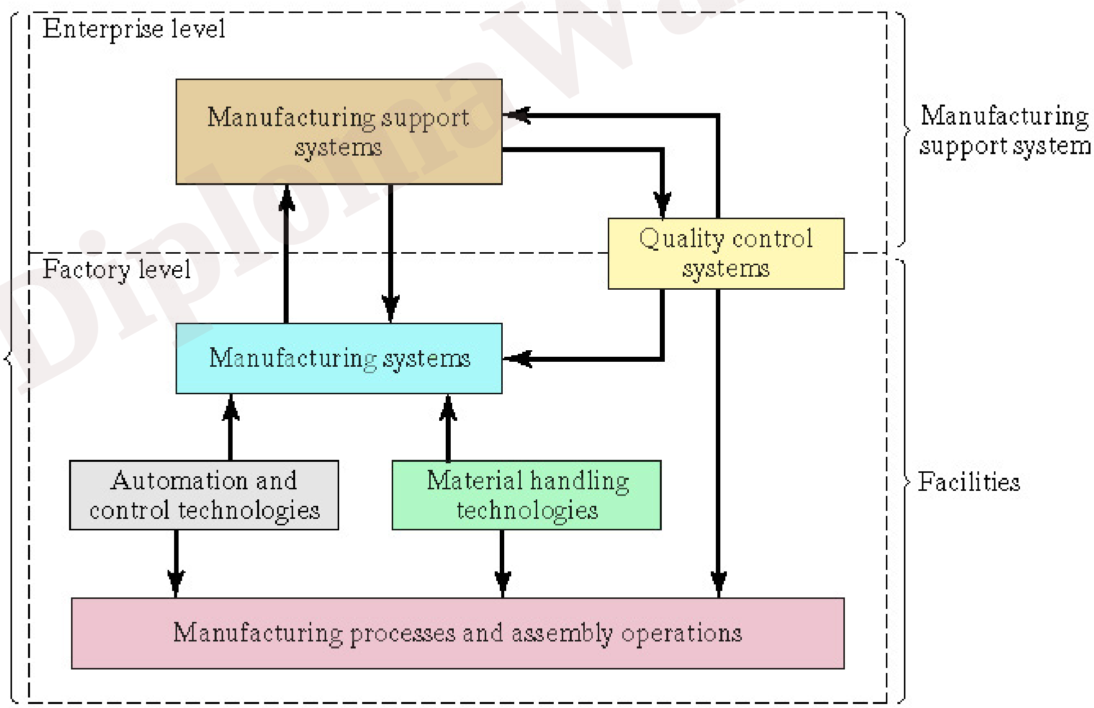


Figure 2.1: Automation and control technologies in the production system

2.1.2 Basic Components of an Automated System

An automated system consists of three basic elements: (1) power to accomplish the process and operate the system, (2) as program of instructions to direct the process, and (3) a control system to actuate the instructions. The relationship amongst these elements is illustrated in Figure 2.2. All systems that qualify as being automated include these three basic elements in one form or another.

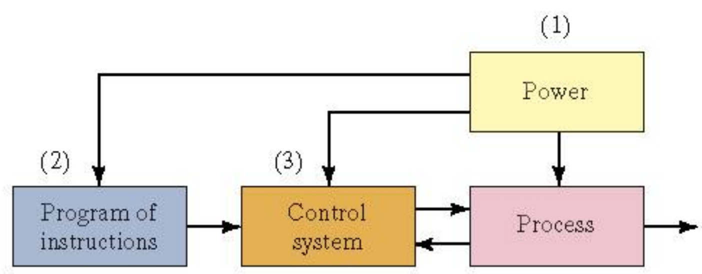


Figure 2.2 Elements of automated system: (1) power, (2) program of instruction and (3) control system

2.1.2.1 Power to Accomplish the Automated Process

An automated system is used to operate some process, and power is required to drive the process; as well as the controls. The principal source of power in automated systems is electricity. Electric power has many advantages in automated as well as non-automated processes:

- Electrical power is widely available at moderate cost. It is an important part of our industrial infrastructure
- Electrical power can be readily converted to alternative energy forms: mechanical, thermal, light, acoustic, hydraulic, and pneumatic.
- Electrical power at low levels can be used to accomplish functions such as signal transmission, information processing, and data storage and communication.
- Electrical energy can be stored in long life batteries for use in locations where an external source of electrical power is not conveniently available.

Alternative power sources include fossil fuels, solar energy, water, and wind. However, their exclusive use is rare in automated systems. In many cases when alternative power sources are used to drive the process itself, electrical power is used for the controls that automate the operation. For example, in casting or heat treatment, the furnace may be heated by fossil fuels but the control system to regulate temperature and time cycle is electrical. In other cases, the energy from these alternative sources is converted to electric power to operate both the process and its automation. When solar energy is used as a power source for an automated system, it is generally converted in this way.

Power for the Process. In production, the term *process* refers to the manufacturing operation that is performed on a work unit. In Table 2.1, a list of common manufacturing processes is compiled along with the form of power required and the resulting action on the work unit. Most of the power in manufacturing plants is consumed by these kinds of operations. The "power form" indicated in the middle column of the table refers to the energy that is applied directly to the process. As indicated above, the power source for each operation is usually converted from electricity. In addition to driving the manufacturing process itself, power is also required for the following material handling functions.

• *Loading and unloading the work unit.* All of the processes listed in Table 2.1 are accomplished on discrete parts. These parts must be moved into the proper position and orientation for the process is performed and power is required for this transport and placement function. At the conclusion of the process, the work unit must similarly be removed. If the process is completely automated, then some form of mechanized power is used. If the process is manually operated or semi-automated, then human power may be used to position and locate the work unit

• *Material transport between operations.* In addition to loading and unloading at a given operation, the work units must be moved between operations.

- *Power for Automation.* Above and beyond the basic power requirements for the power is used for the following functions:

• *Controller unit.* Modern industrial controllers are based on digital computers, which require electrical power to read the program of instructions, make the control calculations, and execute the instructions by transmitting the proper commands to the actuating devices.

• *Power to actuate the control signals.* The commands sent by the controller unit are carried out by means of electromechanical devices, such as switches and motors, called *actuators.* The commands are generally transmitted by means of low-voltage control signals. To accomplish the commands, the actuators require more power and so the control signals must he amplified to provide the proper power level for the actuating device.

• *Data acquisition and information processing.* In most control systems, data must be collected from the process and used as input to the control algorithms. In addition, a requirement of the process may include keeping records of process performance or product quality. These data acquisition and record keeping functions require power, although in modest amounts.

Table 3.1: Common Manufacturing Processes and Their Power Requirements

Process	*Power Form*	*Action Accomplished*
Casting	Thermal	Melting the metal before pouring into a mold cavity where solidification occurs.
Electric discharge machining (EDM)	Electrical	Metal removal accomplished by a series of discrete electrical discharges between electrode (tool and work piece. The electric discharges cause very high localized temperatures that melt the metal.
Forging	Mechanical	Metal work part is deformed by opposing die(s). Work parts are often heated in advance deformation thus thermal power is also required.
Heat treating	Thermal	Metallic work unit is heated to temperature below melting point to effect rnicrostructural changes.
Injection molding	Thermal and mechanical	Heat is used to raise temperature of polymer to highly plastic consistency, and mechanical force is used to inject the polymer melt into a mold cavity.
Laser beam cutting	Light and thermal	A highly coherent light beam is used to cut material by vaporization and melting.
Machining	Mechanical	Cutting of metal is accomplished by relative motion between tool and work piece.
Sheet metal punching and blanking	Mechanical	Mechanical power is used to shear metal sheets and plates.
Welding	Thermal (maybe mechanical)	Most welding processes use heat to cause fusion and coalescence of two (or more) metal parts at their contacting surfaces. Some welding processes also apply mechanical pressure to the surfaces.

2.1.2.2 Program of Instructions

The actions performed by an automated process are defined by a program of instructions. Whether the manufacturing operation involves low, medium, or high production (Section 1.1), each part or product style made in the operation requires one or more processing steps that are unique to that style. These processing steps are performed during a work cycle. A new part is completed during each work cycle (in some manufacturing operations, more than one part is produced during the work cycle; e.g., a plastic injection molding operation may produce multiple parts each cycle using a multiple cavity mold). The particular processing steps for the work cycle are specified in a *work cycle program.* Work cycle programs are called *part programs* in numerical control. Other process control applications use different names for this type of program.

Work Cycle Programs. In the simplest automated processes, the work cycle consists of essentially one step, which is to maintain a single process parameter at a defined level, for example, maintain the temperature of a furnace at a designated value for the duration of a heat treatment cycle. (We assume that loading and unloading of the work units into and from the furnace is performed manually and is therefore not part of the automatic cycle.) In this case, programming simply involves setting the temperature dial on the furnace. To change the program, the operator simply changes the temperature setting. An extension of this simple case is when the single-step process is defined by more than one process parameter, for example, a furnace in which both temperature and atmosphere are controlled.

In more complicated systems, the process involves a work cycle consisting of multiple steps that are repeated with no deviation from one cycle to the next. Most discrete part manufacturing operations are in this category. A typical sequence of steps (simplified) is: (1) load the part into the production machine, (2) perform the process, and (3) unload the part. During each step, there are one or more activities that involve changes in one or more process parameters. *Process parameters* are inputs to the process such as temperature setting of a furnace, coordinate axis value in a positioning system; valve opened or closed in a fluid flow system, and motor on or off. Process parameters are distinguished from *process variables,* which are outputs from the process; for example, the actual temperature of the furnace, the actual position of the axis, the actual flow rate of the fluid in the pipe, and the rotational speed of the motor. As our list of examples suggests, the changes in process parameter values may be continuous (gradual changes during the processing step; for example, gradually increasing temperature during a heat treatment cycle) or discrete (stepwise changes; for example, on/off). Different process parameters filly be involved in each step.

EXAMPLE 2.1: An Automated Turning Operation

Consider an automated turning operation in which a cone-shaped geometry is generated. Assume the system is automated and that a robot is used to load and unload the work unit. The work cycle consists of the following steps: (1) load starting work piece (2) position cutting tool prior to turning (3) turn (4) reposition tool to a safe location at end of turning and (5) unload finished work piece. Identify the activities and process parameter(s) in each step of the operation.

Solution: In step (1), the activities consist of the robot manipulator reaching for the raw work part. Lifting and positioning the part into the chuck jaws of the lathe, then removing the manipulator to a safe position to wait unloading. The process parameters for these activities are the axis values of the robot manipulator (which change continuously), the gripper value (open or closed).and the chuck jaw value (open or closed)

In step (2), the activity involves the movement of the cutting tool to a "ready' position. The process parameters associated with this activity are the r-and z-axis position of the tool

Step (3) is the turning operation. It requires the simultaneous control of three process parameters: rotational speed of the work piece (rev/min), feed (mm/rev), and radial distance of the cutting tool from the axis of rotation. To cut the conical shape, radial distance must be changed continuously at a constant rate for each revolution of the work piece. For a consistent finish on the surface, the rotational speed must be continuously adjusted to maintain a constant surface speed (m/min); and [or equal feed marks on the surface, the feed must be set at a constant value. Depending on the angle of the cone, multiple turning passes may be required to gradually generate the desired contour. Each pass represents an additional step in the sequence.

Steps (4) and (5) involve the reverse activities as steps (2) and (1), respectively and the process parameters are the same.

Many production operations consist of multiple steps, sometimes more complicated than our turning example. Examples of these operations include automatic screw machine cycles, sheet metal stamping operations, plastic injection molding, and die casting. Each of these manufacturing processes has been used for many decades. In earlier versions of these operations, the work cycles were controlled by hardware components, such as limit switches, timers, cams, and electromechanical relays. In effect, the hardware components and their arrangements served (IS the program of instructions that directed the sequence of steps in the processing cycle. Although these devices were quite adequate in performing their sequencing function, they suffered from the following disadvantages: (1) They often required considerable time to design and fabricate, thus forcing the production equipment to be used for batch production only; (2) making even minor changes in the program was difficult and time consuming; and (3) the program was in a physical form that is not readily compatible with computer data processing and communication.

Modern controllers used in automated systems are based on digital computers. Instead of cams, timers, relays, and other hardware devices, the programs for computer-controlled equipment are contained in magnetic tape, diskettes, compact disks (CD-ROMs), computer memory, and other modern storage technologies. Virtually all new equipment that perform the above mass production operations are designed with some type of computer controller to execute their respective processing cycles. The use of digital computer’s as the process controller allows improvements and upgrades to be made in the control programs, such as the addition of control functions not foreseen during initial equipment design. These kinds of control changes are often difficult to make with the previous hardware devices.

The work cycle may include manual steps, where the operator performs certain activities during the work cycle, and the automated system performs the rest. A common example is the loading and unloading of parts by the operator into and from a numerical control machine between machining cycles, where the machine performs the cutting operation under part program control. Initiation of the cutting operation of each cycle is triggered by the operator activating a "start" button after the part has been loaded.

Decision-Making in the Programmed Work Cycle. In our previous discussion of automated work cycles, the only two features of *the* work cycle are (1) the number and sequence of processing steps and (2) the process parameter changes in each step. Each work cycle consists of the same steps and associated process parameter changes with no variation from one cycle to the next. The program of instructions is repealed each work cycle without deviation. In fact, many automated manufacturing operations require decisions to be made during the programmed work cycle to cope with variations in the cycle. In many cases, the variations are routine elements of the cycle, and the corresponding instructions for dealing with them are incorporated into the regular part program. These cases include:

• *Operator interaction.* Although the program of instructions is intended to be carried out without human interaction, the controller unit may require input data from a human operator in order to function. For example, in an automated engraving operation the operator may have to enter the alphanumeric characters that are to be engraved on the work unit (e.g. plaque, trophy, belt buckle). Having entered the characters, the engraving operation is accomplished automatically by the system. (An everyday example of operator interaction with an automated system is a bank customer using an automated teller machine. The customer must enter the codes indicating what transaction is to be accomplished by the teller machine.)

• *Different part or product styles processed by the system.* In this instance, the automated system is programmed to perform different work cycles on different part or product styles. An example is an industrial robot that performs a series of spot welding operations on car bodies in a final assembly plant. These plants are often designed to build different body styles on the same automated assembly line, such as two-door and four-door sedans. As each car body enters a given welding station on the line, sensors identify which style it is, and the robot performs the correct series of welds for that style.

• *Variations in the starting work units.* In many manufacturing operations the starting work units are not consistent. A *good* example is a sand casting as the starting work unit in a machining operation. The dimensional variations in the raw castings sometimes necessitate an extra machining pass to bring the machined dimension to the specified value. The part program must be coded to allow for the additional pass when necessary.

In all of these examples, the routine variations can be accommodated in the regular *work* cycle program. The program can be designed to respond to sensor or operator inputs by executing the appropriate subroutine corresponding to the input. In other cases, the variations in the work cycle are not routine at all. They are infrequent and unexpected, such as the failure of an equipment component. In these instances, the program must include contingency procedures or modifications in the sequence to cope with conditions that lie outside the normal routine. We discuss these measures later in the chapter in the context of advanced automation functions (Section 2.2). A variety of production situations and work cycle programs has been discussed here The features of work cycle programs (part programs) used to direct the operations of an automated system are summarized as in Table 2.2.

TABLE 2.2: Features of Work Cycle Programs Used in Automated Systems

Program Feature	Examples or Alternatives
Steps in work cycle	Example • Typical sequence of steps: (1) load, (2), process, (3) unload
Process parameters (inputs) in each step	Alternatives: • One parameter versus multiple parameters that must be changed during the step • Continuous parameters versus discrete parameters • Parameters that change during the step; for example, a positioning system whose axes values change during the processing step
Manual steps in work cycle	Alternatives: • Manual steps versus no manual steps (completely automated work cycle Example • Operator loading and unloading parts to and from machine
Operator interaction	Alternatives: • Operator interaction versus completely automated work cycle Example: • Operator entering processing information for current work part
Different part or product styles	Alternatives: • Identical part or product style each cycle (mass or batch production) versus different part or product styles each cycle (flexible automation)
Variations in starting work units	Example • Variations in starting dimensions or part features

2.1.2.3 Control System

The control element of the automated system executes the program of instructions. The control system causes the process to accomplish its defined function, which for our purpose is to carry out some manufacturing operation. Let us provide a brief introduction to control systems here. The following chapter describes this important industrial technology in more detail.

The controls in an automated system can be either closed loop or open loop. A *closed loop control system,* also known as a *feedback control system,* is one in which the output variable is compared with an input parameter, and any difference between the two is used to drive the output into agreement with the input. As shown in Figure 2.3, a closed loop control system consists of six basic elements: (1) input parameter, (2) process, (3) output variable (4) feedback sensor (5) controller and (6) actuator. The *input parameter* often referred to as the *set point,* represents the desired value of the output. In a home temperature control system, the set point is the desired thermostat setting. The *process* is the operation or function being controlled. In particular, it is the *output variable* that is being controlled in the loop. In the present discussion, the process of interest is usually a manufacturing operation, and the output variable is some process variable, perhaps a critical performance measure in the process, such as temperature or force or flow rate. A *sensor* is used to measure the output variable and close the loop between input and output. Sensors perform the feedback function in a closed loop control system. The controller compares the output with the input and makes the required adjustment in the process to reduce the difference between them. The adjustment is accomplished using one or more *actuators,* which are the hardware devices that physically carry out the control actions, such as an electric motor or a flow valve. It should be mentioned that our model in Figure 2.3 shows closed loop control system. Most industrial processes require multiple loops, one for each process variable that must be controlled

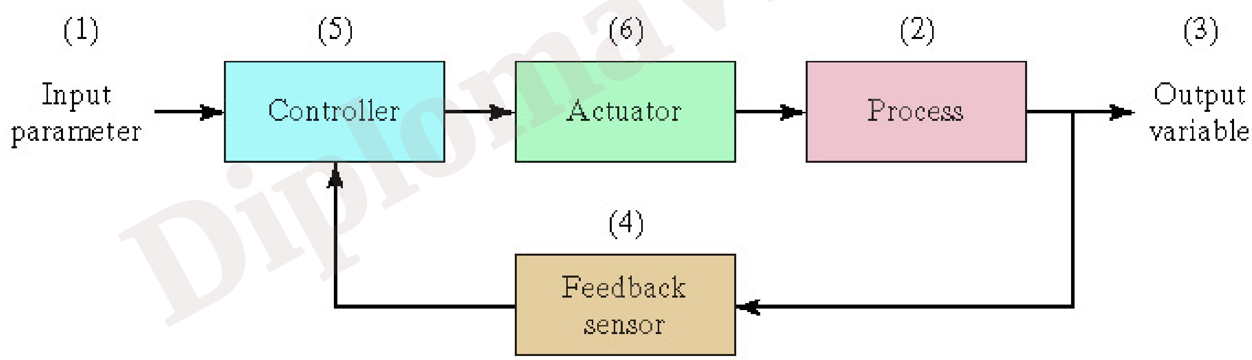


Figure 2.3: A feedback control system

In contrast to the closed loop control system, an *open loop control system* operates without the feedback loop, as in Figure 2.4. In this case, the controls operate without measuring the output variable.so no comparison is made between the actual values of the output and the desired input parameter. The controller relies on an accurate model of the effect of its actuator for the process variable. With an open loop system, there is always the risk that the actuator will not have the intended effect on the process, and that is the disadvantage of an open loop system. Its advantage is that it is generally simpler and less expensive than a closed loop system. Open loop systems are usually appropriate when the following conditions apply: (1) The actions performed by the control system are simple, (2) the actuating function is very reliable, and (3) any reaction forces opposing the actuation are small enough to have no effect on the actuation. If these characteristics are not applicable, then a closed loop control system may be more appropriate.

Consider the difference between a closed loop and open loop system for the case of a positioning system. Positioning systems are common in manufacturing to locate a work part relative to a tool or work head. In operation, the system is directed to move the worktable to a specified location as defined by a coordinate value in a Cartesian (or other) coordinate system. Most positioning systems have at least two axes (e.g. an x-*y* positioning table) with a control system for each axis, but our diagram only illustrates one of these axes. A dc servomotor connected to a lead screw is a common actuator for each axis. A signal indicating the coordinate value (e.g. x-value) is sent from the controller to the motor that drives the lead screw, whose rotation is converted into linear motion of the positioning table. As the table moves closer to the desired x-coordinate value, the difference between the actual x-position and the input revalue is reduced. The actual reposition is measured by a feedback sensor (e.g. an optical encoder). The controller continues to drive the motor until the actual table position corresponds to the input position value.

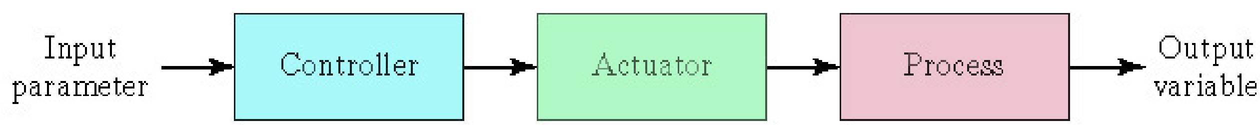


Figure 2.4: An open loop control system

For the open loop case, the diagram for the positioning system would be similar to the preceding, except that no feedback loop is present and a stepper motor is used in place of the dc servomotor. A stepper motor is designed to rotate a precise fraction of a turn for each pulse received from the controller. Since the motor shaft is connected to the lead screw, and the lead screw drives the worktable, each pulse converts into a small constant linear movement of *the* table. To move the table a desired distance, the number of pulses corresponding to that distance is sent to the motor, given the proper application, whose characteristic match the preceding list of operating conditions. An open loop positioning system works with high reliability.

2.2 Automation System in an Application

An automated system in an application consists of a series of workstations connected by a transfer system to move parts between the stations. This is an example of fixed automation, since these lines are typically set up for long production runs, perhaps making millions of product units and running for several years between changeovers. Each station is designed to perform a specific processing operation, so that the part or product is constructed stepwise as it progresses along the line. A raw work part enters at one end of the line, proceeds through each workstation, and emerges at the other end as a completed product. In the normal operation of the line, there is a work part being processed at each station, so that many parts are being processed simultaneously and a finished part is produced with each cycle of the line. The various operations, part transfers, and other activities taking place on an automated transfer line must all be sequenced and coordinated properly for the line to operate efficiently. Modern automated lines are controlled by programmable logic controllers, which are special computers that facilitate connections with industrial equipment (such as automated production lines) and can perform the kinds of timing and sequencing functions required to operate such equipment.

Automated production lines are utilized in many industries, most notably automotive, where they are used for processes such as machining and press work. Machining is a manufacturing process in which metal is removed by a cutting or shaping tool, so that the remaining work part is the desired shape. Machinery and motor components are usually made by this process. In many cases, multiple operations are required to completely shape the part. If the part is mass-produced, an automated transfer line is often the most economical method of production. The many separate operations are divided among the workstations. Transfer lines date back to about 1924.

Press work operations involve the cutting and forming of parts from sheet metal. Examples of such parts include automobile body panels, outer shells of major appliances (e.g., laundry machines and ranges), and metal furniture (e.g., desks and file cabinets). More than one processing step is often required to complete a complicated part. Several presses are connected together in sequence by handling mechanisms that transfer the partially completed parts from one press to the next, thus creating an automated press work line.

2.2.1 Applications Examples: Single Station Manned Cells

Most industrial production operations are based on the use of single station manned and automated cells. Let us expand the list here:

• A CNC machining center. The machine executes a part program for each part. The parts are identical. A worker is required to be at the machine at the end of each program execution to unload the part just completed and load a raw work part onto the machine table.

• A CNC turning center. The machine executes a part program for each part. The parts are identical. A worker is required to unload finished parts and place them in a tote pan and then load raw parts from another tote pan. This is similar to the preceding machining center, but a different machining process is performed.

• Same as the preceding except the parts are not identical. In this case, the machine operator must call the appropriate part program and load it into the CNC control unit for each consecutive work part.

• A duster of two CNC turning centers, each producing the same part but operating independently from its own machine control unit. A single worker attends to the loading and unloading of both machines. The part programs are long enough relative to the load/unload portion of the work cycle that this can be accomplished without forced machine idle time.

• A plastic injection molding machine on semi .automatic cycle, with a worker present to remove the molding, sprue, and runner system when the mold opens each molding cycle. Parts are placed in a box by the worker. Another worker must periodically exchange the tote box and resupply molding compound to the machine.

• A worker at an electronics assembly workstation placing components onto printed circuit boards in a batch operation. The worker must periodically delay production and replace the supply of components that are stored in tote bins at the station. Starting and finished boards are stored in magazines that must be periodically replaced by another worker.

• A worker at an assembly workstation performing mechanical assembly of a simple product (or subassembly of a product) from components located in tote bins at the station.

• A stamping press that punches and forms sheet metal parts from flat blanks in a stack near the press. A worker is required to load the blank into the press, actuate the press, and then remove the stamping each cycle. Completed stampings are stored in four-wheel trucks that have been especially designed for the part.

2.2.2 Applications Examples: Single Station Automated Cells

Following are examples of single station automated cells. We have taken each of the preceding examples:

• A CNC machining center with parts carousel and automatic pallet changer, as in the layout of Figure 2.5. The parts are identical, and the machining cycle is controlled by a part program. Each part is held on a pallet fixture. The machine cuts the parts one-by-one, when all of the parts in the carousel have been machined; a worker removes the finished pieces from the carousel and loads starting work parts. Loading and unloading of the carousel can be performed while the machine is operating.

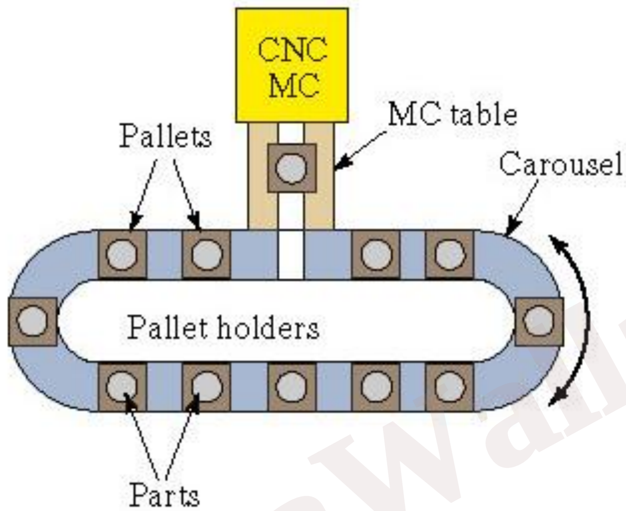


Figure 2.5: A CNC Machining Centre

• A CNC turning center with parts storage tray and robot. The robot is equipped with a dual gripper to unload the completed piece and load a starting work part from the parts storage tray each cycle. The parts storage tray can hold a certain quantity of parts. In effect, this is the same case as the CNC machining center, just a different machining process, Figure 2.6.

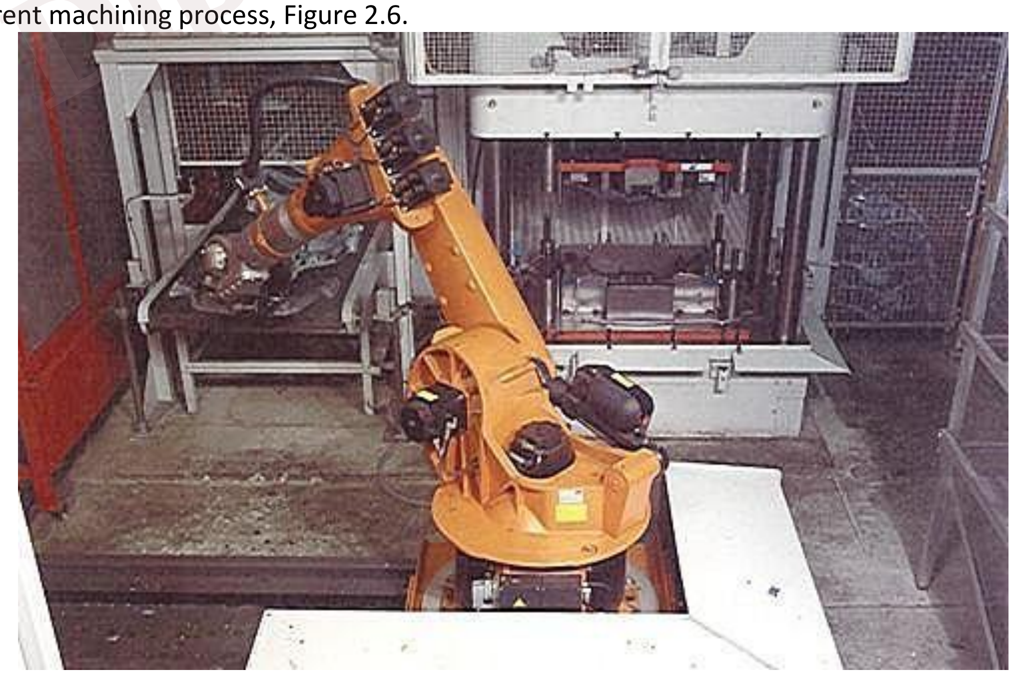


Figure 2.6: Robot in the Single Station Automated Cell

• Same as the preceding except the parts are not identical. In this case, the appropriate part program is automatically downloaded to the CNC control unit for each consecutive work part, based on either a given production schedule or an automatic part recognition system that identifies the raw part.

• A cluster of ten CNC turning centers, each producing a different part. Each workstation has its own parts carousel and robotic arm for loading and unloading between the machine and the carousel. A single worker must attend all ten machines by periodically unloading and loading the storage carousels. The time required to service a carousel is short relative to the time each machine can run unattended, so all ten machines can be serviced with no machine idle time.

• A plastic injection molding machine on automatic cycle, with mechanical and to ensure removal of the molding, sprue, and runner system each molding cycle. Parts are collected in a tote box beneath the mold. A worker must periodically exchange the tote box and resupply molding compound to the machine.

• An automated insertion machine assembling electronic components onto printed circuit boards in a batch operation. Starling boards and finished boards are stored in magazines for periodic replacement by a human worker. The worker must also periodically replace the supply of components, which are stored in long magazines.

• A robotic assembly cell consisting of one robot that assembles 8 simple product (or subassembly of a product) from components presented by several parts delivery systems (e.g. bowl feeders).

• A stamping press that punches and forms small sheet metal parts from a long coil, as depicted in Figure 2.7. The press operates at a rate of 180 cycles/min and 9000 parts can be stamped from each coil. The stampings are collected **in** a tote box on the output side of the press. When the coil runs out, it must be replaced with a new coil, and the tote box is replaced at the same lime.

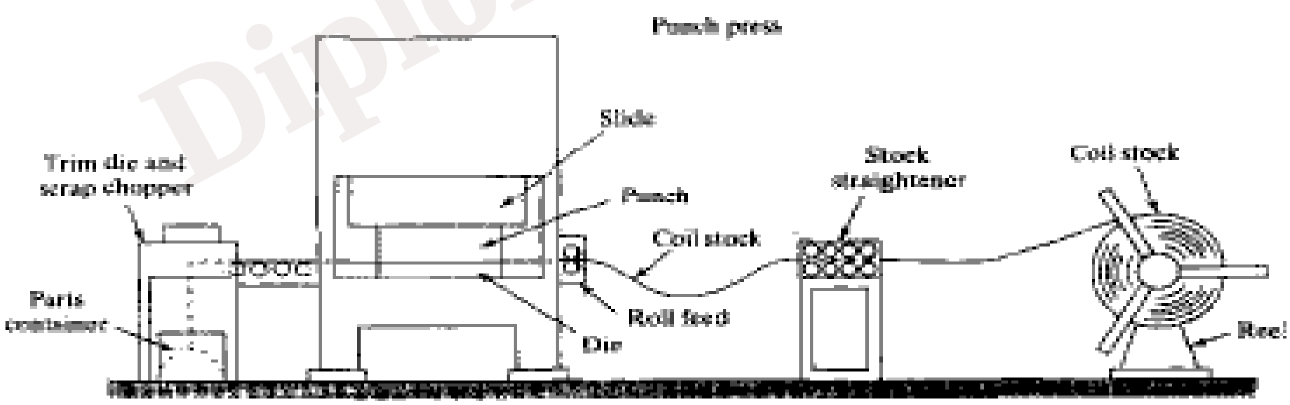


Figure 2.7: Stamping press on automatic cycle producing stampings from a sheet metal coil.

2.2.3 Applications Examples: CNC Machining and Turning Centers

Several of our application examples of single station manufacturing cells consisted of CNC machining centers and turning centers. The *machining center,* developed in the late 1950s before the advent of computer numerical control (CNC), is a machine tool capable of performing multiple machining operations on a work part in one setup under NC program control. Today's machining centers lire CNC. Typical cutting operations performed on a machining center are those that use as rotating cutting tool, such as milling, drilling, reaming, and tapping.

Machining centers are classified as vertical, horizontal, or universal. The designation refers to the orientation of the machine spindle. A vertical machining center has its spindle on a vertical axis relative to the worktable, and a horizontal machining center has its spindle on a horizontal axis. This distinction generally results in a difference in the type of work that is performed on the machine. A vertical machining center is typically used for flat work that requires tool access from the top. A horizontal machining center is used for cube-shaped parts where tool access can best be achieved on the sides of the cube. Universal machining centers have work heads that swivel their spindle axes to any angle between horizontal and vertical, thus making this a very flexible machine tool.

Numerical control machining centers are usually designed with features to reduce non-productive time. These features include the following:

•*Automatic tool-changing.* A variety of machining operations means that a variety of cutting tools is required. The tools are contained in a tool storage unit that is integrated with the machine tool. When a cutter needs to be changed, the tool drum rotates to the proper position, and an automatic tool changer (ATC), operating under part program control, exchanges the tool in the spindle for the tool in the tool storage unit. Capacities of the tool storage unit commonly range from 16 to 80 cutting tools.

• *Automatic work part positioning.* Many horizontal and universal machining centers have the capability to orient the work part relative to the spindle. This is accomplished by means of a rotary table on which the work part is fixture. The table can be oriented at any angle about a vertical axis to permit the cutting tool to access almost the entire surface of the part in a single setup.

• *Automatic pallet changer.* Machining centers are often equipped with two (or more) separate pallets that can be presented to the cutting tool using an automatic pallet changer. While machining is being performed with one pallet in position at the machine, the other pallet is in a safe location away from the spindle. In this safe location, the operator can unload the finished part from the prior cycle and then fixture the raw work part for the next cycle while the current work piece is being machined.

A numerically controlled horizontal machining center, with many of the features described above, is shown in Figure 2.8.

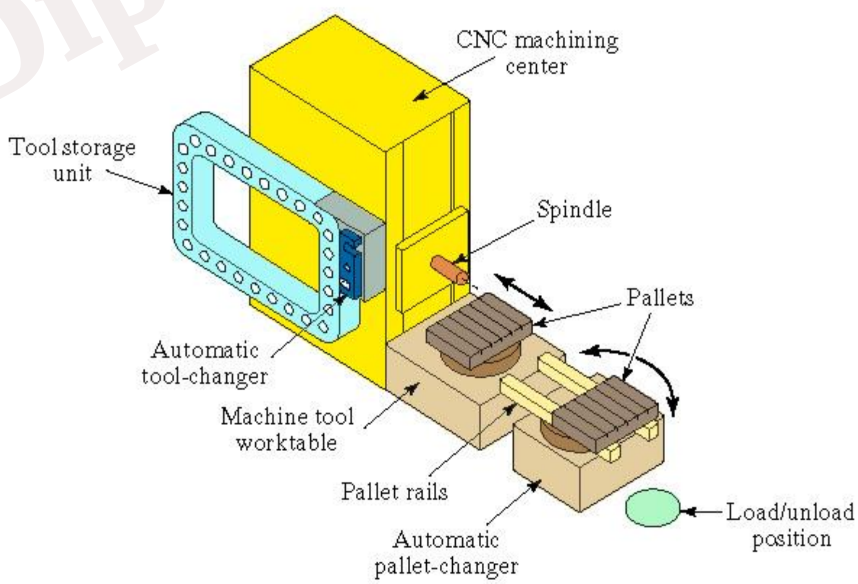


Figure 2.8: Automatic pallet changer integrated with a CNC machining center, set up for manual unloading and loading of work parts.

The success of NC machining centers motivated the development of NC turning centers, A modem *NC turning center,* Figure 2.9, is capable of performing various turning and related operations, contour turning, and automatic tool indexing, all under computer control In addition, the most sophisticated turning centers can accomplish: (1) work part gaging: checking key dimensions after machining, (2) tool monitoring: sensing when the tools are worn, (3) automatic tool changing when tools become worn, and (4) automatic work part changing at the completion of the work cycle.

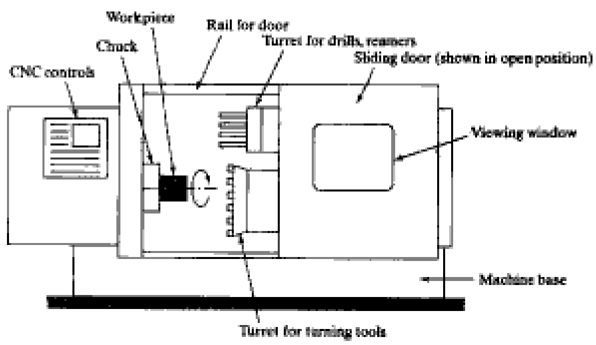


Figure 2.9: Front view of a CNC turning center showing two tool turrets, one for single point turning tools and the other for drills and similar tools. Turrets can be positioned under NC control to cut the work piece.

2.2.3 Application Examples: Automation Storage/Retrieval System

An automated storage/retrieval system (AS/RS) is a storage system that performs storage and retrieval operations with speed and accuracy under a defined degree of automation. Different levels of automation may be applied. At one extreme, the AS/RS is completely automated. This can include a full complement of totally automated, computer-controlled storage functions that are integrated into overall factory or warehouse operations. At the other extreme it may use human workers to control equipment and perform storage/retrieval transactions. Using modular components, available from AS/RS vendors, the AS/RS system is custom-designed to fit the requirements of the plant in which it is installed.

The basic equipment of the AS/RS include a rack structure used for storing loads, plus a storage/retrieval (S/R) mechanism with three dimensions of motion (x, y, z). Additionally, the AS/RS maintains one or more storage aisles that are serviced by the S/R mechanism. The S/R mechanism is used to deliver materials to the storage racks and to retrieve materials from the racks. Each aisle has an input/output station where storage deliveries are transferred into the system, or out-of the system; these stations are known as pickup-and-deposit (P&D) stations. P&D stations may be manually operated or connected to an automated transport system, such as a conveyor or an AGVS (see Figure 2.10).

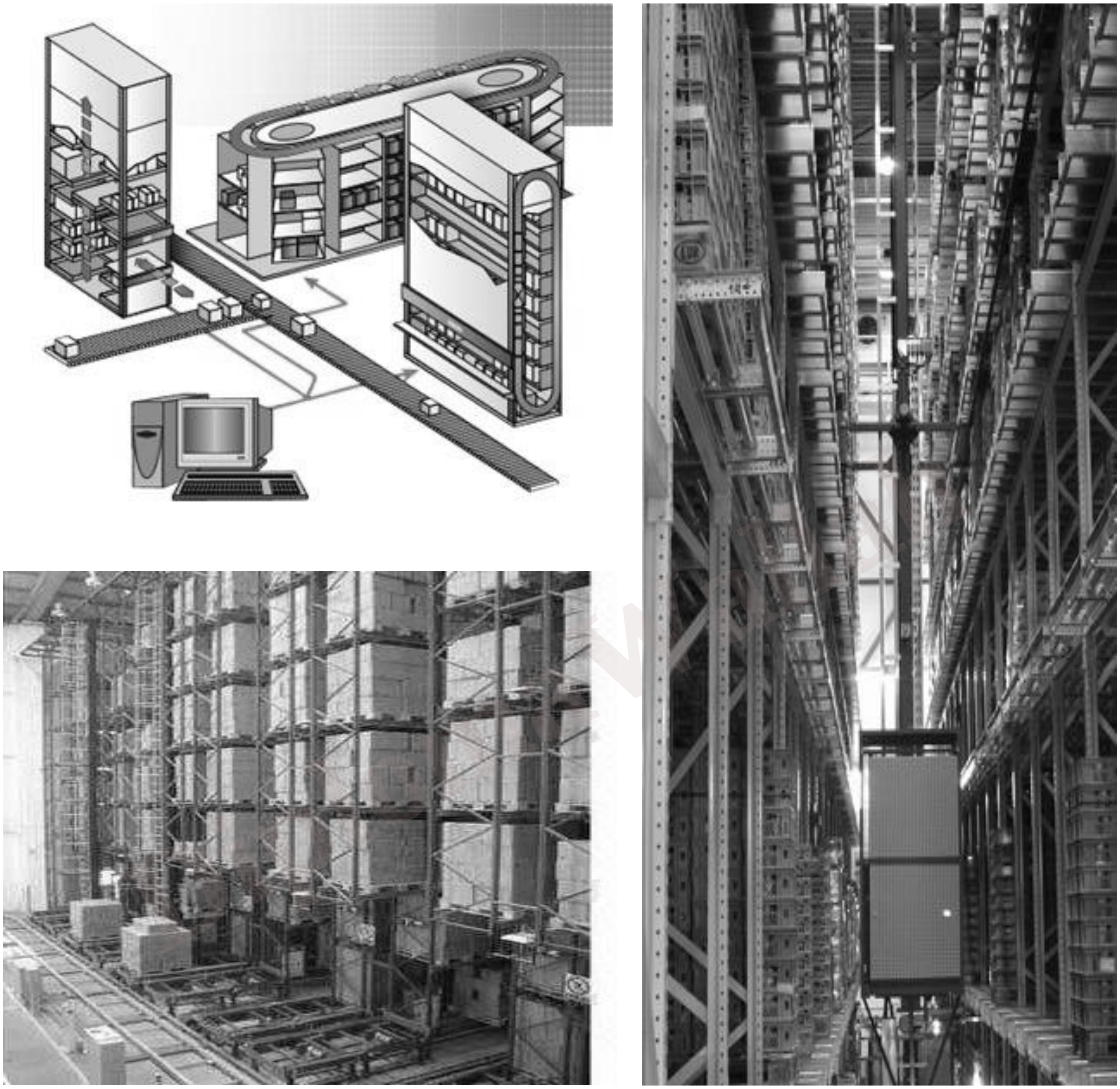


Figure 2.10: Automated Storage and Retrieval System (AS/RS)

2.2.4 Application Examples: Automation Machining with Robot

The system is flexible if we can answer "yes" to all of these questions, with the most important criteria for flexibility being numbers 1 and 2. Numbers 3 and 4 are softer criteria that may be implemented at various levels. In Figure 15.1 the automated manufacturing cell with two machine tools and robot shall be considered flexible if it: (1) can machine different part mixes taken from the carousel by the robot; (2) allows for changes in the production schedule, without affecting the operation of the robotic arm and the two machine tools; (3) is able to carry-on operating even if one machine tool breaks down; and (4) can accommodate new part designs if the numerical control programmed to do so is written off-line and then downloaded by the system for execution.

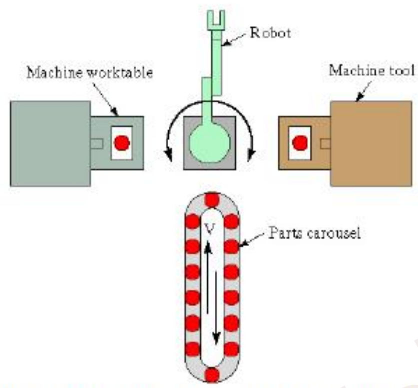


Figure 2.11: Automated manufacturing cell with two machine tools and robot

2.2.5 Application Examples: Automation Loading/ Unloading System

Machine cell—contains one machine (often a CNC machining center) connected to a parts storage system, which can load and unload parts to and from the storage system (as in Figure 2.12). It is designed to operate in batch mode, flexible mode, or a combination of the two. When in batch mode, the system processes parts of a single style in specific lot sizes before physical and programmed changeover to the next batch specifications; in flexible mode the system satisfies three of the four tests for flexibility—the exception being error recovery, since, if the CNC machine center breaks down, the system stops

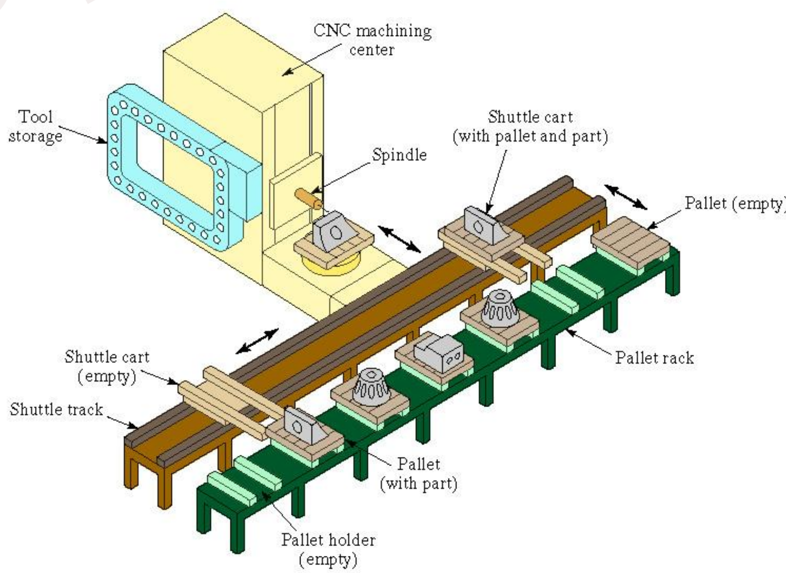


Figure 2.12 Single machine cell with one CNC machining center and parts storage unit

2.2.6 Application Examples: Automation CNC Machining System

Manufacturing cell—contains two or three processing workstations (often CNC machining or turning centers), plus a parts handling system, as in Figure 2.13. This set-up can operate in flexible mode and batch mode, as necessary, and can readily adapt to evolving production schedule and increased production volumes. Since there is more than one machine, error recovery is possible by re-routing the failed machine's intended parts for processing to the other two machines in the system; and new part designs can be introduced with relative ease into the set-up. The flexible manufacturing cell satisfies all four flexibility tests.

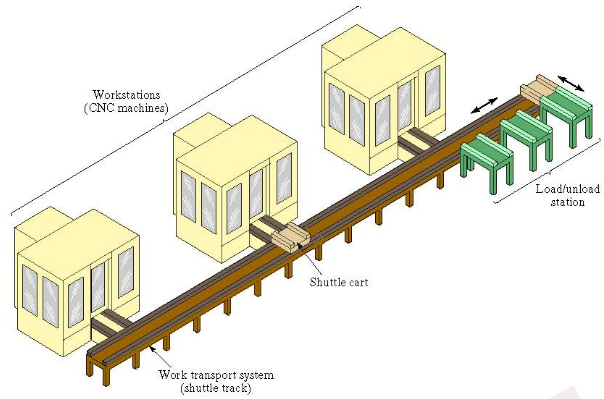


Figure 2.13: Flexible manufacturing cell with three identical processing stations, a load/unload station, and parts handling system

2.2.7 Application Examples: Automation Machining and Inspection System

An example of a non-rotational manufacturing is illustrated in Figure 2.14, which is taken from the Vought Aerospace plant in Dallas, US. This system is used to machine approximately 600 different aircraft components, by means of a manufacturing with eight CNC horizontal machining centers, plus inspection modules. Automated guided vehicles are used as the primary and secondary material handling system. There are two load and unload stations in the system, both of which maintain carousels so that parts may be stored on pallets before and after production. This system can process a sequence of single, one-of-a-kind parts in a continuous mode, so a complete set of components for one aircraft may be made efficiently without batching.

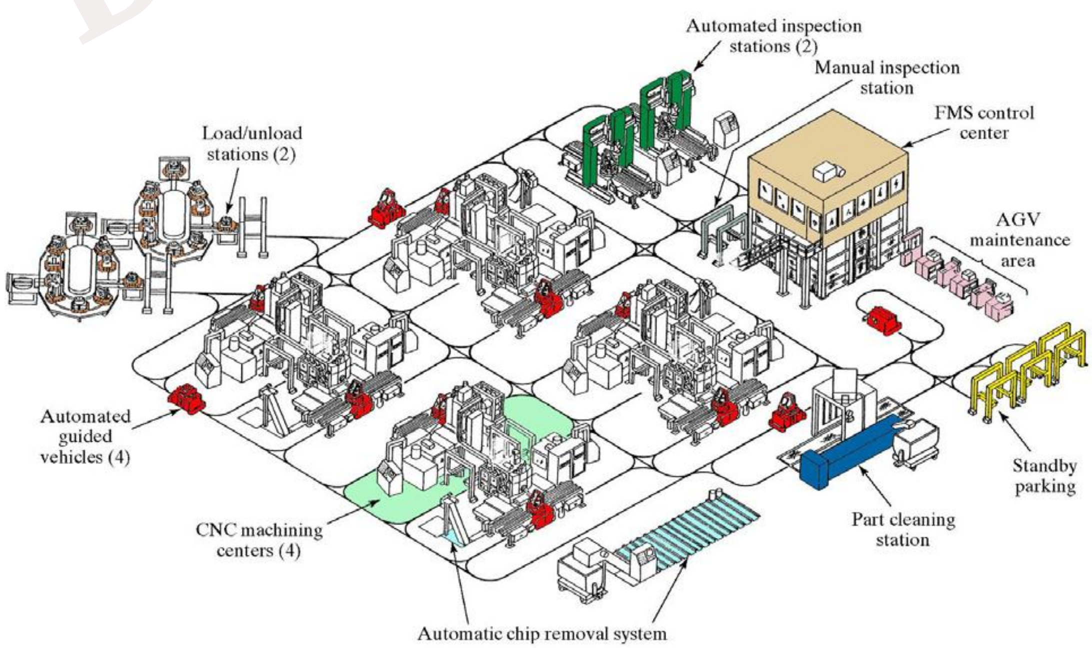


Figure 2.14: Flexible Manufacturing System in Vought Aerospace

2.2.8 Application Examples: Automation Fabricating System

An example of a fabricating flexible manufacturing system for automated sheet-metal processing is shown in Figure 15.9, based upon the flexible manufacturing system (FMS) at the Allen-Bradley Company. This flexible manufacturing system produces motor starters in 125 model styles; with the line having a one-day manufacturing lead time on lot sizes as low as one, and production rates of 600 units per hour. There are 26 workstations in the system: these perform all the processes necessary to complete the product, from assembly, sub-assembly, testing, and packaging. Workstation composition includes the use of linear and dial-indexing assembly machines, with pick-and-place robots performing the handing functions between workstations. Each step uses 100% automated testing to ensure very high quality levels.

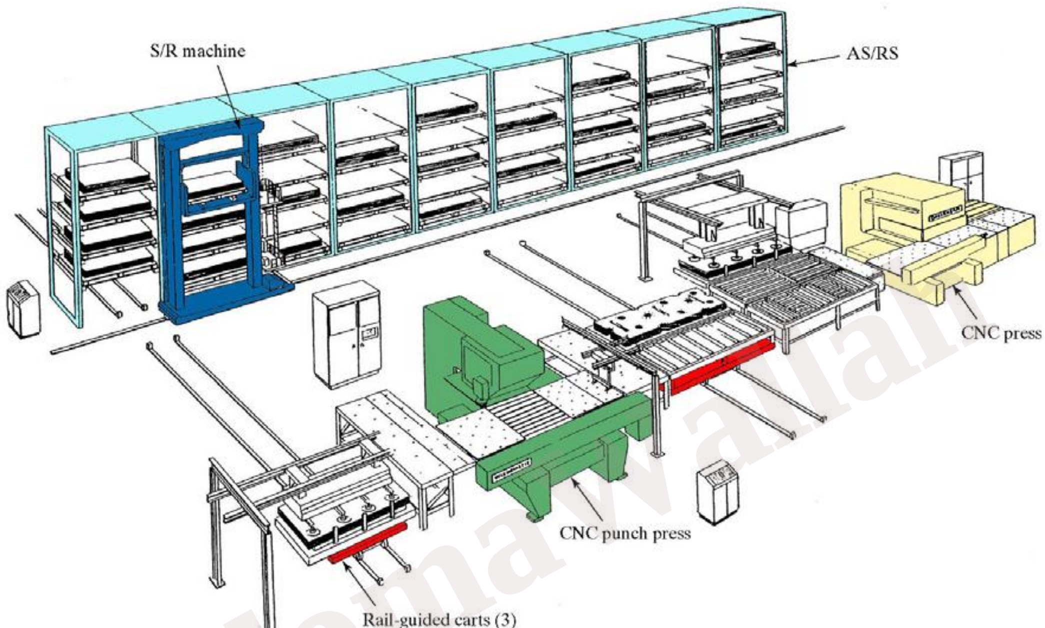


Figure 2.15: Flexible manufacturing system (FMS) at the Allen-Bradley Company

2.3 Functions of Automation Systems

2.3.1 Specifications Functions of Automation Systems

In addition to executing work cycle programs, an automated system may be capable of executing advanced functions that are not specific to a particular work unit. In general, the functions are concerned with enhancing the performance and safety of the equipment. Advanced automation functions include the following: (1) safety monitoring, (2) maintenance and repair diagnostics, and (3) error detection and recovery.

Advanced automation functions are made possible by special subroutines included in the program of instruction. In some cases, the functions provide information only and do not involve any physical actions by the control system. An example of this case includes reporting a list of preventive maintenance tasks that should be accomplished. Any actions taken on the basis of this report are decided by the human operators and managers of the system and not by the system itself. In other cases, the program of instructions must be physically executed by means of the control system using available actuators. A simple example of this case is a safety monitoring system that sounds an alarm when a human worker gets dangerously close to the automated system.

2.3.1.1 Safety Monitoring

One of the significant reasons for automating a manufacturing operation is to remove worker(s) from a hazardous working environment. An automated system is often installed to perform a potentially dangerous operation that would otherwise be accomplished manually by human workers. However, even in automated systems, workers are still needed to service the system at periodic time intervals, if not full-time. Accordingly, it is important that the automated system be designed to operate safely when workers are in attendance. In addition, it is essential that the automated system carry out its process in a way that is not self-destructive. Thus, there are two reasons for providing an automated system with a safety monitoring capability: (1) to protect human workers in the vicinity of the system and (2) to protect the equipment associated with the system.

Safety monitoring means more than the conventional safety measures taken in a manufacturing operation, such as protective shields around the operation or the kinds of manual devices that might be utilized by human workers, such as emergency stop buttons. *Safety monitoring in* an automated system involves the use or sensors to track the system's operation and identify conditions and events that are unsafe or potentially unsafe. The safety monitoring system is programmed to respond to unsafe conditions in some appropriate way. Possible responses to various hazards might include one or more of the following:

• complete stoppage of the automated system
• sounding an alarm
• reducing the operating speed of the process
• taking corrective actions to recover from the safety violation

This last response is the most sophisticated and is suggestive of an intelligent machine performing some advanced strategy. This kind of response is applicable to a variety of possible mishaps, not necessarily confined to safety issues, and is called error detection and recovery.

Sensors are for safety monitoring range from very simple devices to highly sophisticated systems. The following list suggests some of the possible sensors and their applications for safety monitoring:

- Limit switches to detect proper positioning of a part in a work holding device so that the processing cycle can begin.
- Photoelectric sensors triggered by the interruption of a light beam; this could be used to indicate that a part is in the proper position or to detect the presence of a human intruder into the work cell.
- Temperature sensors to indicate that a metal work part is hot enough to proceed with a hot forging operation. If the work part is not sufficiently heated, then the metal's ductility may be too low, and the forging dies might be damaged during the operation.
- Heat or smoke detectors to sense fire hazards.
- Pressure-sensitive floor pads to detect human intruders into the work cell
- Machine vision systems to supervise the automated system and its surroundings.

It should be mentioned that a given safety monitoring system is limited in its ability to respond to hazardous conditions by the possible irregularities that have been foreseen by the system designer. If the designer has not anticipated a particular hazard, and consequently has not provided the system with the sensing capability to detect that hazard, then the safety monitoring system cannot recognize the event if and when it occurs.

2.3.1.2 Maintenance and Repair Diagnostics

Modem automated production systems are becoming increasingly complex and sophisticated, thus complicating the problem of maintaining and repairing them. *Maintenance and repair diagnostics* refers to the capabilities of an automated system to assist in the identification of the source of potential or actual malfunctions and failures of the system. Three modes of operation are typical of a modern maintenance and repair diagnostics subsystem

1. *Status monitoring.* In the status, monitoring mode, the diagnostic subsystem monitor and records the status of key sensors and parameters of the system during normal operation. On request, the diagnostics subsystem can display any of these values and provide an interpretation of current system status, perhaps warning of an imminent failure.

2. Failure diagnostics. The failure diagnostics mode is invoked when a malfunction or failure occurs. Its purpose is to interpret the current values of the monitored variables and to analyze the recorded values preceding the failure so that the cause of the failure can be identified.

3. Recommendation of repair procedure. In the third mode of operation, the subsystem provides a recommended procedure to the repair crew as to the steps that should be taken to effect repairs. Methods for developing the recommendations are sometimes based on the use of expert systems in which the collective judgments of many repair experts are pooled and incorporated into a computer program that uses artificial intelligence techniques.

Status monitoring serves two important functions in machine diagnostics: (1) providing information for diagnosing a current failure and (2) providing data to predict a future malfunction or failure. First, when a failure of the equipment has occurred, it is usually difficult for the repair crew to determine the reason for the failure and what steps should be taken to make repairs. It is often helpful to reconstruct the events leading up to the failure.

The computer is programmed to monitor *and* record the variables and to draw logical inferences from their values about the reason for the malfunction. This diagnosis helps the repair personnel make the necessary repairs and replace the appropriate components. This is especially helpful in electronic repairs where it is often difficult to determine on the basis of visual inspection which components have failed. The second function of status monitoring is to identify signs of an impending failure, so that the affected components can be replaced before failure actually causes the system to go down. These part replacements can be made during the night shift or other time when the process is not operating with the result that the system experiences no loss of regular operation.

2.3.1.3 Error Detection and Recovery

In the operation of any automated system, there are hardware malfunctions and unexpected events that occur during operation. These events can result in costly delays and loss of production until the problem has been corrected and regular operation is restored. Traditionally, equipment malfunctions are corrected by human workers, perhaps with the aid of maintenance and repair diagnostics subroutine. With the increased use of computer control for manufacturing processes, there is a trend toward using the control computer not only to diagnose the malfunctions but also to automatically take the necessary corrective action to restore the system to normal operation. The term *error detection and recovery* is used when the computer performs these functions.

Error Detection. As indicated by the term error detection and recovery consists of two steps: (1) error detection and (2) error recovery. The *error detection* step uses the automated system's available sensor systems to determine when a deviation or malfunction has occurred, correctly interpret the sensor signal(s), and classify the error. Design of the error detection subsystem must begin with the classifications of the possible errors that can occur during system operation. The errors in a manufacturing process tend to be very application specific. They must be anticipated in advance in order to select sensors that will enable their detection.

In analyzing a given production operation, the possible errors can be classified into one of three general categories: (1) random errors, (2) systematic errors and (3) aberrations. *Random errors* occur as a result of the normal stochastic nature of the process. These errors occur when the process is in statistical control. Large variations in part dimensions, even when the production process is in statistical control can cause problems in downstream operations. By detecting these deviations on a part-by-part basis, corrective action can be taken in subsequent operations. *Systematic errors* are those that result from some assignable cause such as a change in raw material properties or a drift in an equipment setting. These errors usually cause the product to deviate from specifications so as to be unacceptable in quality terms. Finally, the third type of error, *aberrations,* results from either an equipment failure or a human mistake. Examples of equipment failures include fracture of a mechanical shear pin, bursts in a hydraulic line, rupture of a pressure vessel, and sudden failure of a cutting tool. Examples of human mistakes include errors in the control program, improper fixture setups, and substitution of the wrong raw materials.

The two main design problems in error detection are: (1) to anticipate all of the possible errors that can occur in a given process and (2) to specify the appropriate sensor systems and associated interpretive software so that the system is capable of recognizing each error. Solving the first problem requires a systematic evaluation of the possibilities under each of the three error classifications. If the error has not been anticipated, then the error detection subsystem cannot correctly detect and identify it.

EXAMPLE 2.2: Error Detection in an Automated Machining Cell
Consider an automated cell consisting of a CNC machine tool, a part storage unit, and a robot for loading and unloading the parts between the machine and the storage unit. Possible errors that might affect this system can be divided into the following categories: (1) machine and process, (2) cutting tools, (3) work holding fixture, (4) part storage unit, and (5) load/unload robot. Develop a list of possible errors (deviations and malfunctions) that might be included in each of these five categories.

Solution: A list of possible errors in the machining cell is presented in Table 2.3.

Table 2.3 Error Detection step in an Automated Machining Cell Error Categories and Possible Malfunctions within Each Category

	Error Categories	Possible Malfunctions
1.	Machine and process	Loss of power, power overload, thermal deflection, cutting temperature too high, vibration, no coolant, chip fouling, wrong part program, defective part
2.	Cutting tools	Tool breakage, tool wear-out, vibration, tool not present, wrong tool
3.	Work holding fixture	Part not in fixture, clamps not actuated, and part dislodged during machining, part deflection during machining, part breakage, chips causing location problems
4.	Part storage unit	Work part not present, wrong work part, oversized or undersized work part
5.	Load/unload robot	Improper grasping of work part, robot drops work part, no part present at pickup

Error Recovery. Error recovery is concerned with applying the necessary corrective action to overcome the error and bring the system back to normal operation. The problem of designing an error recovery system focuses on devising appropriate strategies and procedures that will either correct or compensate for the variety of errors that can occur in the process. Generally, a specific recovery strategy and procedure must be designed for each different error. The types of strategies can be classified as follows:

1. *Make adjustments at the end of the current work cycle.* When the current work cycle is completed, the part program branches to a corrective action subroutine specifically designed for the error detected, executes the subroutine, and then returns to the work cycle program. This action reflects a low level of urgency and is most commonly associated with random errors in the process.
2. *Make adjustments during the current cycle.* This generally indicates a higher level of urgency than the preceding type. In this case, the action to correct or compensate for the detected error is initiated as soon as the error is detected. However, it must be possible to accomplish the designated corrective action while the work cycle is still being executed.
3. *Stop the process to involve corrective action.* In this case, the deviation or malfunction requires that the execution of the work cycle be suspended during corrective action. It is assumed that the system is capable of automatically recovering from the error without human assistance. At the end of the corrective action, the regular work cycle is continued.
4. *Stop the process and call for help.* In this case, the error requiring stoppage of the process cannot be resolved through automated recovery procedures. This situation arises because: (1) the automated cell is not enabled to correct the problem or (2) the error cannot be classified into the predefined list of errors. In either case, human assistance is required to correct the problem and restore the system to fully automated operation. Error detection and recovery requires an interrupt system. When an error in the process is sensed and identified, an interrupt in the current program execution is invoked to branch to the appropriate recovery subroutine, This is done either at the end of the current cycle (type 1 above) or immediately (types 2, 3 and 4). At the completion of the recovery procedure, program execution reverts back to normal operation.

EXAMPLE 2.3: Error Recovery in an Automated Machining Cell
For the automated cell of Example 2.2, develop a list of possible corrective actions that might be taken by the system to address certain of the errors.

Solution: A list of possible corrective actions is presented in Table 2.4.

Table 2.4: Error Recovery in an Automated Machining Cell: Possible Corrective actions That Might Be Taken in Response to Errors Detected during the Operation

Errors Detected	Possible Corrective Actions to Recover
Part dimensions deviating due to thermal deflection of machine tool	Adjust coordinates in part program to compensate (category 1 corrective action)
Part dropped by robot during pickup	Reach for another part (category 2 corrective action)
Part is dimensionally oversized	Adjust part program to take a preliminary machining pass across the work surface (category 2 corrective action)
Chatter (tool vibration)	Increase or decrease cutting speed to change harmonic frequency (category 2 corrective action)
Cutting temperature too high	Reduce cutting speed (category 2 corrective action)
Failure of cutting tool	Replace cutting tool with another sharp tool (category 3 corrective action).

No more parts in parts storage unit	Call operator to resupply starting work parts (category 4 corrective action)
Chips fouling machining operation	Call operator to clear chips from work area (category 4 corrective action)

2.3.2 Summarized of Functions of Automation Systems

Advanced automation functions are made possible by special subroutines included in the programmed of instructions and include safety monitoring, maintenance and repair diagnostics, and error detection and recovery.

Table 2.5: Summarized Functions of Automation System

Function	**Description**
Safety Monitoring	Complex equipment is generally harder to repair and maintain. Maintenance and repair diagnostics uses the system itself to participate in the identification and sourcing of malfunctions and failures of the system. There are three modes: status monitoring—key sensors and parameters are continually examined during system operation, to detect malfunctions as they might occur; failure diagnostics—this is invoked so that the cause of a failure can be identified; and recommendation of repair procedure—suggestions for repair procedures are supplied to the repair crew.
Maintenance and Repair Diagnostics	An error detection and recovery system uses an interrupt subroutine: this allows the main programmed to be interrupted, upon the detection of an error, so that a recovery subroutine may be run instead. In error detection the system's sensors determine, interpret and classify an error when it occurs. Three types of errors may occur: random errors—which result from the normal stochastic nature of the process; systematic errors—which result from an assignable cause in material or equipment; and aberrations—which result from human mistakes or equipment failure. Collecting and classifying all possible errors is the largest problem in error detection.
Error Detection and Recovery	An error detection and recovery system uses an interrupt subroutine: this allows the main programmed to be interrupted, upon the detection of an error, so that a recovery subroutine may be run instead. In error detection the system's sensors determine, interpret and classify an error when it occurs. Three types of errors may occur: random errors—which result from the normal stochastic nature of the process; systematic errors—which result from an assignable cause in material or equipment; and aberrations—which result from human mistakes or equipment failure. Collecting and classifying all possible errors is the largest problem in error detection.

2.4 Levels of Automation

There are various levels at which automation can be applied in the context of the enterprise. A temperature sensor that feeds back information to a regular in a shower is a reasonably low level of automation. On the other hand a high level automation system is required to run a train system in a city. Five levels of automation can be identified and are outlined in Table 2.6. Automation can be examined at five different levels, in a hierarchy that runs from the single device, the machine, the cell or system, the plant, to the enterprise level.

Table 2.6: Automation levels

Level	Description
Device level	The lowest level, it includes hardware components that comprise the machine level, such as actuators and sensors. Control loop devices are predominant here.
Machine level	Hardware at the device level is assembled into individual machines. Control functions at this level include performing the sequence of steps in the programmed of instructions.
Cell or system level	This operates under instructions from the plant level. Consists of a group of machines or workstations connected and supported by a material handling system, computers and other appropriate equipment, including production lines.
Plant level	Factory or production systems level, it receives instruction from the corporate information system and translates them into operational plans for production.
Enterprise level	The highest level, it consists of the corporate information system, and is concerned with all the functions that are necessary to manage and coordinate the entire company.

2.5 Process Industries and Discrete Manufacturing Industries

Having examined the basic elements and contents of automation, we can now look in more detail at industrial control systems. These labels are also useful here to describe the sorts of automation that occurs in each type of industry. Table 2.7 details the two industry types, and their respective automation uses for each level of automation

Table 2.7: Automation levels for process and discrete industries

Level	Process Industry Automation	Discrete Manufacturing Industry Automation
5	Corporate level: Management information systems; strategic planning systems; high-level management of enterprise	Corporate level: Management information systems; strategic planning systems; high-level management of enterprise
4	Plant level: Scheduling; tracking materials; equipment monitoring	Plant or factory level: Scheduling; tracking work-in-process; routing parts through machines; machine utilization
3	Supervisory control level: control and co-ordination of several interconnected unit operations that make up the total process	Manufacturing cell or system level: control and co-ordination of groups of machines and supporting equipment working in co-ordination, including material handling equipment
2	Regulatory control level: Control of unit operations	Machine level: Production machines and workstations for discrete part and product manufacture
1	Device level: Sensors and actuators comprising the basic control loops for unit operations	Device level: Sensors and actuators to accomplish control of machine actions

Significant variations exist at lower levels owing to the differences in devices and equipment used in the two industries. At higher levels, the control of the unit, or the control of machinery provides the basis for the differences between levels two and three. Levels four and five are, in comparison, fairly similar across both process industries and discrete manufacturing industries.

Figure 2.5: Process and discrete industries respectively

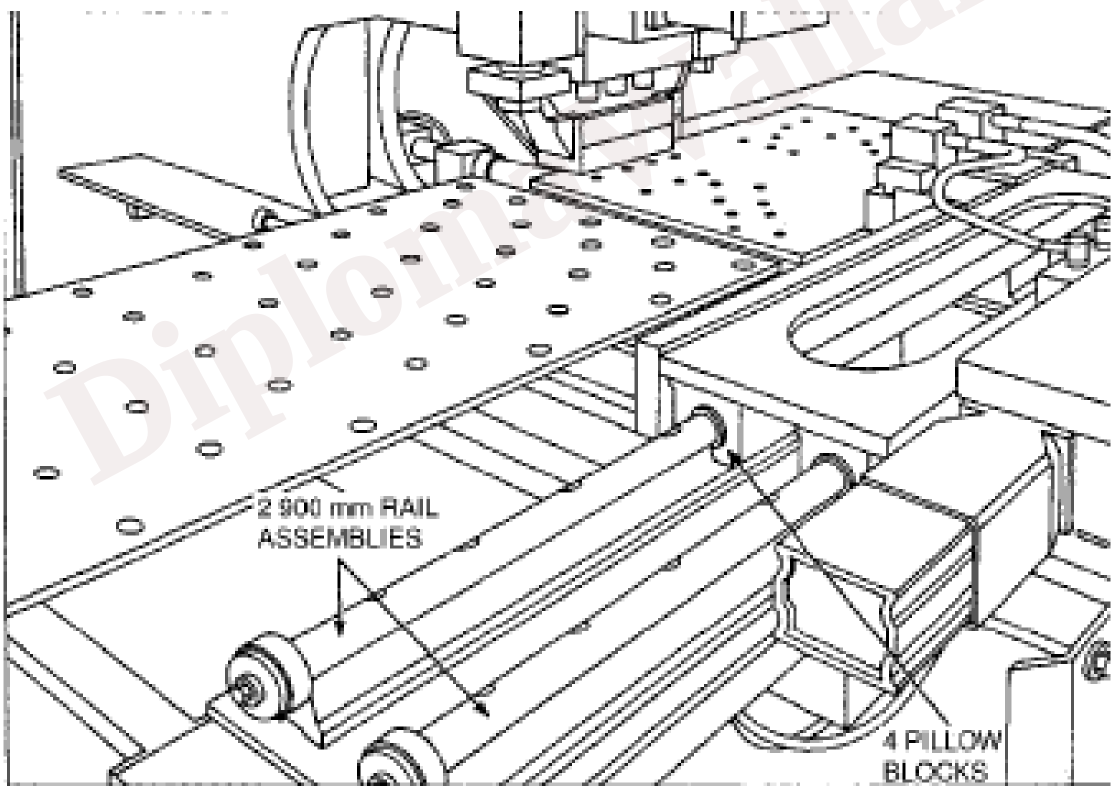
END OF CHAPTER 2

1. What are the three basic elements of automation?
2. Explain the difference between process parameters and process variables.
3. Outline the difference between open and closed loop control.
4. What makes advanced automation functions possible?
5. What is the automation levels hierarchy? Describe it briefly.
6. Draw the block diagram of a typical industrial control system.
7. Consider a motor driven position control system, as commonly found in CNC Machine drives. Identify the main feedback sensors in the system. Identify the major sources of disturbance. How is such a drive different from that of an automated conveyor system?
8. State the major aspect in which sequence/logic control systems differ from analog control systems.
9. State three major functions of a Production Control System.
10. Explore and find out concrete activities for production control under at least two of the above major functions in any typical factory such as a Power Plans or a Steel Plant.
11. Explain examples of application examples indicate automation.
12. Draw the Automation Level from bottom to top and identify the level.
13. Give examples of the above major functional level in any typical factor.
14. What are the types of manufactured products that may be produced by industry?
15. Define the difference between processing operations and assembly operations.
16. What are the ranges of production quantity? How is this related to production plants themselves?
17. What are the types of product variety that may be defined?
18. Explain three (3) modes of operation are typical of a modern maintenance and repair diagnostics subsystem.
19. Explain further on how safety monitoring can be determined as functions of Automation Systems?
20. Given three (3) examples of error recovery in an automated machining cell?
21. Use the internet or a company with which you are familiar to identify one example of an open loop and one example of a closed loop control system. Identify the various elements used in each system and how they operate. Describe how the control system works.

CHAPTER 3
MECHANICAL SYSTEM: COMPONENTS, DYNAMICS AND MODELING

Upon completion of this course, students should be able to:-

- Explain the Elementary Mechanical Concepts
- Describe the Motion Conversion
- Explain the Modeling of Mechanical System
- Define the End Effectors



3.1 Elementary Mechanical Concepts

A mechanism is used to produce mechanical transformation in a machine. This transformation could be any of the following.

- It may convert one speed to a another speed
- It may convert one force to another force.
- It may convert one torque to another torque.
- It may convert force into torque.
- It may convert one angular motion to another angular motion
- It may convert rotation motion into liner motion
- It may convert linear motion into rotation motion

.

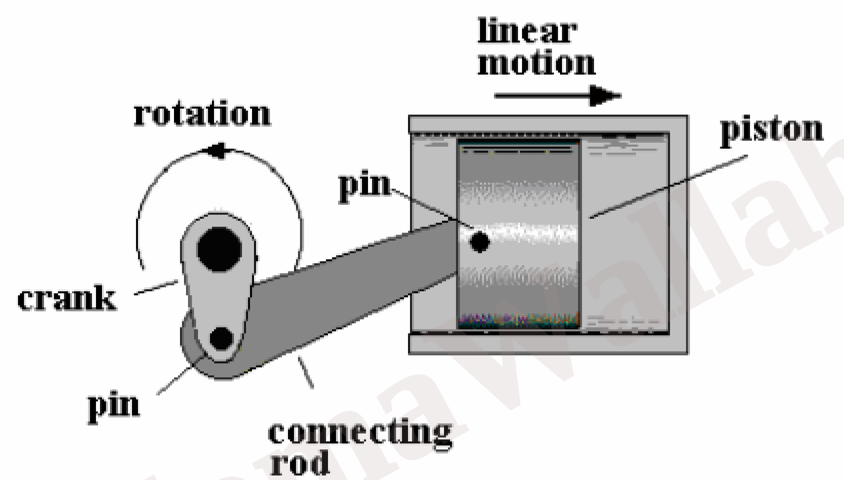


Figure 3.1: A good example is a crank, connecting rod and piston mechanism

If the crank is turned, angular motion is converted into linear motion of the piston and input torque is transformed into force on the piston. If the piston is forced to move, the linear motion is converted into rotary motion and the force into torque. The piston is a sliding joint and this is called PRISMATIC in some fields of engineering such as robotics. The pin joints allow rotation of one part relative to another. These are so called REVOLUTE joints in others areas of engineering.

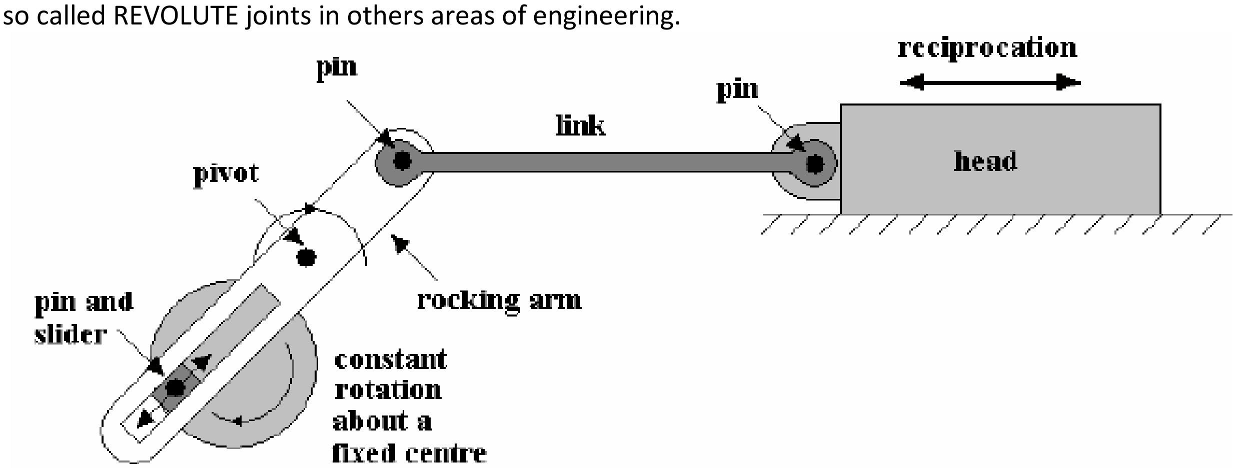


Figure 3.2: Mechanism used in shaping machines and also known as the Whitworth quick-return mechanism.

The input is connected to a motor turning at constant speed. This makes the rocking arm move back and forth and the head (that carries the cutting tool) reciprocates back and forth. Depending on the lengths of the various parts, the motion of the head can be made to move forwards at a fairly constant cutting speed but the return stroke is quick. Note that that the pin and slider must be able to slide in the slot or the mechanism would jam. This causes problems in the solution because of the sliding link.

The main point is that the motion produced is anything but simple harmonic motion and at any time the various parts of the mechanism have a displacement, velocity and acceleration. The acceleration gives rise to inertia forces and this puts stress on the parts in addition to the stress produced by the transmission of power. For example the acceleration of a piston in an internal combustion engine can be enormous and the connecting rod is subjected to high stresses as a result of the inertia as well as due to the power transmission.

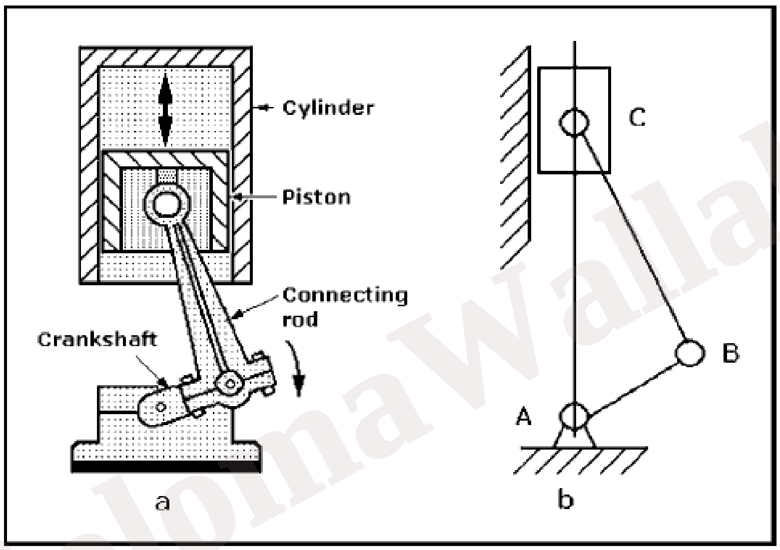


Figure 3.3: Cross section of a cylinder of an internal combustion engine showing piston reciprocation (a) and the skeleton outline of the linkage mechanism that moves the piston (b).

Machines and Mechanisms. *Machines* are devices used to alter, transmit, and direct forces to accomplish a specific objective. A chain saw is a familiar machine that directs forces to the chain with the objective of cutting wood. A *mechanism* is the mechanical portion of a machine that has the function of transferring motion and forces from a power source to an output. It is the heart of a machine. For the chain saw, the mechanism takes power from a small engine and delivers it to the cutting edge of the chain. Figure 3.4 illustrates an adjustable height platform that is driven by hydraulic cylinders. Although the entire device could be called a machine, the parts that take the power from the cylinders and drive the raising and lowering of the platform comprise the mechanism.

A mechanism can be considered rigid parts that are arranged and connected so that they produce the desired motion of the machine. The purpose of the mechanism in Figure 3.4 is to lift the platform and any objects that are placed upon it. *Synthesis* is the process of developing a mechanism to satisfy a set of performance requirements for the machine. *Analysis* ensures that the mechanism will exhibit motion that will accomplish the set of requirements.

Figure 3.4: Adjustable height platform (Courtesy Advance Lifts)

As stated, mechanisms consist of connected parts with the objective of transferring motion and force from a power source to an output. A *linkage* is a mechanism where rigid parts are connected together to form a chain. One part is designated the *frame* because it serves as the frame of reference for the motion of all other parts. The frame is typically a part that exhibits no motion. A popular elliptical trainer exercise machine is shown in Figure 3.5. In this machine, two planar linkages are configured to operate out-of-phase to simulate walking motion, including the movement of arms.

Since the base sits on the ground and remains stationary during operation, the base is considered the frame. *Links* are the individual parts of the mechanism. They are considered rigid bodies and are connected with other links to transmit motion and forces. Theoretically, a true rigid body does not change shape during motion. Although a true rigid body does not exist, mechanism links are designed to minimally deform and are considered rigid. The footrests and arm handles on the exercise machine comprise different links and, along with connecting links, are interconnected to produce constrained motion.

Elastic parts, such as springs, are not rigid and, therefore, are not considered links. They have no effect on the kinematics of a mechanism and are usually ignored during kinematic analysis. They do supply forces and must be included during the dynamic force portion of analysis.
A *joint* is a movable connection between links and allows relative motion between the links. The two *primary joints,* also called *full joints*, are the revolute and sliding joints. The *revolute joint* is also called a *pin* or *hinge joint*. It allows pure rotation between the two links that it connects. The *sliding joint* is also called a *piston* or *prismatic joint*. It allows linear sliding between the links that it connects. Figure 3.6 illustrates these two primary joints, kinematic analysis. They do supply forces and must be included during the dynamic force portion of analysis.

A *joint* is a movable connection between links and allows relative motion between the links. The two *primary joints,* also called *full joints*, are the revolute and sliding joints. The *revolute joint* is also called a *pin* or *hinge joint*. It allows pure rotation between the two links that it connects. The *sliding joint* is also called a *piston* or *prismatic joint*. It allows linear sliding between the links that it connects. Figure 3.6 illustrates these two primary joints.

A cam joint is shown in Figure 3.7a. It allows for both rotation and sliding between the two links that it connects. Because of the complex motion permitted, the cam connection is called a *higher-order joint*, also called *half joint*. A gear connection also allows rotation and sliding between two gears as their teeth mesh. This arrangement is shown in Figure 3.7b. The gear connection is also a higher-order joint.

A *simple link* is a rigid body that contains only two joints, which connect it to other links. Figure 3.7a illustrates a simple link. A *crank* is a simple link that is able to complete a full rotation about a fixed center. A *rocker* is a simple link that oscillates through an angle, reversing its direction at certain intervals. A *complex link* is a rigid body that contains more than two joints. Figure 3.7b illustrates a complex link. A *rocker arm* is a complex link, containing three joints, that is pivoted near its center. A *bell crank* is similar to a rocker arm, but is bent in the center. The complex link shown in Figure 3.7b is a bell crank.

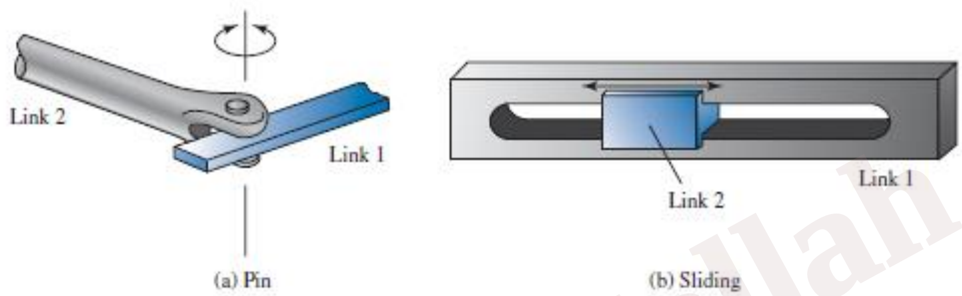


Figure 3.5: Primary joints (a) Pin and (b) Sliding

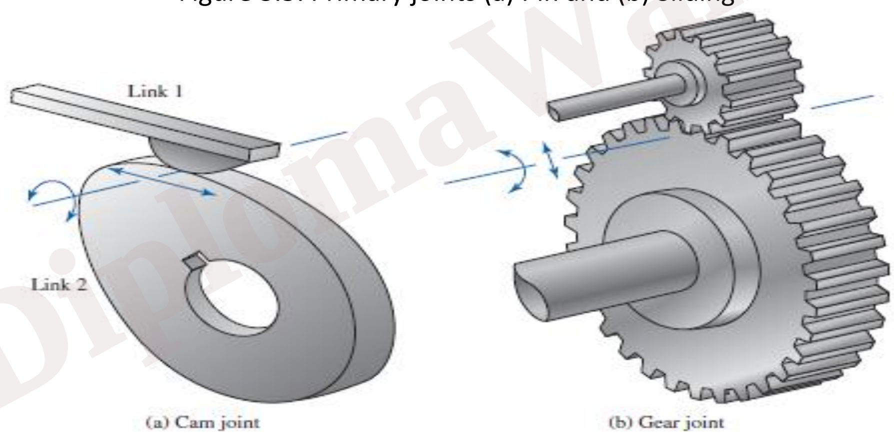


Figure 3.6: Higher-order joints (a) Cam joint and (b) Gear joint.

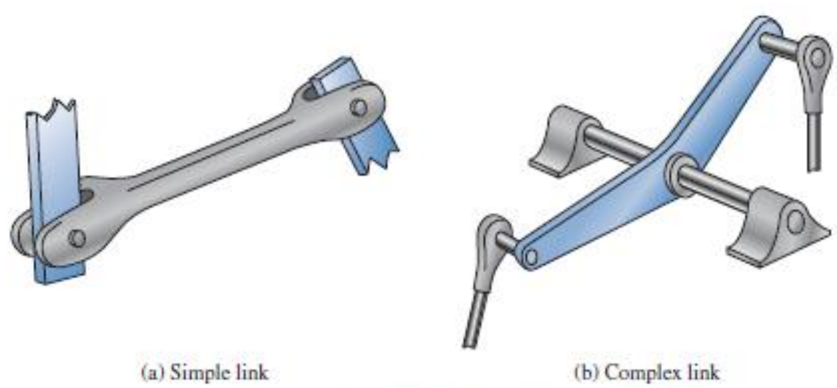


Figure 3.7: Links (a) Simple link and (b) Complex link

Many different kinds of mechanical components are listed in manufacturers' catalogs for speeding the design and assembly of motion control systems. These drawings illustrate what, where, and how one manufacturer's components were used to build specialized systems (Figure 3.8).

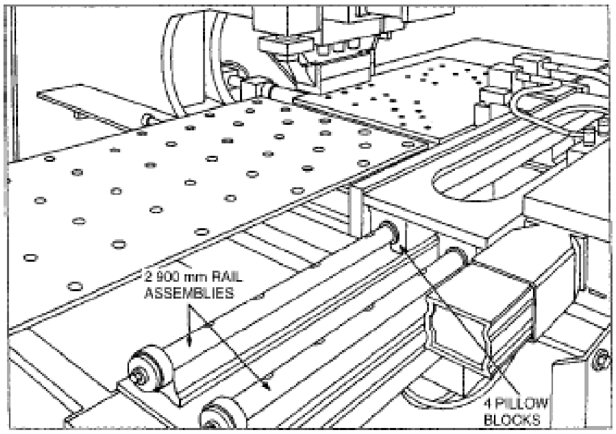


Fig. 1 Punch Press: Catalog pillow blocks and rail assemblies were installed in this system for reducing the deflection of a punch press plate loader to minimize scrap and improve its cycle speed.

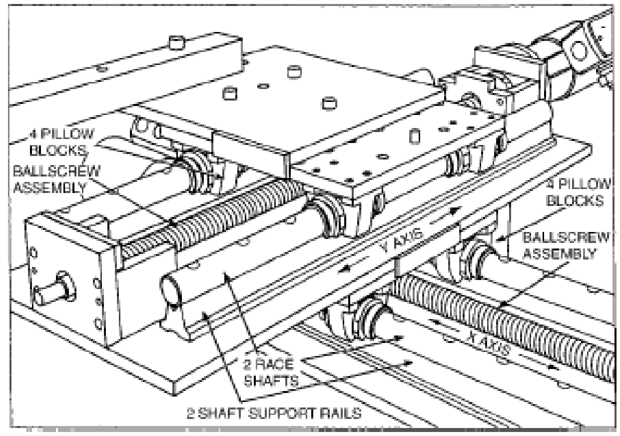


Fig. 2 Microcomputer-Controlled X-Y Table: Catalog pillow blocks, rail guides, and ballscrew assemblies were installed in this rigid system that positions workpieces accurately for precise milling and drilling on a vertical milling machine.

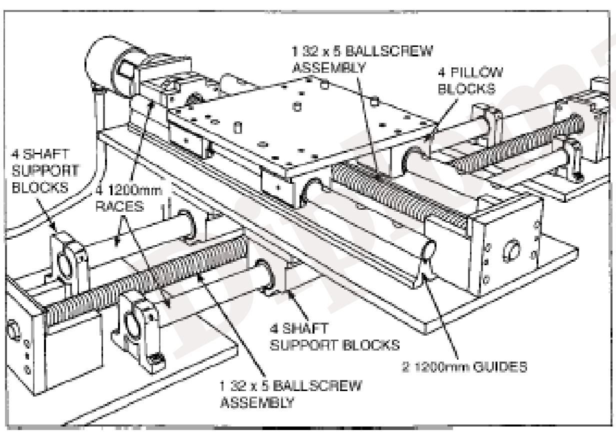


Fig. 3 Pick and Place X-Y System: Catalog support and pillow blocks, ballscrew assemblies, races, and guides were in the assembly of this X-Y system that transfers workpieces between two separate machining stations.

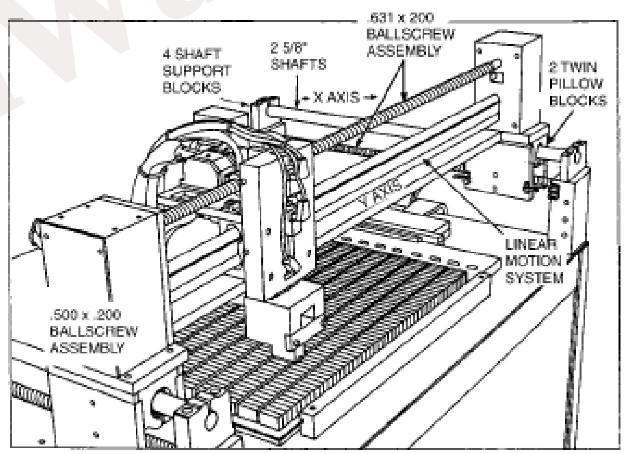


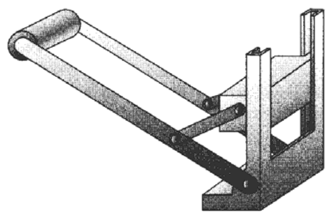
Fig. 4 X-Y Inspection System: Catalog pillow and shaft-support blocks, ballscrew assemblies, and a preassembled motion system were used to build this system, which accurately positions an inspection probe over small electronic components.

Figure 3.8: Different type of Automation in mechanical system

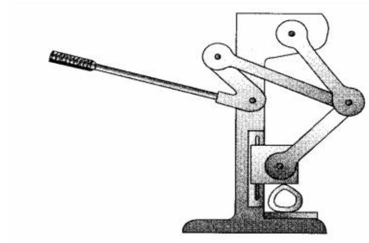
Mechanical systems may be modeled as systems of lumped masses (rigid bodies) or as distributed mass (continuous) systems. The latter are modeled by partial differential equations, whereas the former are represented by ordinary differential equations. Of course, in reality all systems are continuous, but, in most cases, it is easier and therefore preferred to approximate them with lumped mass models and ordinary differential equations.

Machines are mechanical devices used to accomplish work. A mechanism is a heart of a machine. It is the mechanical portion of the machine that has the function of transferring motion and forces from a power source to an output. Mechanism is a system of rigid elements (linkages) arranged and connected to transmit motion in a predetermined fashion. *Mechanism consists of linkages and joints* (Figure 3.9)

The motion of mechanical elements can be described in various dimensions as translational, rotational, or a combination of both. The equations governing the motion of mechanical systems are often directly or indirectly formulated from Newton's law of motion.

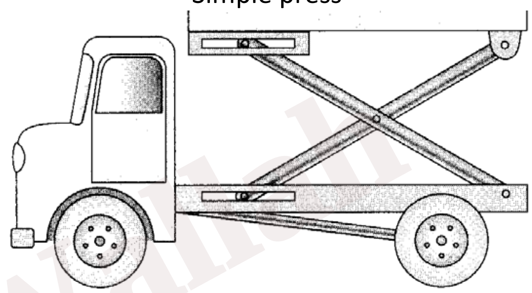


Can crusher

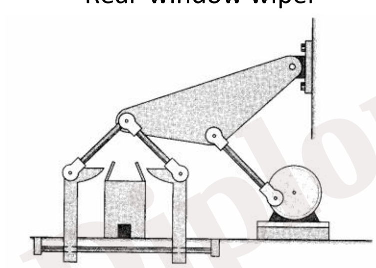


Simple press

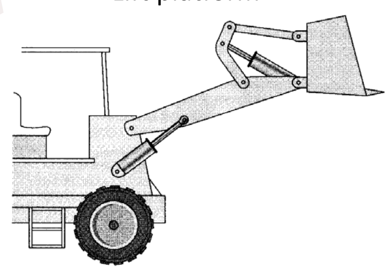
Rear-window wiper



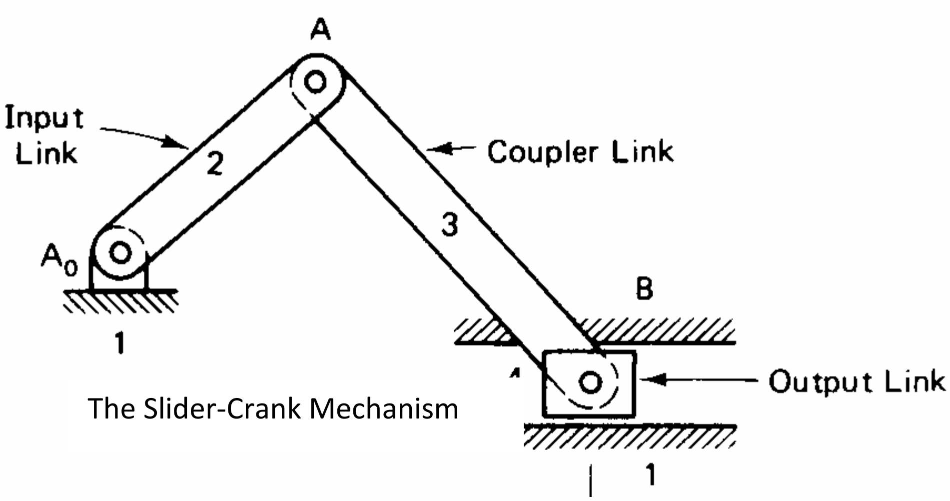
Lift platform



Device to close the top flap of boxes



Loader Device



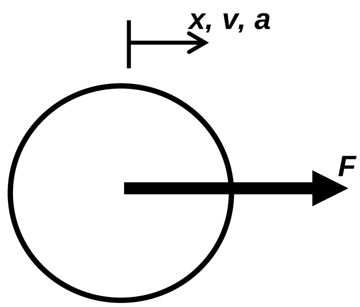
The Slider-Crank Mechanism

Figure 3.9: Examples of Mechanism

3.1.1 Translation or Linear Motion

If the velocity and acceleration of a body are both zero then the body will be static. When forces act on the body they may cause motion. If the applied forces are balanced, and cancel each other out, the body will not accelerate. If the forces are unbalanced then the body will accelerate. If all of the forces act through the center of mass then the body will only translate. Forces that do not act through the center of mass will also cause rotation to occur. This chapter will focus only on translational systems. The equations of motion for translating bodies are shown in Figure 3.10.

These state simply that velocity is the first derivative of position, and acceleration is the first derivative of acceleration. Conversely the acceleration can be integrated to find velocity, and the velocity can be integrated to find position. Therefore, if we know the acceleration of a body, we can determine the velocity and position. Finally, when a force is applied to a mass, acceleration can be found by dividing the net force by the mass.



Equations of motion

$$v(t) = \left(\frac{d}{dt}\right) x\,(t)$$

$$a(t) = \left(\frac{d}{dt}\right)^2 x(t) = \left(\frac{d}{dt}\right) v(t), \quad OR$$

$$x(t) = \int v(t)dt = \iint a(t)dt$$

$$v(t) = \int a(t)dt$$

$$a(t) = \frac{F(t)}{M}$$

where,

x, v, a* = *position , velocity and acceleration

M* = *mass of the body

F* = *an applied force

Figure 3.10: Velocity and acceleration of a translating mass

An example application of these fundamental laws is shown in Figure 2. The initial conditions of the system are supplied (and are normally required to solve this type of problem). These are then used to find the state of the system after a period of time. The solution begins by integrating the acceleration, and using the initial velocity value for the integration constant. So at $t = 0$ the velocity will be equal to the initial velocity. This is then integrated once more to provide the position of the object. As before the initial position is used for the integration constant. This equation is then used to calculate the position after a period of time. Notice that the units are used throughout the calculations. This is good practice for any engineer.

Given an initial (t=0) state of x=5m, v=2m/s, a=3ms^{-2}, find the system state 5 seconds later assuming constant acceleration.

The initial conditions for the system at time t=0 are,

$$x_0 = 5m$$
$$v_0 = 2ms^{-1}$$
$$a_0 = 3ms^{-2}$$

Note: units are very important and should normally be carried through all calculations.

The constant acceleration can be integrated to find the velocity as a function of time.

$$v(t) = \int a_0 dt = a_0 t + C = a_0 t + v_0 \qquad (6)$$

Note:
$$v(t) = a_0 t + C$$
$$v_0 = a_0(0) + C$$
$$v_0 = C$$

Next, the velocity can be integrated to find the position as a function of time.

$$x(t) = \int v(t) dt = \int (a_0 t + v_0) dt = \frac{a_0}{2} t^2 + v_0 t + x_0 \qquad (7)$$

This can then be used to calculate the position of the mass after 5 seconds.

$$x(5) = \frac{a_0}{2} t^2 + v_0 t + x_0$$
$$\therefore \quad = \frac{3ms^{-2}}{2}(5s)^2 + 2ms^{-1}(5s) + 5m$$
$$\therefore \quad = 37.5m + 10m + 5m = 52.5m$$

Figure 3.11: Sample state calculation for a translating mass, with initial conditions

The motion of translation is defined as a motion that takes place along a straight or curved path. The variables that are used to describe translational motion are **acceleration, velocity,** and **displacement.** Newton's law of motion states that the algebraic sum of external forces acting on a rigid body in a given direction is equal to the product of the mass of the body and its acceleration in the same direction. The law can be expressed as

$$\sum_{external} forces = Ma$$

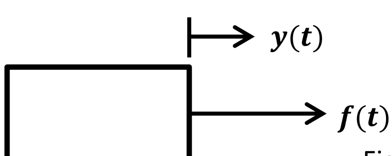


Figure 3.12: Force-mass system

where M denotes the mass, and
a is the acceleration in the direction considered.

Fig. 3.12: Illustrates the situation where a force is acting on a body with mass M.

For translation motion, the following elements are usually involved:

1. ***Mass.*** Mass is *considered a property of an element that stores the kinetic energy of translational motion.* Mass is analogous to the inductance of electric networks. If *W* denotes the weight of a body, then *M* is given by

$$M = \frac{W}{g} \tag{3.1}$$

where *g* is the acceleration of free fall of the body due to gravity (g = 32.174 ft./sec^2 in British units, and g = 9.8066 m/sec^2 in SI units).

The equations of a linear mechanical system are written by first constructing a model of the system containing interconnected linear elements and then by applying Newton's law of motion to the free-body diagram (FBD).

The force equation is written as

$$f(t) = Ma(t) = M\frac{d^2y(t)}{dt^2} = M\frac{dv(t)}{dt} \tag{3.2}$$

where *a(t)* is the acceleration, *v(t)* denotes linear velocity, and *y(t)* is the displacement of mass *M,* respectively.

For linear translational motion, in addition to the mass, the following system elements are also involved.

2. **Linear spring.** In practice, a linear spring may be a model of an actual spring or a compliance of a cable or a belt. In general, *a spring is considered to be an element that stores potential energy.*

$$f(t) = Ky(t) \tag{3.3}$$

where *K* is the spring constant, or simply stiffness. Eq. (3.3) implies that the force acting on the spring is directly proportional to the displacement (deformation) of the spring. The model representing a linear spring element is shown in Fig. 3.13. If the spring is preloaded with a preload tension of *T,* then Eq. (3.3) should be modified to

$$f(t) - T = Ky(t) \tag{3.4}$$

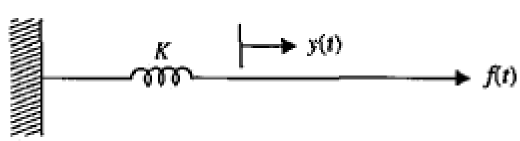


Figure 3.13: Force-spring system

Friction for translation motion. Whenever there is motion or tendency of motion between two physical elements, frictional forces exist. The frictional forces encountered in physical systems are usually of a nonlinear nature. The characteristics of the frictional forces between two contacting surfaces often depend on such factors as the composition of the surfaces, the pressure between the surfaces, and their relative velocity among others, so an exact mathematical description of the frictional force is difficult. Three different types of friction are commonly used in practical systems: **viscous friction, static friction,** and **Coulomb friction.** These are discussed separately in the following paragraphs.

- **Viscous friction.** *Viscous friction represents a retarding force that is a linear relationship between the applied force and velocity.* The schematic diagram element for viscous friction is often represented by a dashpot, such as that shown in Fig. 3.14. The mathematical expression of viscous friction is

$$f(t) = B\frac{dy(t)}{dt} \qquad \textbf{(3.5)}$$

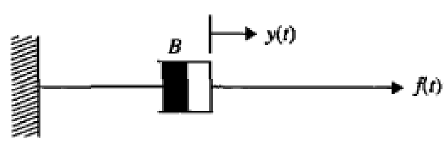


Figure 3.14: Dashpot for viscous friction

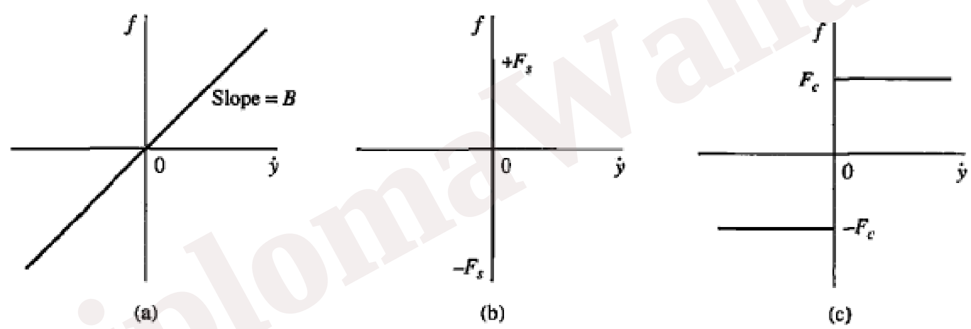


Figure 3.15: Graphical representation of linear and nonlinear frictional forces. (a) Viscous friction. (b) Static friction. (c) Coulomb friction.

where *B* is the viscous frictional coefficient. Fig. 3.15(a) shows the functional relation between the viscous frictional force and velocity.

- **Static friction.** *Static friction represents a retarding force that tends to prevent motion from beginning.* The static frictional force can be represented by the expression

$$f(t) = \pm(F)_s|_{y=0} \qquad \textbf{(3.6)}$$

  which is defined as a frictional force that exists only when the body is stationary but has a tendency of moving. The sign of the friction depends on the direction of motion or the initial direction of velocity. The force-to-velocity relation of static friction is illustrated in Fig. 3.15(b). Notice that, once motion begins, the static frictional force vanishes and other frictions take over.

- **Coulomb friction.** Coulomb friction is a retarding force that has constant amplitude with respect to the change of velocity, but the sign of the frictional force changes with the reversal of the direction of velocity. The mathematical relation for the Coulomb friction is given by

$$f(t) = F_c \frac{\left(\frac{dy(t)}{dx}\right)}{\left|\left(\frac{dy(t)}{dx}\right)\right|} \tag{3.7}$$

where F_c is the Coulomb friction coefficient. The functional description of the friction-to-velocity relation is shown in Fig. 3.15(c).

It should be pointed out that the three types of frictions cited here are merely practical models that have been devised to portray frictional phenomena found in physical systems. They are by no means exhaustive or guaranteed to be accurate. In many unusual situations, we have to use other frictional models to represent the actual phenomenon accurately. One such example is rolling dry friction, which is used to model friction in high-precision ball bearings used in spacecraft systems. It turns out that rolling dry friction has nonlinear hysteresis properties that make it impossible for use in linear system modeling.

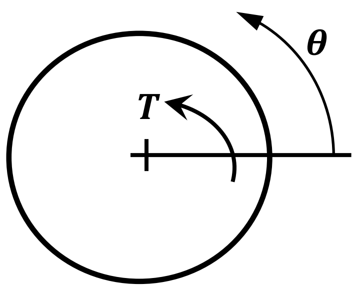
Table 3-1 shows the basic translational mechanical system properties with their corresponding basic SI and other measurement units.

Table 3.1: Basic Translational Mechanical System Properties and Their Units

Parameter	Symbol Used	SI Units	Other Units	Conversion Factors
Mass	M	kilogram (kg)	slug ft/sec^2	1 kg = 1000 g = 2.2046 lb(mass) = 35.274 oz(mass) = 0.06852 slug
Distance	y	meter (m)	ft in	1 m = 3.2808 ft = 39.37 in 1 in. = 25.4 mm 1 ft = 0.3048 m
Velocity	v	m/sec	ft/sec in/sec	
Acceleration	a	m/sec^2	ft/sec^2 in/sec^2	
Force	f	Newton (N)	pound (lb force) dyne	1 N = 0.2248 lb(force) = 3.5969 oz(force) 1 N = 1 kg–m/s^2 1 dyn = 1 g–cm/s^2
Spring Constant	K	N/m	lb/ft	
Viscous Friction Coefficient	B	N/m/sec	lb/ft/sec	

3.1.2 Rotational Motion

The equations of motion for a rotating mass are shown in Figure 3.16. Given the angular position, the angular velocity can be found by differentiating once, the angular acceleration can be found by differentiating again. The angular acceleration can be integrated to find the angular velocity; the angular velocity can be integrated to find the angular position. The angular acceleration is proportional to an applied torque, but inversely proportional to the mass moment of inertia.



equation of motion

$$\omega = \left(\frac{d}{dt}\right)\theta$$

$$\alpha = \left(\frac{d}{dt}\right)\omega = \left(\frac{d}{dt}\right)^2\theta, \quad OR$$

$$\theta(t) = \int \omega(t)dt = \iint \alpha(t)dtdt$$

$$\omega(t) = \int \alpha(t)dt$$

$$\alpha(t) = \frac{T(t)}{J}$$

where

θ, ω, α = position, velocity and acceleration

J = second mass

T = torque applied

Figure 3.16: Basic properties of rotation

The rotational motion of a body can be defined as motion about a fixed axis. The extension of Newton's law of motion for rotational motion states that the *algebraic sum of moments or torque about a fixed axis is equal to the product of the inertia and the angular acceleration about the axis.* Or

$$\sum torques = J\alpha \qquad (3.8)$$

where ***J*** denotes the inertia and ***α*** is the angular acceleration. The other variables generally used to describe the motion of rotation are ***torque T*, angular velocity, ω** and **angular displacement, θ**. The elements involved with the rotational motion are as follows:

1. **Inertia.** *Inertia, J, is considered a property of an element that stores the kinetic energy of rotational motion.* The inertia of a given element depends on the geometric composition about the axis of rotation and its density. For instance, the inertia of a circular disk or shaft, of radius *r* and mass *M,* about its geometric axis is given by

$$J = \frac{1}{2}Mr^2 \qquad (3.9)$$

When a torque is applied to a body with inertia *J,* as shown in Fig. 3.17, the torque equation is written

$$T(t) = J\alpha(t) = J\frac{d\omega(t)}{dt} = J\frac{d^2\theta(t)}{dt^2} \qquad \textbf{(3.10)}$$

where (θ) is the angular displacement; , (ω) the angular velocity; and (α) the angular acceleration.

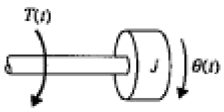


Figure 3.17: Torque-inertia system

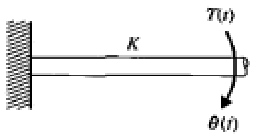


Figure 3.18: Torque torsional spring system

Torsional spring. As with the linear spring for translational motion, a **torsional** spring **constant *K,*** in torque-per-unit angular displacement, can be devised to represent the compliance of a rod or a shaft when it is subject to an applied torque. Fig. 3.18 illustrates a simple torque-spring system that can be represented by the equation

$$T(t) = K\theta(t) \qquad \textbf{(3.11)}$$

If the torsional spring is preloaded by a preload torque of TP, Eq. (3.11) is modified to

$$T(t) - TP = K\theta(t) \qquad \textbf{(3.12)}$$

Friction for rotational motion. The three types of friction described for translational motion can be carried over to the motion of rotation. Table 3.2 shows the SI and other measurement units for inertia and the variables in rotational mechanical systems.

Table 3.2: Basic rotational Mechanical System Properties and Their Units

Parameter	Symbol Used	SI Units	Other Units	Conversion Factors
Inertia	*J*	kg-m^2	slug-ft^2 lb-ft-sec^2 oz-in.-sec^2	1 g-cm = 1.417×10^{-5} oz-in.-sec^2 1 lb-ft-sec^2 = 192 oz-in.-sec^2 = 32.2 lb-ft^2 1 oz-in.-sec^2 = 386 oz-in^2 1 g-cm-sec^2 = 980 g-cm^2
Angular Displacement	*T*	Radian	Radian	1 rad $= \frac{180}{\pi} = 57.3$ deg
Angular Velocity	*O*	radian/sec	radian/sec	1 rpm $= \frac{2\pi}{60}$ = 0.1047 rad/sec 1 rpm = 6 deg/sec
Angular Acceleration	*A*	radian/sec^2	radian/sec^2	
Torque	*T*	(N-m) dyne-cm	lb-ft oz-in.	1 g-cm = 0.0139 oz-in. 1 lb-ft = 192 oz-in. 1 oz-in. = 0.00521 lb-ft
Spring Constant	*K*	N-m/rad	ft-lb/rad	
Viscous Friction Coefficient	*B*	N-m/rad/sec	ft-lb/rad/sec	
Energy	*Q*	J (joules)	Btu Calorie	1 J = 1 N-m 1 Btu = 1055 J 1 cal = 4.184 J

3.1.3 Mechanical Work and Power

One way to consistently partition and connect subsystem models is by using power and energy variables to quantify the system interaction, as illustrated for a mechanical system in Fig. 3.19(a). In this figure, one **port** is shown at which power flow is given by the product of force and velocity, ***F.V*** and another for which power is the product of torque and angular velocity, $\boldsymbol{T \cdot \omega}$. These power conjugate variables (i.e., those whose product yields power) along with those that would be used for electrical and hydraulic energy domains are summarized in Table 3.3. Similar effort (***E***) and flow (***f***) variables can be identified for other energy domains of interest (e.g., thermal, magnetic, chemical). This basis assures energetically correct models, and provides a consistent way to connect system elements together.

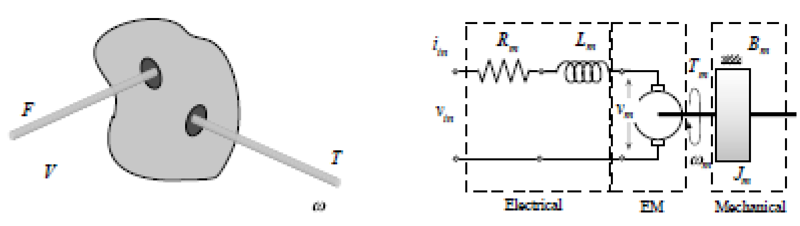
In modeling energetic systems, energy continuity serves as a basis to classify and to quantify systems. Paynter's shows how the energy continuity equation, together with a carefully defined port concept, provides a basis for a generalized modeling framework that eventually leads to a bond graph approach. Paynter's reticulated equation of energy continuity,

$$-\sum_{i=1}^{l} \boldsymbol{P_i} = \sum_{j=1}^{m} \frac{\boldsymbol{dE_j}}{\boldsymbol{dt}} + \sum_{k=1}^{n} (\boldsymbol{P_d})_k \qquad \textbf{(3.13)}$$

concisely identifies the l distinct flows of power, $\boldsymbol{P_i}$, ***m*** distinct stores of energy, $\boldsymbol{E_j}$, and the ***n*** distinct dissipators of energy, $\boldsymbol{P_d}$. Modeling seeks to refine the descriptions from this point. For example, in a simple mass–spring–damper system, the mass and spring store energy, a damper dissipates energy, and

Table 3.3: Power and Energy Variables for Mechanical System

Energy Domain	Effort, e	Flow, f	Power, P
General	e	f	$e \cdot f$ [W]
Translational	Force, F [N]	Velocity, V [m/sec]	$F \cdot V$ [N m/sec, W]
Rotational	Torque, T or τ [N m]	Angular velocity, ω [rad/sec]	$T \cdot \omega$ [N m/sec, W]
Electrical	Voltage, v [V]	Current, i [A]	$v \cdot i$ [W]
Hydraulic	Pressure, P [Pa]	Volumetric flowrate, Q [m^3/sec]	$P \cdot Q$ [W]



(a) (b)

Figure 3.19: Basic interconnection of systems using power variables

the interconnection of these elements would describe how power flows between them. Some of the details for accomplishing these modeling steps are presented in later sections. One way to proceed is to define and categorize types of system elements based on the reticulated energy continuity Eq. (3.13). For example, consider a system made up only of rigid bodies as energy stores (in particular of kinetic energy) for which P_d = 0 (we can add these later), and in general there can be l ports that could bring energy into this purely (kinetic) energy-storing system which has m distinct ways to put energy into the rigid bodies. This is a very general concept, consistent with many other ways to model physical systems. However, it is this foundation that provides for a generalized way to model and integrate different types of energetic systems. The schematic of a permanent-magnet dc (PMDC) motor shown in Fig. 3.19(b) illustrates how power variables would be used to identify interconnection points. This example also serves to identify the need for modeling mechanisms, such as the electromechanical (EM) interaction, that can represent the exchange of energy between two parts of a system. This model represents a simplified relationship between electrical power flow, $v \cdot i$, and mechanical power flow, $T \cdot \omega$, which forms the basis for a motor model. Further, this is an ideal power-conserving relationship that would only contain the power flows in the energy continuity equation; there are no stores or dissipators.

Mechanical work and power play an important role in the design of electromechanical systems. Stored energy in the form of kinetic and potential energy controls the dynamics of the system, whereas dissipative energy usually is spent in the form of heat, which must be closely control.

The position of a particle P of a mass, m in curvilinear motion is specified by the coordinate s measured along its path from a reference point O (Fig. 3.20a). The velocity of the particle is

$$\boldsymbol{v} = \frac{\boldsymbol{ds}}{\boldsymbol{dt}}\boldsymbol{\tau} = \boldsymbol{s}\dot{\boldsymbol{\tau}}, \qquad \textbf{(3.14)}$$

where $\boldsymbol{\tau}$ is the tangential unit vector.

Using the relation $\boldsymbol{v} = \frac{\boldsymbol{d\tau}}{\boldsymbol{dt}}$, the infinitesimal displacement ***dr*** along the path is

$$\boldsymbol{dr} = \boldsymbol{v}\,\boldsymbol{dt} = \frac{\boldsymbol{ds}}{\boldsymbol{dt}}\boldsymbol{\tau}\,\boldsymbol{dt} = \boldsymbol{ds}\,\boldsymbol{\tau} \qquad \textbf{(3.15)}$$

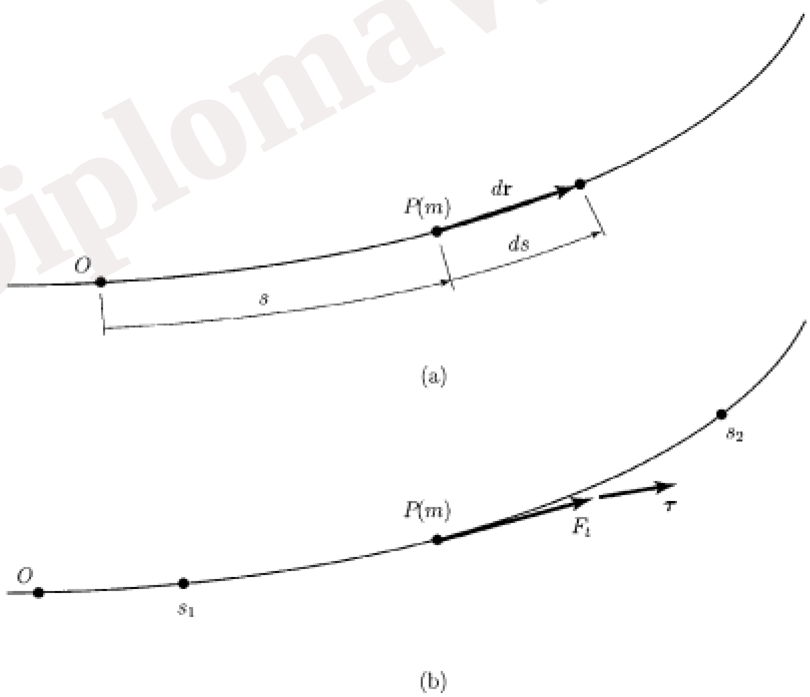


Figure 3.20: Work done by external forces

The work done by the external forces acting on the particle as result of the displacement ***dr*** is

$$\boldsymbol{F}.\boldsymbol{dr} = \boldsymbol{F}.\boldsymbol{ds}\,\boldsymbol{\tau} = \boldsymbol{F}.\boldsymbol{\tau}\,\boldsymbol{ds} = \boldsymbol{F_l ds}, \qquad \textbf{(3.16)}$$

where $F_l = F.\tau$ is the tangential component of the total force.

The work as the particle moves from a position s_1 to position s_2 is (Fig.3.20b)

$$\boldsymbol{U_{12}} = \int_{S_1}^{s_2} \boldsymbol{F_l\ ds}. \qquad \textbf{(3.17)}$$

The work is equal to the integral of tangential component of the total force with respect to distance along the path. Components of force perpendicular to the path do not do any work. The work done by the external force acting on a particle during an infinitesimal displacement ***dr*** is

$$\boldsymbol{dU} = \mathbf{F}.\mathbf{dr}. \qquad \textbf{(3.18)}$$

The power, ***P***, is the rate at which work is done. The power ***P*** is obtained by dividing the expression of the work by the interval of time ***dt*** during which the displacement takes place:

$$\boldsymbol{P} = \boldsymbol{F}.\frac{\boldsymbol{dr}}{\boldsymbol{dt}} = \boldsymbol{F}.\boldsymbol{v} \qquad \textbf{(3.19)}$$

In SI units, the power is expressed in newton meters per second, which is joules per second (J/s) or watts (W). In U.S. customary units, power is expressed in foot pounds per second or in house power (hp), which is 746W or approximately 550ft Ib/s.

The power is also the rate of change of the kinetic energy of the object

$$\boldsymbol{P} = \frac{\boldsymbol{d}}{\boldsymbol{dt}}\left(\frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{mv}^{\mathbf{2}}\right) \qquad \textbf{(3.20)}$$

Work Done On A Particle By a Linear Spring

A linear spring connects a particle ***P*** of mass m to a fixed support (Fig 3.21). The force exerted on the particle is

$$\boldsymbol{F} = -\boldsymbol{k}(\boldsymbol{r} - \boldsymbol{r}_0)\boldsymbol{u}_r \qquad \textbf{(3.21)}$$

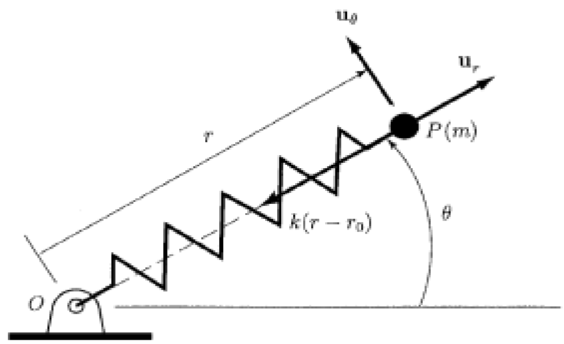


Figure 3.21: Linear spring

Where $\boldsymbol{k}$ is the spring constant, $\boldsymbol{r_0}$ is the unstretched length of the spring and $\boldsymbol{u_r}$is the polar unit vector. If we use the expression for the velocity in polar coordinates, the vector ***dr = vdt*** is

$$\boldsymbol{dr} = \left(\frac{\boldsymbol{dr}}{\boldsymbol{dt}}\boldsymbol{u}_r + \boldsymbol{r}\frac{\boldsymbol{d\theta}}{\boldsymbol{dt}}\boldsymbol{u}_\theta\right)\boldsymbol{dt} = \boldsymbol{dru}_r + \boldsymbol{rd\theta u}_\theta \qquad \textbf{(3.22)}$$

$$\boldsymbol{F}.\boldsymbol{dr} = [-\boldsymbol{k}(\boldsymbol{r} - \boldsymbol{r}_0)\boldsymbol{u}_r] \cdot (\boldsymbol{dru}_r + \boldsymbol{nd\theta u}_\theta) = -\boldsymbol{k}(\boldsymbol{r} - \boldsymbol{r}_0)\boldsymbol{dr} \qquad \textbf{(3.23)}$$

One many express the work done by the spring in terms of its stretch, defined by $\boldsymbol{\delta} = \boldsymbol{r} - \boldsymbol{r}_0$. In terms of these variables, **F**· $\boldsymbol{dr} = -\boldsymbol{k\delta}$, and the work is

$$\boldsymbol{U}_{12} = \int_{r_2}^{r_1} \boldsymbol{F} \cdot \boldsymbol{dr} = \int_{\delta_1}^{\delta_2} -\boldsymbol{k\delta} = -\frac{1}{2}\,\boldsymbol{k}\left(\boldsymbol{\delta}_2^2 - \boldsymbol{\delta}_1^2\right), \qquad \textbf{(3.24)}$$

where $\delta_1\ and\ \delta_2$ are the values of the stretech at the initial and final position.

Work Done On A Particle By Weight

A particle ***P*** of mass ***m*** (Fig. 3.22) moves from position 1 with coordinates (x_1, y_1, z_1) to position 2 with coordinates (x_2, y_2, z_2) in a Cartesian reference frame with the y axis upword. The force exerted by the weight is

$$\boldsymbol{F = -mgj} \qquad \textbf{(3.25)}$$

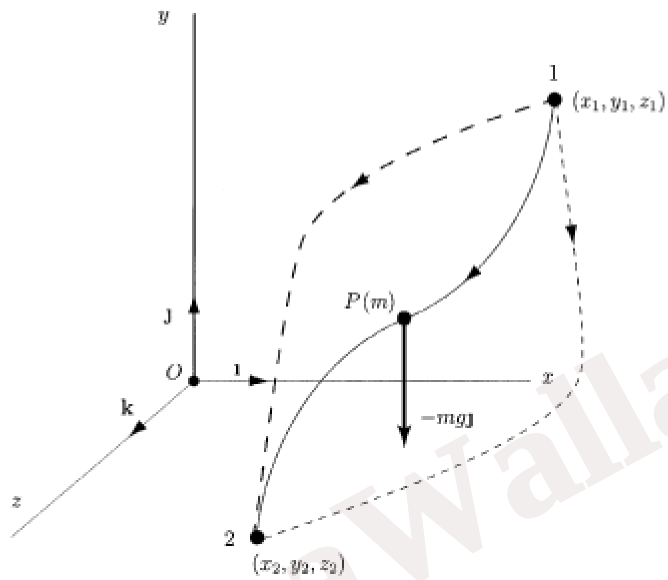


Figure 3.22: Work done on a particle

Because ***v = dr/dt,*** the expression for the vector ***dr*** is

$$\boldsymbol{dr} = \left(\frac{\boldsymbol{dx}}{\boldsymbol{dt}}\boldsymbol{i} + \frac{\boldsymbol{dy}}{\boldsymbol{dt}}\boldsymbol{j} + \frac{\boldsymbol{dz}}{\boldsymbol{dt}}\boldsymbol{k}\right)\boldsymbol{dt} = \boldsymbol{dxi} + \boldsymbol{dyj} + \boldsymbol{dzk}. \qquad \textbf{(3.26)}$$

The dot product of ***F*** and ***dr*** is

$$\boldsymbol{F.dr} = (\boldsymbol{-mgj})\cdot(\boldsymbol{dxi} + \boldsymbol{dyj} + \boldsymbol{dzk}) = \boldsymbol{-mg\,dy}. \qquad \textbf{(3.27)}$$

The work done by as ***P*** moves from position 1 to position 2 is

$$\boldsymbol{U_{12}} = \int_{\tau_1}^{\tau_2} \boldsymbol{F \cdot dr} = \int_{y_1}^{y_2} \boldsymbol{-mg\,dy} = \boldsymbol{-mg\,(y_2 - y_1)}. \qquad \textbf{(3.28)}$$

The work is the product of the weight and the change in height of the particle. The work done is negative if the height increase and positive if it decreases. The work done is the same no matter what path particle follows from position 1 to position 2. To determine the work done by the weight of the particle, only the relative heights of the initial and final positions must be known.

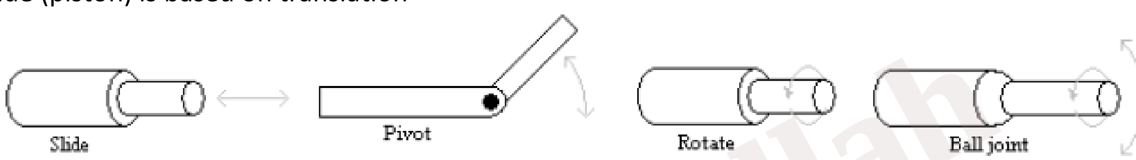
3.2 Motion Conversion

Mechanical systems often include mechanisms such as levers, gears, linkages, cams, chains, and belts. They all serve a common basic function, the transformation of the motion of an input member into the kinematically related motion of an output member.

The actual system may be simplified in many cases to a fictitious but dynamically equivalent one. This is accomplished by “referring” all the elements (masses, springs, dampers) and driving inputs to a single location, which could be the input, the output, or some selected interior point of the system. A single equation can then be written for this equivalent system, rather than having to write several equations for the actual system. This process is not necessary, but often speeds the work and reduces errors.

Link is basic kinematic component. Link is rigid moving part. Linkage is set of links combined via joints. Joint is movable connection. 2 types of joints:

1. Pivot is based on rotation
2. Slide (piston) is based on translation



Kinematic model is kinematic chain and geometry that surrounds chain

1. Rigid geometry does no deform
2. Flexible geometry deforms and is called skin or envelope

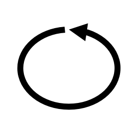
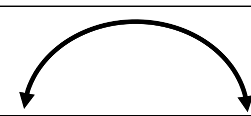
Simple machine alters magnitude and/or direction of applied force

1. Lever
2. Pulley
3. Inclined plane
4. Wedge
5. Screw

Machine (mechanism is system of connected (usually) rigid bodies that alter transmit, and direct force in predetermined way.

Machine parts characterized by their motion

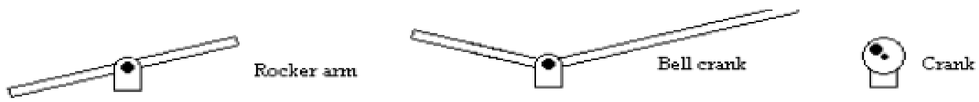
- Translatory motion follows straight motion
- Rotary motion follows curve or circular path
- Motion can be
    - Continuous and omnidirectional
    - Reciprocal and oscillatory

Motion	Translatory	Rotary
Continuous		
Reciprocal		

Device for transforming motion:

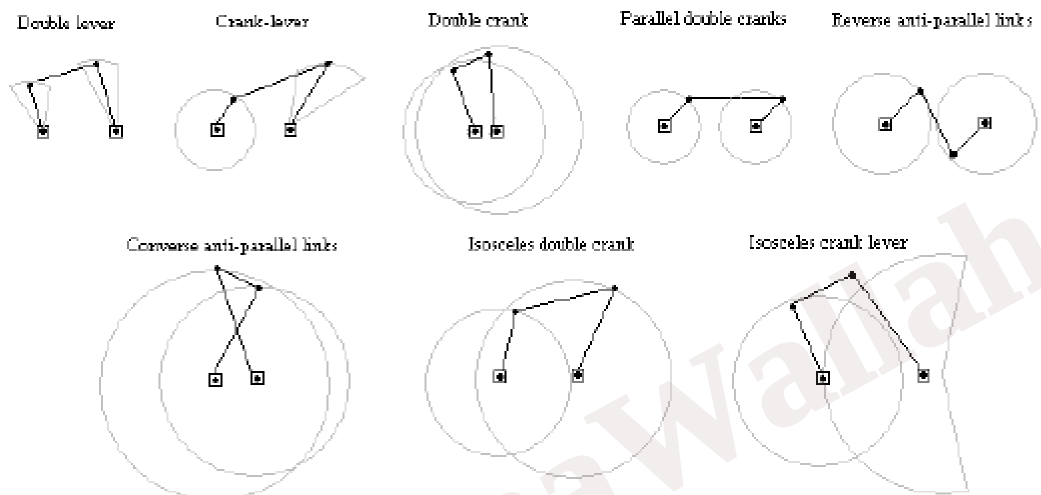
- rocker arm reverses motion
- Bell crank amplifies motion

- Crank transmits rotary motion



8 basic 4 bar linkages associated with cranks. They are characterized by

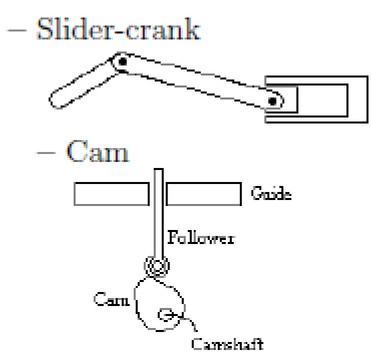
1. Number of crank
2. Conversion of motion
3. Velocity ratio of driver and follower



Linkage	Number of crank	Motion converted	Velocity ratio
Double lever	0	Reciprocating arc to reciprocating arc	Variable
Crank lever (crank rocker)	1	Continuous circular to reciprocating arc	Variable
Double crank (drag link)	2	Continuous circular to Continuous circular to	Variable
Parallel double crank	2	Continuous circular to reverse continuous circular	Constant equal velocity
Reverse anti-parallel	2	Continuous circular to reverse continuous circular	Constant opposite velocity
Converse anti-parallel	2	Continuous circular to reverse Continuous circular	Constant equal velocity
Isosceles double crank	2	Continuous circular to continuous circular to	Constant variable velocity
Isosceles crank lever	1	Continuous circular to reciprocating arc	Variable

Devices for converting one type of motion to another:

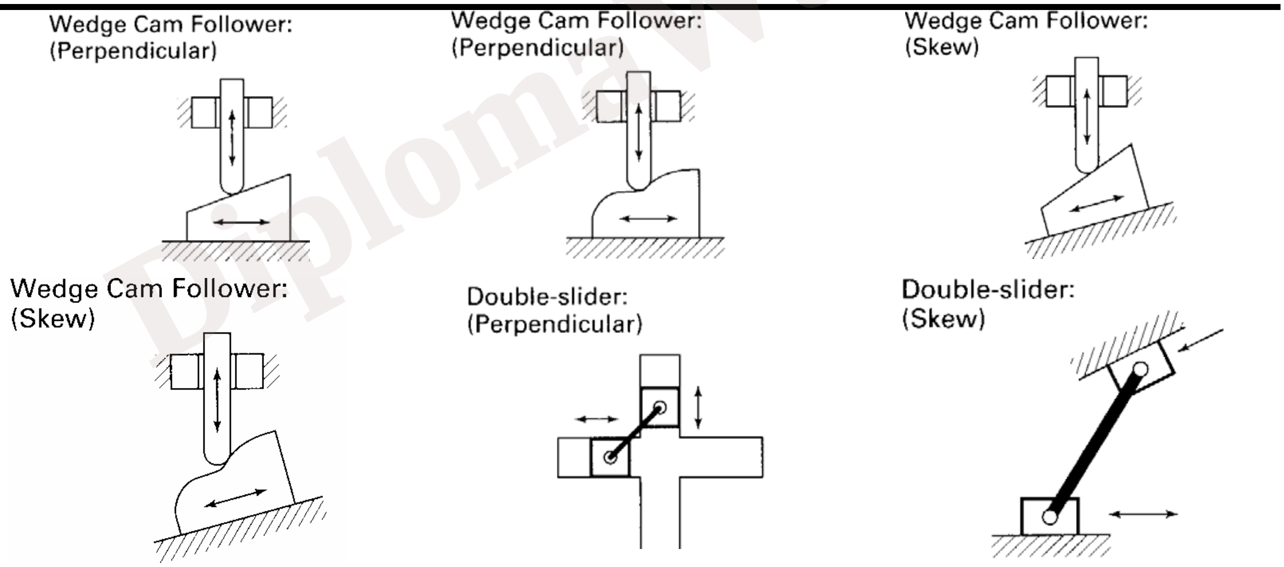
		TO			
		Continuous		Reciprocating	
From		Trans	Rotary	Trans	Rotary arc
Continuous	Rotary		Double crank Parallel crank	Cam Crank slider	Crank lever
Reciprocating	Trans		Slider crank	Rocker arm	Slider lever (toggle)
	Rotary arc		Lever crank	Lever slider	Double lever



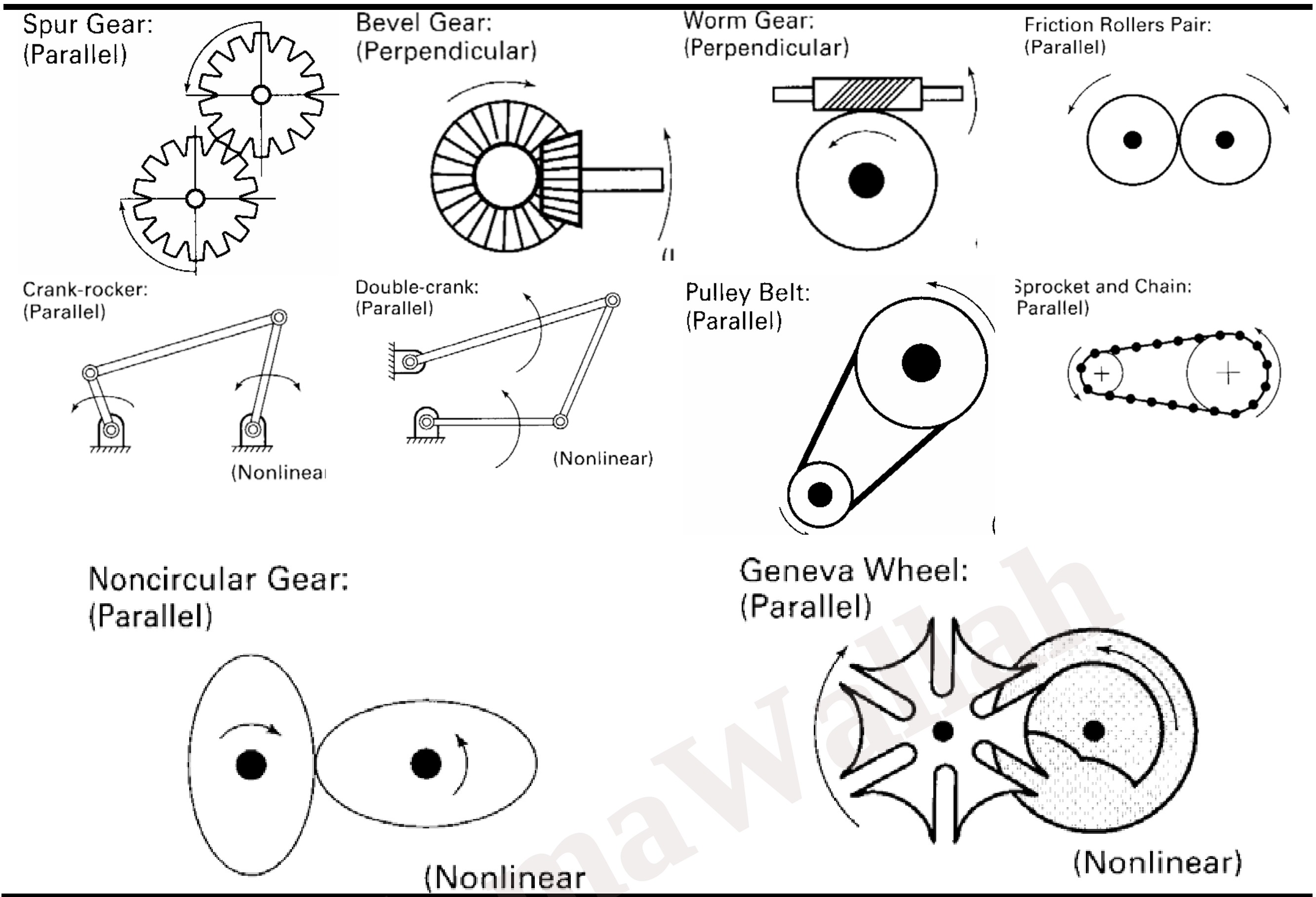
- Connecting rod converts reciprocating to rotary motion

- Lifts rod up and down
- 2 types :
  - Continuous motion that is discontinuous at extremes
  - Continuous acceleration

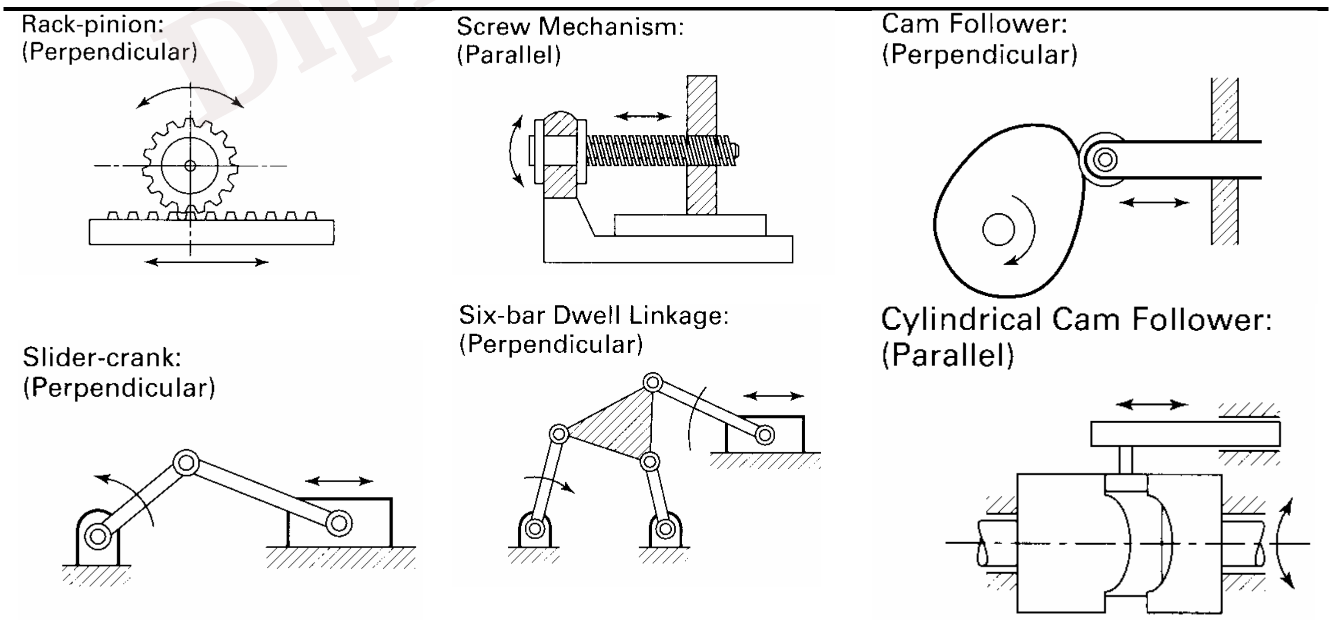
1. Translation motion to translation motion



2. Rotational to Rotational



3. Rotation to Translation



3.2.1 Rotary to Rotary Motion Conversion (Gear trains, Levers and Timing Belts)

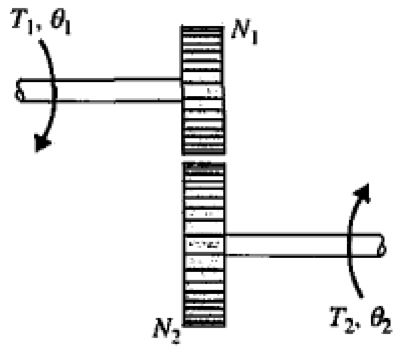


Figure 3.23: Gear train

A gear train, lever, or timing belt over a pulley is a mechanical device that transmits energy from one part of the system to another in such a way that force, torque, speed, and displacement may be altered. These devices can also be regarded as matching devices used to attain maximum power transfer. Two gears are shown coupled together in Fig. 3-23. The inertia and friction of the gears are neglected in the ideal case considered. The relationships between the torques T_1 and T_2, angular displacement θ_1 and θ_2, and the teeth numbers N_1 and N_2 of the gear train are derived from the following facts:

1. The number of teeth on the surface of the gears is proportional to the radius r_1 and r_2 of the gears; that is,

$$r_1 N_2 = r_2 N_1 \qquad (3.29)$$

2. The distance traveled along the surface of each gear is the same. Thus,

$$\theta_1 r_1 = \theta_2 r_2 \qquad (3.30)$$

3. The work done by one gear is equal to that of the other since there are assumed to be no losses. Thus,

$$T_1 \theta_1 = T_2 \theta_2 \qquad (3.31)$$

If the angular velocities of the two gears ω_1 and ω_2 are brought into the picture, Eqs. (3-29) through (3-31) lead to

$$\frac{T_1}{T_2} = \frac{\theta_2}{\theta_1} = \frac{N_1}{N_2} = \frac{\omega_2}{\omega_1} = \frac{r_1}{r_2} \qquad (3.32)$$

In practice, gears do have inertia and friction between the coupled gear teeth that often cannot be neglected. An equivalent representation of a gear train with viscous friction, Coulomb friction, and inertia considered as lumped parameters is shown in Fig. 3-24, where T denotes the applied torque, T_1 and T_2 are the transmitted torque, F_{c1} and F_{c2} are the Coulomb friction coefficients, and B_1 and B_2 are the viscous friction coefficients. The torque equation for gear 2 is

$$T_2(t) = J_2 \frac{d^2\theta_2(t)}{dt^2} + B_2 \frac{d\theta_2(t)}{dt} + F_{c2} \frac{\omega_2}{|\omega_2|} \qquad (3.33)$$

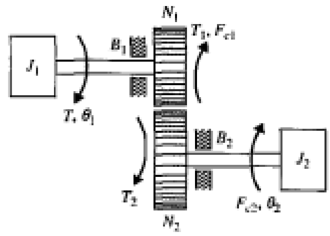


Figure 3.24: Gear train with friction and inertia

The torque equation on the side of gear 1 is

$$\boldsymbol{T(t) = J_1 \frac{d^2\theta_1(t)}{dt^2} + B_1 \frac{d\theta_1(t)}{dt} + F_{c1}\frac{\omega_1}{|\omega_2|} + T_1(t)} \qquad \textbf{(3.33)}$$

Using Eq. (3-32), Eq. (3-33) is converted to

$$\boldsymbol{T_1(t) = \frac{N_1}{N_2}T_2(t) = \left(\frac{N_1}{N_2}\right)^2 J_2 \frac{d^2\theta_1(t)}{dt^2} + \left(\frac{N_1}{N_2}\right)^2 B_2 \frac{d\theta_1(t)}{dt} + \frac{N_1}{N_2} F_{c2}\frac{\omega_2}{|\omega_2|}} \qquad \textbf{(3.34)}$$

Eq. (3-34) indicates that it is possible to reflect inertia, friction, compliance, torque, speed, and displacement from one side of a gear train to the other. The following quantities are obtained when reflecting from gear 2 to gear 1:

Inertia: $\left(\frac{N_1}{N_2}\right)^2 J_2$

Viscous-friction coefficient: $\left(\frac{N_1}{N_2}\right)^2 B_2$

Torque: $\frac{N_1}{N_2} T_2$ **(3.35)**

Angular displacement: $\frac{N_1}{N_2}\theta_2$

Angular velocity: $\frac{N_1}{N_2}\omega_2$

Coulomb friction torque: $\frac{N_1}{N_2} F_{c2}\frac{\omega_2}{|\omega_2|}$

Similarly, gear parameters and variables can be reflected from gear 1 to gear 2 by simply interchanging the subscripts in the preceding expressions. If a torsional spring effect is present, the spring constant is also multiplied by (N_1/N_2) in reflecting from gear 2 to gear

1. Now substituting Eq. (3-34) into Eq. (3-33), we get

$$\boldsymbol{T(t) = J_{le}\frac{d^2\theta_1(t)}{dt^2} + B_{le}\frac{d\theta_1(t)}{dt} + T_\theta} \qquad \textbf{(3.36)}$$

where

$$\boldsymbol{J_{le} = J_1 + \left(\frac{N_1}{N_2}\right)^2 J_2} \qquad \textbf{(3.37)}$$

$$B_{le} = B_1 + \left(\frac{N_1}{N_2}\right)^2 B_2 \quad \textbf{(3.38)}$$

$$T_\theta = F_{c1}\frac{\omega_1}{|\omega_2|} + \frac{N_1}{N_2}F_{c2}\frac{\omega_2}{|\omega_2|} \quad \textbf{(3.39)}$$

Example: 3.1: Given a load that has inertia of 0.05oz-in.-sec^2 and a Coulomb friction of 2 oz-in, find the inertia and frictional torque reflected through a 1:5 gear train (N_1:N_2 = 1/5, with N_2 on the load side). The reflected inertia on the side of N_1 is $(1/5)^2$ x0.05 = 0.002 oz-in.-sec^2. The reflected Coulomb friction is (1/5) x 2 = 0.4 oz-in.

3.2.2 Rotary to Linear Motion Conversion

In motion-control systems, it is often necessary to convert rotational motion into translational motion. For instance, a load may be controlled to move along a straight line through a rotary motor-and-lead screw assembly, such as that shown in Fig. 3-25. Fig. 3-26 shows a similar situation in which a rack-and-pinion assembly is used as a mechanical linkage. Another familiar system in motion control is the control of a mass through a pulley by a rotary motor, as shown in Fig. 3 27. The systems shown in Figs. 3-25, 3-26, and 3-27 can all be represented by a simple system with an equivalent inertia connected directly to the drive motor. For instance, the mass in Fig. 3-27 can be regarded as a point mass that moves about the pulley, which has a radius *r.* By disregarding the inertia of the pulley, the equivalent inertia that the motor sees is

$$J = Mr^2 = \frac{W}{g} r^2 \quad \textbf{(3.39)}$$

If the radius of the pinion in Fig. 3-26 is r, the equivalent inertia that the motor sees is also given by Eq. (3-39).
Now consider the system of Fig. 3-25. The lead of the screw, L, is defined as the linear distance that the mass travels per revolution of the screw. In principle, the two systems in Fig. 3-26 and Fig. 3-27 are equivalent. In Fig. 3-26, the distance traveled by the mass per revolution of the pinion is 27rr. By using Eq. (3-39) as the equivalent inertia for the system of Fig. 3-25, we have

$$J = \frac{W}{g}\left(\frac{L}{2\pi}\right)^2 \quad \textbf{(3.40)}$$

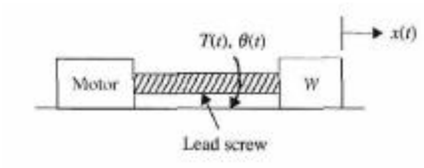


Figure 3.25: Rotary-to-linear motion control system (lead screw)

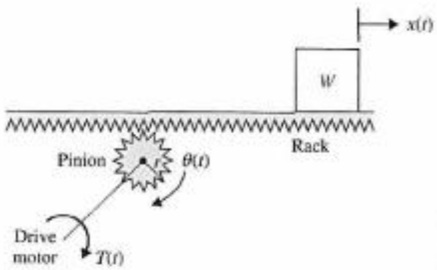


Figure 3.26: Rotary-to-linear motion control system (rack and pinion)

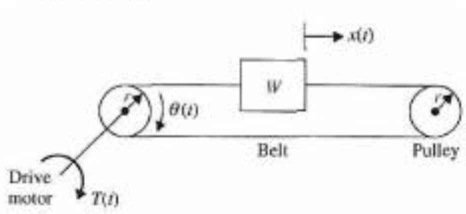


Figure 3.27:Rotary-tolinera motion control system (belt and pulley)

3.2.3 Linkages

A *linkage* is a mechanism formed by connecting two or more of a force or make two or more objects move at the same time. Many different fasteners are used to connect linkages together yet allow them to move freely such as pins, end-threaded bolts with nuts, and loosely fitted rivets. There are two general classes of linkages: *simple planar linkages* and more complex *specialized linkages*; both are capable of performing tasks such as describing straight lines or curves and executing motions at differing speeds. The names of the linkage mechanisms given here are widely but not universally accepted in all textbooks and references.

Linkages can be classified according to their primary functions:

- *Function generation*: the relative motion between the links connected to the frame
- *Path generation*: the path of a tracer point
- *Motion generation*: the motion of the coupler link

Simple Planar Linkages

Four different simple planar linkages shown in Fig. 3.28 are identified by function:

• ***Reverse-motion linkage***, Fig. 3.28a, can make objects or force move in opposite directions; this can be done by using the input link as a lever. If the fixed pivot is equidistant from the moving pivots, output link movement will equal input link movement, but it will act in the opposite direction. However, if the fixed pivot is not centered, output link movement will not equal input link movement. By selecting the position of the fixed pivot, the linkage can be designed to produce specific mechanical advantages. This linkage can also be rotated through 360°.

• ***Push-pull linkage***, Fig. 3.28b, can make the objects or force move in the same direction; the output link moves in the same direction as the input link. Technically classed as a four-bar linkage, it can be rotated through 360° without changing its function.

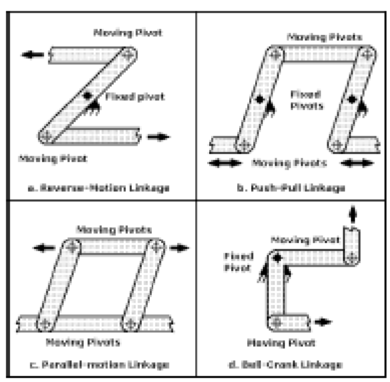


Figure 3.28: Functions of four basic planar linkage mechanisms

• ***Parallel-motion linkage***, Fig. 3.28c, can make objects or forces move in the same direction, but at a set distance apart. The moving and fixed pivots on the opposing links in the parallelogram must be equidistant for this linkage to work correctly. Technically classed as a four-bar linkage, this linkage can also be rotated through 360° without changing its function. Pantographs that obtain power for electric trains from overhead cables are based on parallel-motion linkage. Drawing pantographs that permit original drawings to be manually copied without tracing or photocopying are also adaptations of this linkage; in its simplest form it can also keep tool trays in a horizontal position when the toolbox covers are opened.

• ***Bell-crank linkage***, Fig. 3.28d, can change the direction of objects or force by 90°. This linkage rang doorbells before electric clappers were invented. More recently this mechanism has been adapted for bicycle brakes. This was done by pinning two bell cranks bent 90° in opposite directions together to form tongs. By squeezing the two handlebar levers linked to the input ends of each crank, the output ends will move together. Rubber blocks on the output ends of each crank press against the wheel rim, stopping the bicycle. If the pins which form a fixed pivot are at the midpoints of the cranks, link movement will be equal. However, if those distances vary, mechanical advantage can be gained.

Specialized Linkages

In addition to changing the motions of objects or forces, more complex linkages have been designed to perform many specialized functions: These include drawing or tracing straight lines; moving objects or tools faster in a retraction stroke than in an extension stroke; and converting rotating motion into linear motion and vice versa. The simplest specialized linkages are four-bar linkages. These linkages have been versatile enough to be applied in many different applications. Four-bar linkages actually have only three moving links but they have one fixed link and four pin joints or pivots. A useful mechanism must have at least four links but closed-loop assemblies of three links are useful elements in structures. Because any linkage with at least one fixed link is a mechanism, both the parallel-motion and push-pull linkages mentioned earlier are technically machines.

Four-bar linkages share common properties: three rigid moving links with two of them hinged to fixed bases which form a *frame.* Link mechanisms are capable of producing rotating, oscillating, or reciprocating motion by the rotation of a crank.

Linkages can be used to convert:

- Continuous rotation into another form of continuous rotation, with a constant or variable angular velocity ratio
- Continuous rotation into oscillation or continuous oscillation into rotation, with a constant or variable velocity ratio
- One form of oscillation into another form of oscillation, or one form of reciprocation into another form of reciprocation, with a constant or variable velocity ratio

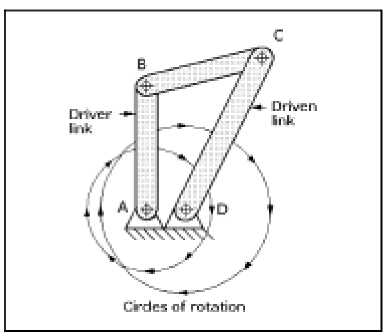


Figure 3.29: Four-bar drag-link mechanism. Both the driver link AB and driven link CD rotate through 360°. Link AD is the foundation link

There are four different ways in which four-bar linkages can perform inversions or complete revolutions about fixed pivot points. One pivoting link is considered to be the *input* or *driver* member and the other is considered to be the *output* or *driven member.* The remaining moving link is commonly called a *connecting link*. The fixed link, hinged by pins or pivots at each end, is called the *foundation link*.

Three inversions or linkage rotations of a four-bar chain are shown in Figs. 9, 10, and 11. They are made up of links AB, BC, CD, and AD. The forms of the three inversions are defined by the position of the shortest link with respect to the link selected as the foundation link. The ability of the driver or driven links to make complete rotations about their pivots determines their functions.

Drag-link mechanism, Fig. 3.29, demonstrates the first inversion. The shortest link AD between the two fixed pivots is the foundation link, and both driver link AB and driven link CD can make full revolutions. *Crank-rocker mechanism*, Fig. 3.30, demonstrates the second inversion. The shortest link AB is adjacent to AD, the foundation link. Link AB can make a full 360° revolution while the opposite link CD can only oscillate and describe an arc.

Double-rocker mechanism, Fig. 3.31, demonstrates the third inversion. Link AD is the foundation link, and it is opposite the shortest link BC. Although link BC can make a full 360_ revolution, both pivoting links AB and CD can only oscillate and describe arcs.

The fourth inversion is another *crank-rocker mechanism* that behaves in a manner similar to the mechanism shown in Fig. 3.30, but the longest link, CD, is the foundation link. Because of this similarity between these two mechanisms, the fourth inversion is not illustrated here. A drag-link mechanism can produce either a non-uniform output from a uniform input rotation rate or a uniform output from a non-uniform input rotation rate.

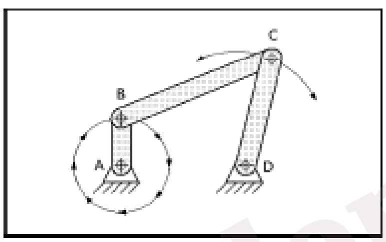


Figure 3.30: Crank-rocker mechanism. Link AB can make a 360° revolution while link CD oscillate with C describing an arc. Link AD is the foundation link.

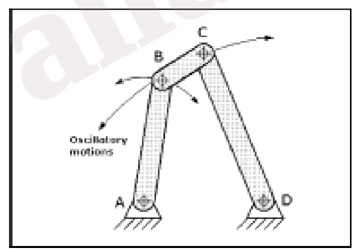


Figure 3.31: Double-rocker mechanism: Short link BC can make a 360° revolution, but pivoting links AB and CD can only oscillate describing arcs.

Straight-Line Generators

Figures 3.32 to 3.35 illustrate examples of classical linkages capable of describing straight lines, a function useful in many different kinds of machines, particularly machine tools. The dimensions of the rigid links are important for the proper functioning of these mechanisms.

Watt's straight-line generator, illustrated in Fig. 3.32, can describe a short vertical straight line. Equal length links AB and CD are hinged at A and D, respectively. The midpoint E of connecting link BC traces a figure eight pattern over the full mechanism excursion, but a straight line is traced in part of the excursion because point E diverges to the left at the top of the stroke and to the right at the bottom of the stroke. This linkage was used by Scottish instrument maker, James Watt, in a steam-driven beam pump in about 1769, and it was a prominent mechanism in early steam-powered machines.

Scott Russell straight-line generator, shown in Fig. 3.33, can also describe a straight line. Link AB is hinged at point A and pinned to link CD at point B. Link CD is hinged to a roller at point C which restricts it to horizontal oscillating movement.

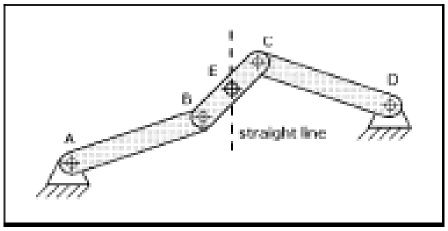


Figure 3.32: Watt's straight-line generator. The center point E of link BC describe a straight line when driven by another links AB or CD

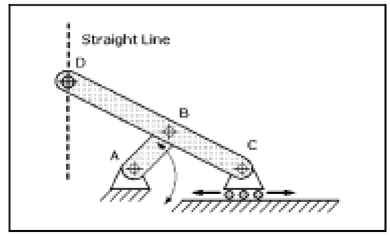


Figure 3.33: Scott Russell straight-line generator: Point D of link DC describe a straight line as driver link AB oscillates causing the slider at C to reciprocate left and right

This configuration confines point D to a motion that traces a vertical straight line. Both points A and C lie in the same horizontal plane. This linkage works if the length of link AB is about 40 percent of the length of CD, and the distance between points D and B is about 60 percent of the length of CD.

Peaucellier's straight-line linkage, drawn as Fig. 3.34, can describe more precise straight lines over its range than either the Watt's or Scott Russell linkages. To make this linkage work correctly, the length of link BC must equal the distance between points A and B set by the spacing of the fixed pivots; in this figure, link BC is 15 units long while the lengths of links CD, DF, FE, and EC are equal at 20 units. As links AD and AE are moved, point F can describe arcs of any radius. However, the linkage can be restricted to tracing straight lines (infinite radiuses) by selecting link lengths for AD and AE. In this figure they are 45 units long. This linkage was invented in 1873 by the French engineer, Captain Charles-Nicolas Peaucellier.

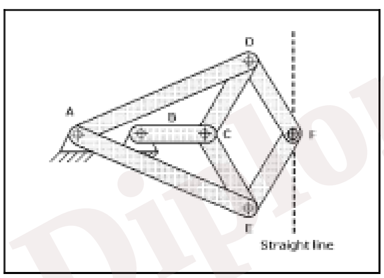


Figure 3.34: Peaucellier's straight-line generator: point F describe a straight line when either link AD or AE acts as the driver

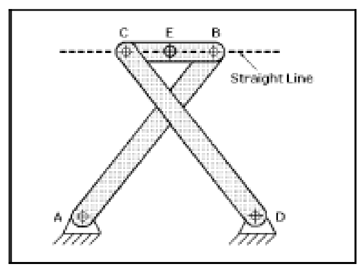


Fig. 3.35 Tchebicheff's straight-line generator: Point E of link CB describes a straight line when driven by either link AB or DC. Link CB moves into a vertical position at both extremes of its travel.

Tchebicheff's straight-line generator, shown in Fig. 3.35, can also describe a horizontal line. Link CB with E as its midpoint traces a straight horizontal line for most of its transit as links AB and DC are moved to the left and right of center. To describe this straight line, the length of the foundation link AD must be twice the length of link CB. To make this mechanism work as a straight-line generator, CB is 10 units long, AD is 20 units long, and both AB and DC are 25 units long. With these dimensions, link CB will assume a vertical position when it is at the right and left extremes of its travel excursion. This linkage was invented by nineteenth-century Russian mathematician, Pafnuty Tchebicheff or Chebyshev.

Rotary/Linear Linkages

Slider-crank mechanism (or a simple crank), shown as Fig. 3.36, converts rotary to linear motion and vice versa, depending on its application. Link AB is free to rotate 360° around the hinge while link BC oscillates back and forth because point C is hinged to a roller which restricts it to linear motion. Either the slider or the rotating link AB can be the driver. This mechanism is more familiar as the piston, connecting rod, and crankshaft of an internal combustion engine, as illustrated in Fig. 3.3. The piston is the slider at C, the connecting rod is link BC, and the crankshaft is link AB. In a four-stroke engine, the piston is pulled down the cylinder by the crankshaft, admitting the air-fuel mixture; in the compression stroke the piston is driven back up the cylinder by the crankshaft to compress the air-fuel mixture. However, the roles change in the combustion stroke when the piston drives the crankshaft. Finally, in the exhaust stroke the roles change again as the crankshaft drives the piston back to expel the exhaust fumes.

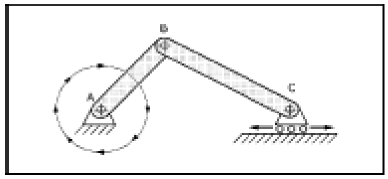


Fig. 3.36 Slider-crank mechanism: This simple crank converts the 360° rotation of driver link AB into linear motion of link BC, causing the slider at C to reciprocate.

Scotch-yoke mechanism, pictured in Fig. 3.37, functions in a manner similar to that of the simple crank mechanism except that its linear output motion is sinusoidal. As wheel A, the driver, rotates, the pin or roller bearing at its periphery exerts torque within the closed yoke B; this causes the attached sliding bar to reciprocate, tracing a sinusoidal waveform. Part a shows the sliding bar when the roller is at 270° and part b shows the sliding bar when the roller is at 0°. *Rotary-to-linear mechanism*, drawn in Fig. 3.38, converts a uniform rotary motion into an intermittent reciprocating motion. The three teeth of the input rotor contact the steps in the frame or yoke, exerting torque 3 times per revolution, moving the yoke with attached bar. Full linear travel of the yoke is accomplished in 30° of rotor rotation followed by a 30° delay before returning the yoke. The reciprocating cycle is completed 3 times per revolution of the input. The output is that of a step function.

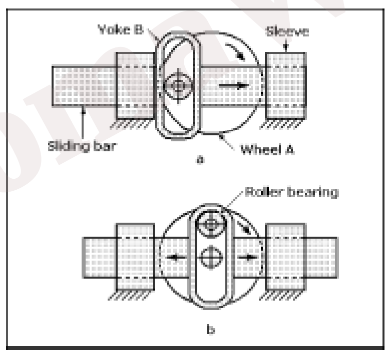


Figure 3.37: Scotch-yoke translates the rotary motion of the wheel with a peripheral roller into reciprocating motion of the yoke.

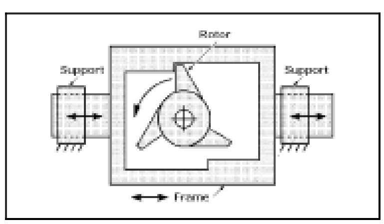


Figure 3.38: Rotary-to-linear mechanism converts the uniform rotation of the 3 tooth rotor into a reciprocating motion of the frame and supporting bars.

3.2.4 Couplers

The term *coupling* refers to a device used to connect two shafts together at their ends for the purpose of transmitting power. There are two general types of couplings: rigid and flexible.

At some point in a mobile robot designer's career there will be a need to couple two shafts together. Fortunately, there are many commercially available couplers to pick from, each with its own strengths and weaknesses. Couplers are available in two major styles: solid and flexible. Solid couplers must be strong enough to hold the shafts' ends together as if they were one shaft. Flexible couplers allow for misalignment and are used where the two shafts are already running in their own bearings, but might be slightly out of alignment. The only other complication is that the shafts may be different diameters, or have different end details like splined, keyed, hex, square, or smooth. The coupler simply has different ends to accept the shafts it is coupling.

Solid couplers are very simple devices. They clamp onto each shaft tight enough to transmit the torque from one shaft onto the other. The shafts styles in each end of the coupler can be the same or different. For shaped shafts, the coupler need only have the same shape and size as the shaft and bolts or other clamping system to hold the coupler to the shaft. For smooth shafts, the coupler must clamp to the shaft tight enough to transmit the torque through friction with the shaft surface. This requires very high clamping forces, but is a common method because it requires no machining of the shafts.

A coupling is used wherever there is a need to connect a prime mover to a piece of driven machinery. The principal purpose of a coupling is to transmit rotary motion and torque from one piece of equipment to another. Couplings may perform other secondary functions, such as accommodating misalignment between shafts, compensating for axial shaft movement, and helping to isolate vibration, heat, and electrical eddy currents from one shaft to another.

1. ***Rigid Couplings*** Rigid couplings are used to connect machines where it is desired to maintain shafts in precise alignment. They are also used where the rotor of one machine is used to support and position the other rotor in a drive train. Because a rigid coupling cannot accommodate misalignment between shafts, precise alignment of machinery is necessary when one is used.

   TYPES. There are two commonly used types of rigid couplings. One type consists of two flanged rigid members, each mounted on one of the connected shafts (Figure 3.39). The flanges are provided with a number of bolt holes for the purpose of connecting the two half-couplings. Through proper design and installation of the coupling, it is possible to transmit the torque load entirely through friction from one flange to the other, which assures that the flange bolts do not experience a shearing stress. This type of arrangement is especially desirable for driving systems where torque oscillations occur, as it avoids subjecting the flange bolts to a shearing stress.

   A second type of rigid coupling, known as the *split rigid,* is split along its horizontal centerline (Figure 3.40). The two halves are clamped together by a series of bolts arranged axially along the coupling. The rigid coupling and machine shafts may be equipped with conventional keyways, which are in turn fitted with keys to transmit the torque load, or in certain cases the frictional clamping force may be sufficient to permit transmitting the torque by means of friction between shaft and rigid coupling. This type of coupling is commonly used to connect sections of line shafting in a drive train.

   A variation to the flanged rigid coupling is known as the adjustable rigid coupling (Figure 3.41). This coupling is designed along the lines of conventional rigid couplings, except that a threaded adjusting ring is placed between the two flanges. This ring engages a threaded extension on one of the shaft ends. By means of this ring, it is possible to position the pump shaft axially with respect to the driver.

APPLICATIONS. A common application for rigid couplings in the pump industry is in vertical drives, where the prime mover (generally an electric motor) is positioned above the pump. In such cases, both machines can employ a common thrust bearing, which is generally located in the motor. The coupling flange bolts must be capable of transmitting any down thrust from the pump to the motor. In applications where the thrust from the pump is toward the motor, it is common practice to provide shoulders on the shafts to transmit the axial force.

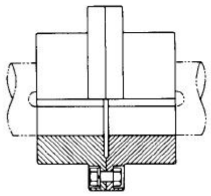


Figure 3.39: Flanged rigid coupling

Figure 3.40: Split rigid coupling

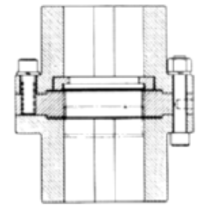


Figure 3.41: Adjustable flanged rigid coupling

Figure 3.42: Gear-type mechanically flexible coupling

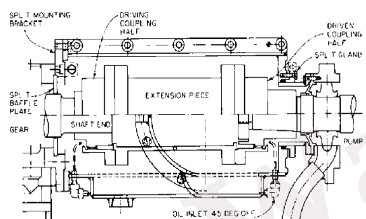


Figure 3.43: Continuously lubricated coupling

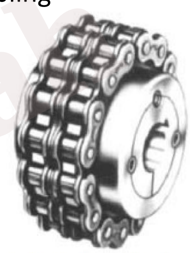


Figure 3.44: Roller chain mechanically flexible coupling

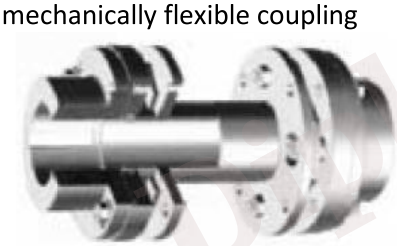


Figure 3.45: Spacer metal disk coupling

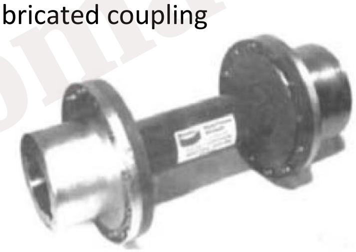


Figure 3.46: Diaphragm material-flexible coupling

Figure 3.47: Pin-and-bushing elastomer coupling

Figure 3.48: Sleeve-type elastomer coupling

Figure 3.49: Spacer elastomer coupling

Figure 12: Compression-loaded

Many pump drive systems require a rigid coupling that is capable of providing axial adjustment to compensate for wear in the pump impeller or impellers. The adjustable rigid coupling is used for this purpose. The threaded adjusting ring attached to a mating threaded extension of the pump shaft permits vertical positioning of the impeller or impellers. The hub, which is mounted on the pump shaft, is equipped with a clearance lit and feathered key that permit the hub to slide with respect to the shaft. The load capacity of a coupling of this type is generally limited by the pressure on the pump shaft key because there is no possibility of load being transmitted by interference fit.

A few words of caution should be noted about the use of rigid couplings. First, precise alignment of machine bearings is absolute necessary because there is no flexibility in the coupling to accommodate misalignment between shafts. Secondly, accuracy of manufacture is extremely important. The coupling surfaces that interface between driving and driven shafts must be manufactured with high degrees of concentricity and squareness, to avoid the transmittal of eccentric motion from one machine to the other.

2. ***Flexible Couplings*** Flexible couplings accomplish the primary purpose of any coupling; that is, to transmit a driving torque between prime mover and driven machine. In addition, they perform a second important function: they accommodate unavoidable misalignment between shafts. A proliferation of designs exists for flexible couplings, which may be classified into two types: mechanically flexible and materially flexible.

   a. ***MECHANICALLY FLEXIBLE COUPLINGS***. Mechanically flexible couplings compensate for misalignment between two connected shafts by means of clearances incorporated in the design of the coupling. The most commonly used type of mechanically flexible coupling is the gear, or dental, coupling (Figure 3.42). This coupling essentially consists of two pair of clearance fit splines. In the most common configuration, the two machine shafts are equipped with hub members having external splines cut integrally on the hubs. The two hubs are connected by a sleeve member having mating internal gear teeth. Backlash is intentionally built into the spline connection, and it is this backlash that compensates for shaft misalignment. Sliding motion occurs in a coupling of this type, and so a supply of clean lubricant (grease or oil, depending on the design) is necessary to prevent wear of the rubbing surfaces.

      If operation cannot be interrupted to lubricate the couplings, constantly lubricated couplings are used, as shown in Figure 3.43. These consist of an oil tight enclosure bolted at one end to the stationary portion of either the driving or driven piece of equipment. The other end of the enclosure has a slip fit inside a cover that is bolted to the other piece of equipment.

      Some form of packing is used to prevent loss of lubricant at the slip joint. Oil under pressure is brought through the enclosure and impinges upon the meshing gear teeth of the coupling, the excess being collected at the bottom of the enclosure and returned to the oil reservoir.

      A second type of mechanically flexible coupling that sees wide usage, especially in low cost drive systems, is known as the roller-chain flexible coupling (Figure 3.44). This coupling employs two sprocket-like members mounted one on each of the two machine shafts and connected by an annulus of roller chain. The clearance between sprocket and roller and, in some cases, the crowning of the rollers provides mechanical flexibility for misalignment.

      This type of coupling is generally limited to low-speed machinery.

   b. ***MATERIAL-FLEXIBLE COUPLINGS*** These couplings rely on flexing of the coupling element to compensate for shaft misalignment. The flexing element may be of any suitable material (metal, elastomer, or plastic) that has sufficient resistance to fatigue failure to provide acceptable life. Some materials, such as steel, have a finite fatigue limit. A coupling made of such material must be operated under conditions of load and misalignment that assure that the stress developed in the coupling element is within that limit. Other materials, such as elastomers, generally do not have a well-defined fatigue limit. In these cases, however, heat developed in the material when the coupling flexes can cause failure if excessive.

One type of material-flexible coupling is the metal-disk coupling (Figure 3.45). This coupling consists of two sets of thin sheet-metal disks bolted to the driving and driven hub members. Each set of disks is made up of a number of thin laminations that are individually flexible and compensate for shaft misalignment by means of this flexibility. These disks may be stacked together as required to obtain the desired torque transmission capability. This type of coupling requires no lubrication; however, alignment of the equipment must be maintained within acceptable limits so as not to exceed the fatigue limit of the material.

Another example of an all-metal material-flexible coupling is the flexible diaphragm coupling, shown in Figure 3.46. This coupling is similar in function to the metal-disk coupling in that the disk flexes to accommodate misalignment. However, the diaphragm type consists of a single element with a hyperbolic contour that is designed to produce uniform stress in the member from inner to outer diameter. By more efficiently utilizing the material, the weight is reduced correspondingly, thus making this coupling suitable for high speed applications.

Material-flexible couplings employing elastomer materials are numerous and their designs are varied. By definition, an elastomer is a material that has a high degree of elasticity and resiliency and will return to its original shape after undergoing large-amplitude deformations. One example of an elastomer coupling is the pin-and-bushing coupling (Figure 3.47). This design comprises two flanged hub members, one mounted on each machine shaft. The flange of one hub is fitted with pins that extend axially toward the adjacent shaft. The other flange is equipped with rubber bushings, which generally have a metal sleeve at the center. The pins fit into these sleeves and provide transmission of torque through the bushings. Because the bushings are made of flexible material, they can accept slight angularity or offset conditions between the two flanges.

A second group of elastomer couplings employs a sleeve-like element that is connected to a hub member on each shaft and transmits torque through shearing of the flexible element. The flexible element may be attached to the machine hubs by a number of different means: it may be chemically bonded, it may be mechanically connected by means of loose fitting splines (Figure 3.48), or it may be clamped to the hubs and held in place by friction (Figure 3.49). Misalignment between shafts is accommodated through flexing of the elastomer sleeve.

A third group of couplings utilizes an elastomer member that is loaded in compression to transmit load from one shaft to the other. The elastomer material is placed loosely into cavities formed by members rigidly mounted onto the two shafts (Figure 3.50). Again, the elastomer material deflects to compensate for shaft misalignment, and is usually used when torsional damping is required.

Another type of elastomer coupling that is commonly used on low-power drive systems is the rubber jaw coupling (Figure 3.51). The heart of this coupling is a “spider” member, having a plurality (usually three) of segments extending radially from a central section. The hubs, which are mounted on driving and driven shafts, each have a set of jaws corresponding to the number of spiders on the flexible element. The spider fits between the two sets of jaws and provides a flexible “cushion” between them. This cushion transmits the torque load as well as compensating for misalignment.

c. ***SPRING-GRID COUPLING*** There is a type of commercially available flexible coupling that combines the characteristics of mechanically flexible and material-flexible couplings. This is the spring-grid coupling (Figure 3.52). This design has two hubs, one mounted on each machine shaft. Each hub has a raised portion on which tooth like slots are cut. A spring steel grid member is fitted, or woven, between the slots on the two hubs. The grid element can slide in the slots to accommodate shaft misalignment and flexes like a leaf spring to transmit torque from one machine to the other. Unlike most material-flexible couplings, this design requires periodic lubrication to prevent excessive wear of the grid member.

APPLICATIONS. The type of flexible coupling most suitable for a particular application depends upon a number of factors, including power, speed of rotation, shaft separation, amount of misalignment, cost, and reliability. In the design of a system, it is the goal of the designer to use the least expensive coupling that will do the job. In low-cost systems, cost alone may be the most important criterion, and the least expensive coupling that transmits the rated power and accepts some small degree of misalignment is generally the choice, albeit at some sacrifice of reliability and durability. On the other hand, high power, high-speed machinery generally represents a critical piece of equipment for a power station, sewage plant, or other vital process, and in these cases a coupling should be selected that will not compromise the overall reliability of the system.

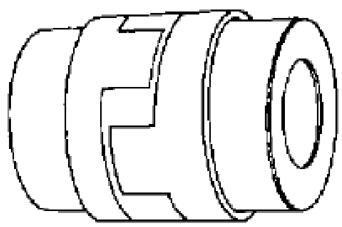


Figure 3.51:Rubber jaw coupling

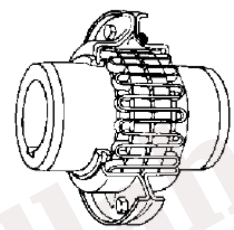


Figure 3.52: Spring-grid coupling

Low-power pumps (up to about 200 hp, 150 kW) driven by electric motors can usually be coupled successfully by any of the couplings described here. Selection procedures vary from manufacturer to manufacturer, but generally the following data are required: power rating, speed, anticipated misalignment, and type of pump (reciprocating, vane, centrifugal, and so on).

Pumps of the same power range are very often driven by reciprocating engines (diesel, gasoline, natural gas). This is quite common in remote areas, such as at pipeline pumping stations, where a source of electric power is not available. Because this type of prime mover produces a pulsating type of power, it is often necessary to perform a torsional vibration analysis of the drive system to ensure that the normal operating speed is well removed from a speed that may produce a torsional resonant vibration. Such an analysis requires that the torsional stiffness of the coupling be known. It is quite often possible to tune the drive system to avoid operating at a resonant condition by selecting the proper coupling stiffness. The selection data required for a system of this type are the same as listed above. Most coupling manufacturers will, however, assign a higher service factor to an application involving a reciprocating prime mover, to compensate for fatigue effects due to torque fluctuations. In addition, the remote location of many engine-driven pumps indicates a special need to ensure a high degree of reliability of the system.

FOUR COUPLINGS FOR PARALLEL SHAFTS

Fig. 1 One method of coupling shafts makes use of gears that can replace chains, pulleys, and friction drives. Its major limitation is the need for adequate center distance. However, an idler can be used for close centers, as shown. This can be a plain pinion or an internal gear. Transmission is at a constant velocity and there is axial freedom.

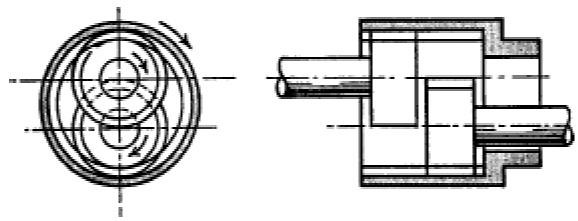


Fig. 2 This coupling consists of two universal joints and a short shaft. Velocity transmission is constant between the input and output shafts if the shafts remain parallel and if the end yokes are arranged symmetrically. The velocity of the central shaft fluctuates during rotation, but high speed and wide angles can cause vibration. The shaft offset can be varied, but axial freedom requires that one shaft be spline mounted.

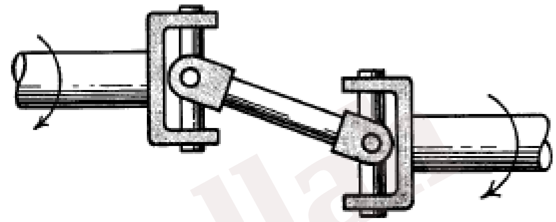


Fig. 3 This crossed-axis yoke coupling is a variation of the mechanism shown in Fig. 2. Each shaft has a yoke connected so that it can slide along the arms of a rigid cross member. Transmission is at a constant velocity, but the shafts must remain parallel, although the offset can vary. There is no axial freedom. The central cross member describes a circle and is thus subjected to centrifugal loads.

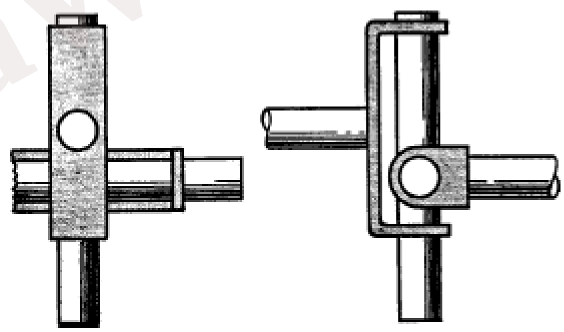
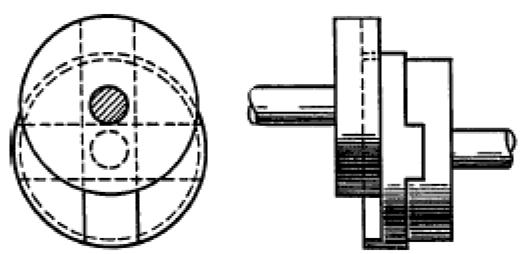


Fig. 4 This Oldham coupling provides motion at a constant velocity as its central member describes a circle. The shaft offset can vary, but the shafts must remain parallel. A small amount of axial freedom is possible. A tilt in the central member can occur because of the offset of the slots. This can be eliminated by enlarging its diameter and milling the slots in the same transverse plane.



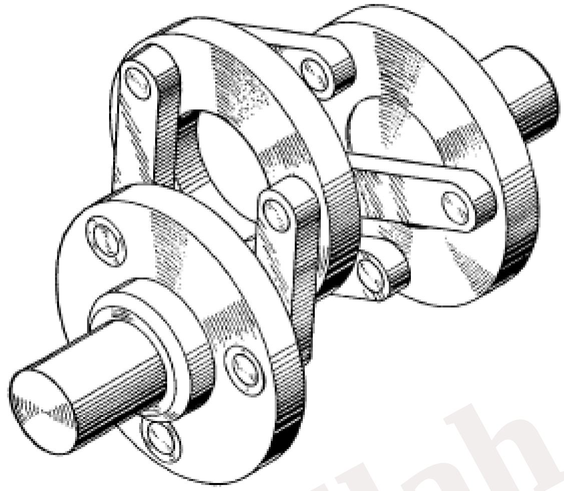
LINKS AND DISKS COUPLE OFFSET SHAFTS

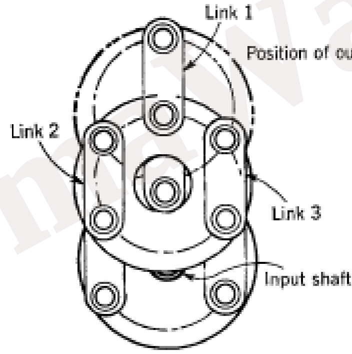
An unorthodox yet remarkably simple arrangement of links and disks forms the basis of a versatile parallel-shaft coupling. This coupling—essentially three disks rotating in unison and interconnected in series by six links (se drawing)—can adapt to wide variations in axial displacement while it is running under load.

Changes in radial displacement do not affect the constant-velocity relationship between the input and output shafts, nor do they affect initial radial reaction forces that might cause imbalance in the system. Those features open up unusual applications for it in automotive, marine, machine-tool, and rolling-mill machinery (see drawings).

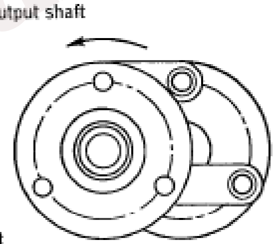
How it works. The inventor of the coupling, Richard Schmidt of Madison, Alabama, said that a similar link arrangement had been known to some German engineers for years. But those engineers were discouraged from applying the theory because they erroneously assumed that the center disk had to be retained by its own bearing. Actually, Schmidt found that the center disk is free to assume its own center of rotation. In operation, all three disks rotate with equal velocity.

The bearing-mounted connections of links to disks are equally spaced at 120° on pitch circles of the same diameter. The distance between shafts can be varied steplessly between zero (when the shafts are in line) and a maximum that is twice the length of the links (see drawings.) There is no phase shift between shafts while the coupling is undulating.



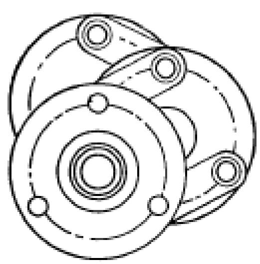


Maximum displacement
(rear view)

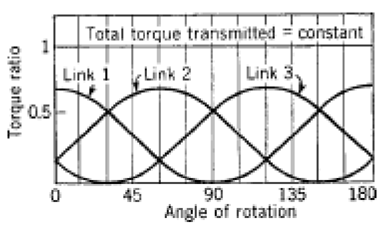


Input and output shafts in line
Zero displacement

Parallel-link connections between disks (see upper drawing) exactly duplicate the motion between the input and output shafts—the basis of this principle in coupling. The lower diagrams show three positions of the links as one shaft is shifted with respect to the other shaft in the system.



Midway position



Torque transmitted by three links in the group adds up to a constant value, regardless of the angle of rotation.

291

Hooke's Joints

The commonest form of a universal coupling is a *Hooke's joint*. It can transmit torque efficiently up to a maximum shaft alignment angle of about 36°. At slow speeds, on hand-operated mechanisms, the permissible angle can reach 45°. The simplest arrangement for a Hooke's joint is two forked shaft-ends coupled by a cross-shaped piece. There are many variations and a few of them are included here.

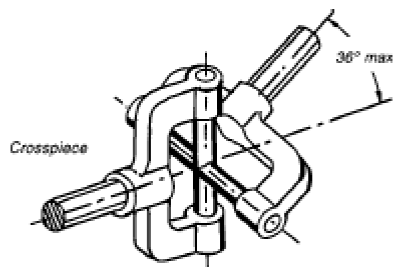


Fig. 1 The Hooke's joint can transmit heavy loads. Anti-friction bearings are a refinement often used.

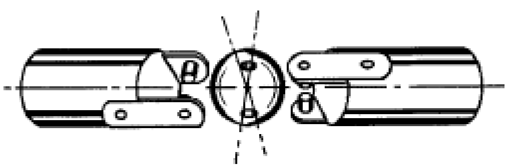


Fig. 2 A pinned sphere shaft coupling replaces a cross-piece. The result is a more compact joint.

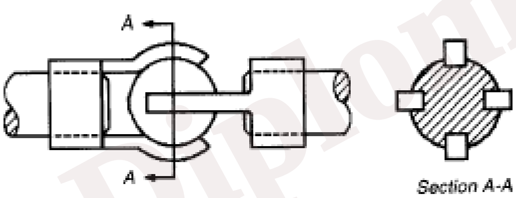


Fig. 3 A grooved-sphere joint is a modification of a pinned sphere. Torques on fastening sleeves are bent over the sphere on the assembly. Greater sliding contact of the torques in grooves makes simple lubrication essential at high torques and alignment angles.

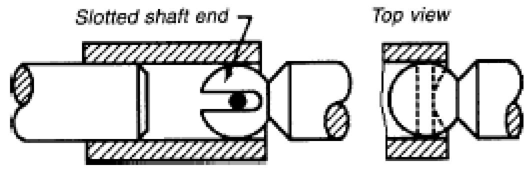


Fig. 4 A pinned-sleeve shaft-coupling is fastened to one shaft that engages the forked, spherical end on the other shaft to provide a joint which also allows for axial shaft movement. In this example, however, the angle between shafts must be small. Also, the joint is only suitable for low torques.

Constant-Velocity Couplings

The disadvantages of a single Hooke's joint is that the velocity of the driven shaft varies. Its maximum velocity can be found by multiplying driving-shaft speed by the secant of the shaft angle; for minimum speed, multiply by the cosine. An example of speed variation: a driving shaft rotates at 100 rpm; the angle between the shafts is 20°. The minimum output is 100×0.9397, which equals 93.9 rpm; the maximum output is 1.0642×100, or 106.4 rpm. Thus, the difference is 12.43 rpm. When output speed is high, output torque is low, and vice versa. This is an objectionable feature in some mechanisms. However, two universal joints connected by an intermediate shaft solve this speed-torque objection.

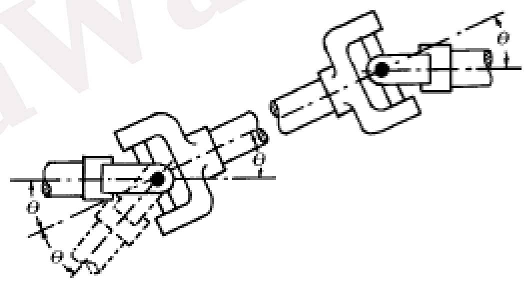


Fig. 5 A constant-velocity joint is made by coupling two Hooke's joints. They must have equal input and output angles to work correctly. Also, the forks must be assembled so that they will always be in the same plane. The shaft-alignment angle can be double that for a single joint.

This single constant-velocity coupling is based on the principle (Fig. 6) that the contact point of the two members must always lie on the homokinetic plane. Their rotation speed will then always be equal because the radius to the contact point of each member will always be equal. Such simple couplings are ideal for toys, instruments, and other light-duty mechanisms. For heavy duty, such as the front-wheel drives of military vehicles, a more complex coupling is shown diagrammatically in Fig. 7A. It has two joints close-coupled with a sliding member between them. The exploded view (Fig. 7B) shows these members. There are other designs for heavy-duty universal couplings; one, known as the *Rzeppa*, consists of a cage that keeps six balls in the homokinetic plane at all times. Another constant-velocity joint, the *Bendix-Weiss*, also incorporates balls.

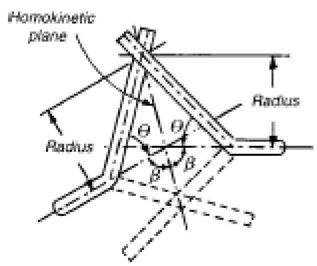


Fig. 6

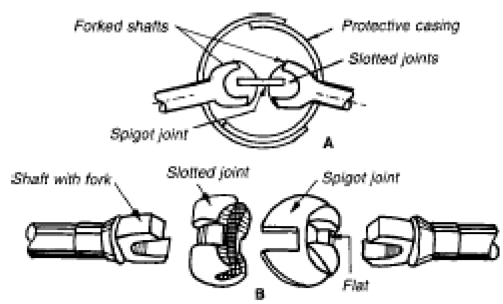


Fig. 7

Fig. 8 This flexible shaft permits any shaft angle. These shafts, if long, should be supported to prevent backlash and coiling.

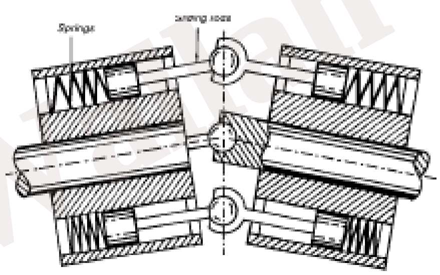


Fig. 9 This pump-type coupling has the reciprocating action of sliding rods that can drive pistons in cylinders.

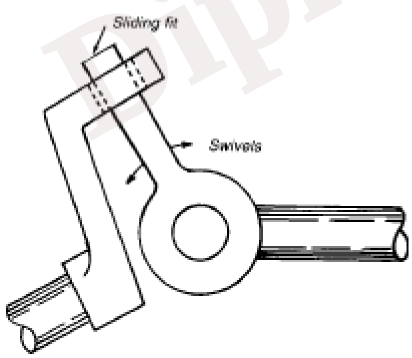


Fig. 10 This light-duty coupling is ideal for many simple, low-cost mechanisms. The sliding swivel-rod must be kept well lubricated at all times.

3.2.5 The Concept of Power Transfer

Power transfer mechanisms are normally divided into five general categories:

1. belts (flat, round, V-belts, timing)
2. chain (roller, ladder, timing)
3. plastic-and-cable chain (bead, ladder, pinned)
4. friction drive
5. gears (spur, helical, bevel, worm, rack and pinion, and many others)

Some of these, like V-belts and friction drives, can be used to provide the further benefit of mechanically varying the output speed. This ability is not usually required on a mobile robot, indeed it can cause control problems in certain cases because the computer does not have direct control over the actual speed of the output shaft. Other power transfer devices like timing belts, plastic-and-cable chain, and all types of steel chain connect the input to the output mechanically by means of teeth just like gears. These devices could all be called synchronous because they keep the input and output shafts in synch, but roller chain is usually left out of this category because the rollers allow some relative motion between the chain and the sprocket. The term *synchronous* is usually applied only to toothed belts which fit on their sprockets much tighter than roller chain.

For power transfer methods that require attaching one shaft to another, like motor-mounted gearboxes driving a separate output shaft, a method to deal with misalignment and vibration should be incorporated. This is done with shaft couplers and flexible drives. In some cases where shock loads might be high, a method of protecting against overloading and breaking the power transfer mechanism should be included. This is done with torque limiters and clutches. Let's take a look at each method. We'll start with mechanisms that transfer power between shafts that are not inline, and then look at couplers and torque limiters. Each section has a short discussion on how well that method applies to mobile robots.

Transmission of power from a source, such as an engine or motor, through a machine to an output actuation is one of the most common machine tasks. An efficient means of transmitting power is through rotary motion of a shaft that is supported by bearings. Gears, belt pulleys, or chain sprockets may be incorporated to provide for torque and speed changes between shafts. Most shafts are cylindrical (solid or hollow), and include stepped diameters with shoulders to accommodate the positioning and support of bearings, gears, etc. The design of a system to transmit power requires attention to the design and selection of individual components (gears, bearings, shaft, etc.). However, as is often the case in design, these components are not independent. For example, in order to design the shaft for stress and deflection, it is necessary to know the applied forces. If the forces are transmitted through gears, it is necessary to know the gear specifications in order to determine the forces that will be transmitted to the shaft. But stock gears come with certain bore sizes, requiring knowledge of the necessary shaft diameter. It is no surprise that the design process is interdependent and iterative, but where should a designer start?

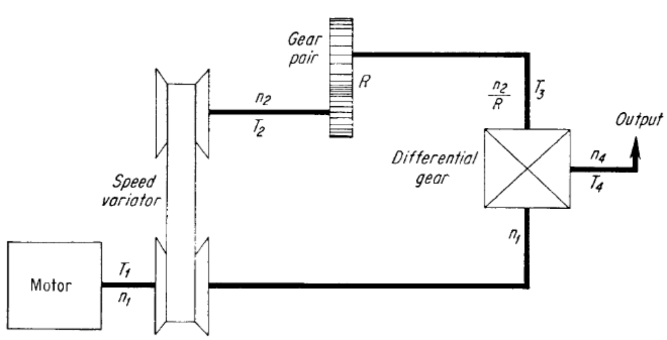
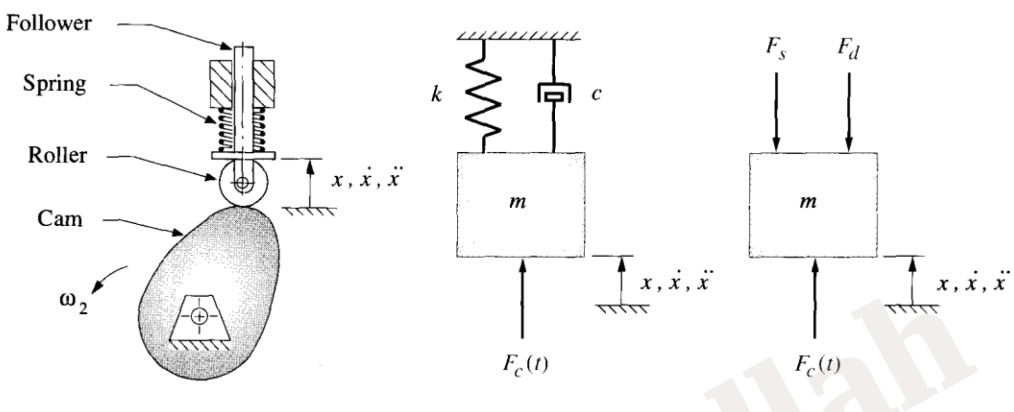


Figure 3.53: The concept of power transfer

3.3 Modeling of Mechanical System

If the dynamic behavior of a physical system can be represented by an equation, or a set of equation, this is referred to as the mathematical model of the system. Such models can be constructed from knowledge of the physical characteristics of the system, i.e. mass for a mechanical system or resistance for an electrical system. Alternatively, a mathematical model may be determined by experimentation, by measuring how the system output responds to known inputs.



(a) Physical system (b) Lumped model (c) Free-body diagram

Figure 3.54: One-DOF lumped parameter model of a cam-follower system

Table 3-4: Lists the through and across variables **for various** types of dynamic systems.

TABLE 10-1 Through and Across Variables in Dynamic Systems

System Type	Through Variable	Across Variable	Power Units
Electrical	Current (i)	Voltage (e)	ei = watts
Mechanical	Force (F)	Velocity (v)	Fv = (in-lb)/sec
Fluid	Flow (Q)	Pressure (P)	PQ = (in-lb)/sec

TABLE 10-2 Physical Analogs in Dynamic Systems

System Type	Energy Dissipator	Energy Storage	Energy Storage
Electrical	Resistor (R)	Capacitor (C)	Inductor (L)
Mechanical	Damper (c)	Mass (m)	Spring (k)
Fluid	Fluid resistor (R_f)	Accumulator (C_f)	Fluid inductor (L_f)

TABLE 10-3 Relationships Between Variables in Dynamic Systems

System Type	Resistance	Capacitance	Inductance
Electrical	$i = \frac{1}{R} e$	$i = C \frac{de}{dt}$	$i = \frac{1}{L} \int e\, dt$
Mechanical	$F = c\, v$	$F = m \frac{dv}{dt}$	$F = k \int v\, dt$
Fluid	$Q = \frac{1}{R_f} P$	$Q = C_f \frac{dP}{dt}$	$Q = \frac{1}{L_f} \int P\, dt$

1. Stiffness in mechanical systems

An elastic element is assumed to produce an extension proportional to the force (or torque) applied to it.

For the translational spring

Force ∝ Extension

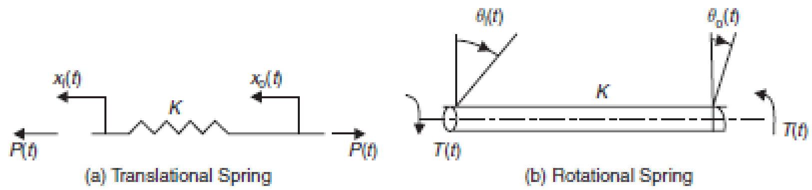


Fig. 3.55: Linear elastic elements.

If $x_i(t) > x_0(t)$, then

$$P(t) = K(x_i(t) - x_0(t) \quad \textbf{(3.41)}$$

And for the rotational spring

Torque ∝ Velocity

If $\theta_i(t) > \theta_0(t)$, then

$$T(t) = K(\theta_i(t) - \theta_o(t)) \quad \textbf{(3.42)}$$

Note that ***K,*** the spring stiffness, has units of (N/m) in equation (3.41) and (Nm/rad) in equation (3.42)

2. Damping in mechanical systems

A damping element (sometimes called a dashpot) is assumed to produce a velocity proportional; to the force (or torque) applied to it.

Force ∝ Velocity

$$P(t) = C_{v(t)} = C\,\frac{dx_o}{dt} \quad \textbf{(3.43)}$$

And for the rotational damper

Torque ∝ Angular velocity

$$T(t) = C_{\omega(t)} = C\frac{d\theta_o}{dt} \quad \textbf{(3.44)}$$

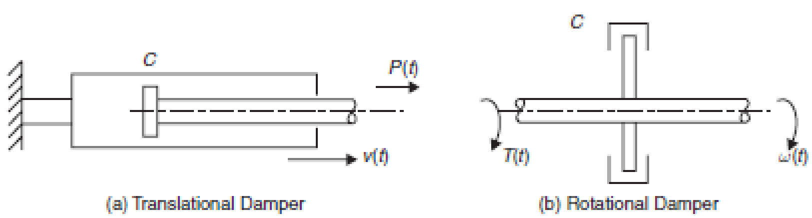


Figure 3.56: Linear damping elements

Note that *C*, the damping coefficient, has units of (*Ns/m*) in equation (3.43) and (*Nm s/rad*) in equation (3.44).

3. Mass in mechanical systems

The force to accelerate a body is the product of its mass and acceleration (Newton's second law).
For the translational system

Force ∝ Acceleration

$$P(t) = ma(t) = m\frac{dv}{dt} = m\frac{d^2x_o}{dt^2} \qquad \textbf{(3.45)}$$

For the rotational system

Torque ∝ Angular acceleration

$$T(t) = I\alpha(t) = I\frac{d\omega_\omega}{dt} = I\frac{d^2\theta_o}{dt^2} \qquad \textbf{(3.46)}$$

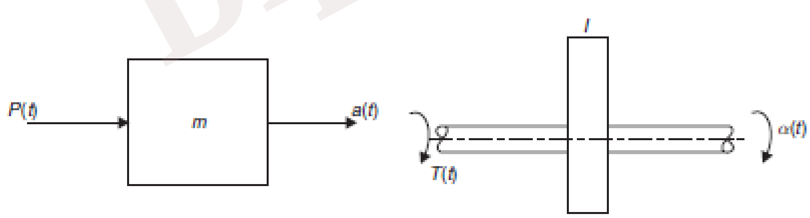
If equation (3.46) I is the moment of inertia about the rotational axis.
when analyzing mechanical systems, it is usual to identify all external forces by the use of a 'free-biody diagram and the apply Newton's second law of motion in the form:

$$\sum F = ma \qquad \text{for translational systems}$$

or

$$\sum M = I\alpha \qquad \text{for rotational systems} \qquad \textbf{(3.47)}$$

Example 2.1
Find the differential equation relating the displacements $x_i(t)$ and $x_o(t)$ for the spring–mass–damper system shown in Figure 2.5. What would be the effect of neglecting the mass?



(a) Translational Acceleration (b) Angular Acceleration

Fig. 2.4 Linear mass elements.

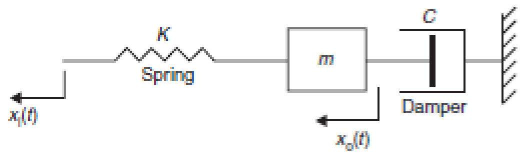


Fig. 2.5 Spring–mass–damper system.

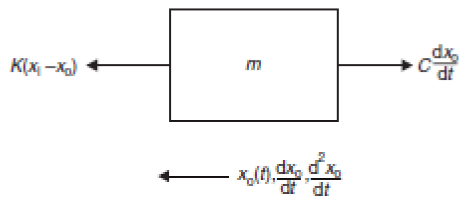


Fig. 2.6 Free-body diagram for spring–mass–damper system.

Solution

Using equations (2.12) and (2.14) the free-body diagram is shown in Figure 2.6. From equation (2.18), the equation of motion is

$$\sum F_x = ma_x$$

$$K(x_i - x_o) - C\frac{dx_o}{dt} = m\frac{d^2x_o}{dt^2}$$

$$Kx_i - Kx_o = m\frac{d^2x_o}{dt^2} + C\frac{dx_o}{dt}$$

Putting in the form of equation (2.10)

$$m\frac{d^2x_o}{dt^2} + C\frac{dx_o}{dt} + Kx_o = Kx_i(t) \qquad (2.19)$$

Hence a spring–mass–damper system is a second-order system.

If the mass is zero then

$$\sum F_x = 0$$

$$K(x_i - x_o) - C\frac{dx_o}{dt} = 0$$

$$Kx_i - Kx_o = C\frac{dx_o}{dt}$$

Hence

$$C\frac{dx_o}{dt} + Kx_o = Kx_i(t) \qquad (2.20)$$

Thus if the mass is neglected, the system becomes a first-order system.

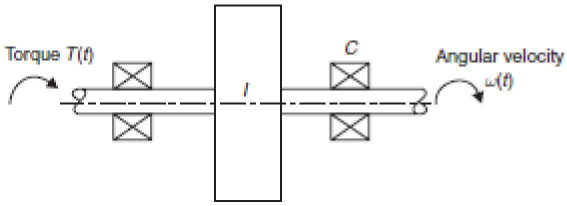


Fig. 2.7 Flywheel in bearings.

Example 2.2
A flywheel of moment of inertia I sits in bearings that produce a frictional moment of C times the angular velocity $\omega(t)$ of the shaft as shown in Figure 2.7. Find the differential equation relating the applied torque $T(t)$ and the angular velocity $\omega(t)$.

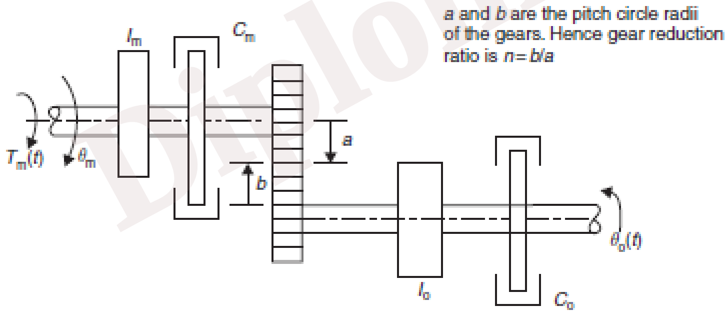
Solution
From equation (2.18), the equation of motion is

$$\sum M = I\alpha$$

$$T(t) - C\omega = I\frac{d\omega}{dt}$$

$$I\frac{d\omega}{dt} + C\omega = T(t) \qquad (2.21)$$

Example 2.3
Figure 2.8 shows a reduction gearbox being driven by a motor that develops a torque $T_m(t)$. It has a gear reduction ratio of 'n' and the moments of inertia on the motor and output shafts are I_m and I_o, and the respective damping coefficients C_m and C_o. Find the differential equation relating the motor torque $T_m(t)$ and the output angular position $\theta_o(t)$.



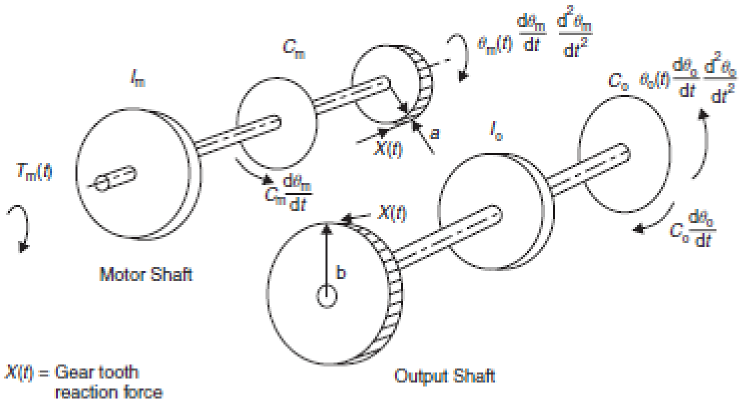


Fig. 2.9 Free-body diagrams for reduction gearbox.

Gearbox parameters

$$I_m = 5 \times 10^{-6}\ \text{kg m}^2$$
$$I_o = 0.01\ \text{kgm}^2$$
$$C_m = 60 \times 10^{-6}\ \text{Nm s/rad}$$
$$C_o = 0.15\ \text{Nm s/rad}$$
$$n = 50{:}1$$

Solution

The free-body diagrams for the motor shaft and output shaft are shown in Figure 2.9.
Equations of Motion are

(1) Motor shaft

$$\sum M = I_m \frac{d^2\theta_m}{dt^2}$$

$$T_m(t) - C_m \frac{d\theta_m}{dt} - aX(t) = I_m \frac{d^2\theta_m}{dt^2}$$

re-arranging the above equation,

$$X(t) = \frac{1}{a}\left(T_m(t) - I_m \frac{d^2\theta_m}{dt^2} - C_m \frac{d\theta_m}{dt}\right) \quad (2.22)$$

(2) Output shaft

$$\sum M = I_o \frac{d^2\theta_m}{dt^2}$$

$$bX(t) - C_o \frac{d\theta_o}{dt} = I_o \frac{d^2\theta_o}{dt^2}$$

re-arranging the above equation,

$$X(t) = \frac{1}{b}\left(I_o \frac{d^2\theta_o}{dt^2} + C_o \frac{d\theta_o}{dt}\right) \tag{2.23}$$

Equating equations (2.22) and (2.23)

$$\frac{b}{a}\left(T_m(t) - I_m \frac{d^2\theta_m}{dt^2} - C_m \frac{d\theta_m}{dt}\right) = \left(I_o \frac{d^2\theta_o}{dt^2} + C_o \frac{d\theta_o}{dt}\right)$$

Kinematic relationships

$$\frac{b}{a} = n \quad \theta_m(t) = n\theta_o(t)$$

$$\frac{d\theta_m}{dt} = n\frac{d\theta_o}{dt}$$

$$\frac{d^2\theta_m}{dt^2} = n\frac{d^2\theta_o}{dt^2}$$

Hence

$$n\left(T_m(t) - nI_m \frac{d^2\theta_o}{dt^2} - nC_m \frac{d\theta_o}{dt}\right) = \left(I_o \frac{d^2\theta_o}{dt^2} + C_o \frac{d\theta_o}{dt}\right)$$

giving the differential equation

$$\left(I_o + n^2 I_m\right)\frac{d^2\theta_o}{dt^2} + \left(C_o + n^2 C_m\right)\frac{d\theta_o}{dt} = nT_m(t) \tag{2.24}$$

The terms $(I_o + n^2 I_m)$ and $(C_o + n^2 C_m)$ are called the equivalent moment of inertia I_e and equivalent damping coefficient C_e referred to the output shaft.

Substituting values gives

$$I_e = (0.01 + 50^2 \times 5 \times 10^{-6}) = 0.0225\,\text{kg m}^2$$
$$C_e = (0.15 + 50^2 \times 60 \times 10^{-6}) = 0.3\,\text{Nm s/rad}$$

From equation (2.24)

$$0.0225\frac{d^2\theta_o}{dt^2} + 0.3\frac{d\theta_o}{dt} = 50T_m(t) \tag{2.25}$$

3.3.1 Elements, Rules and Nomenclature

Three Basic Mechanical Elements are:

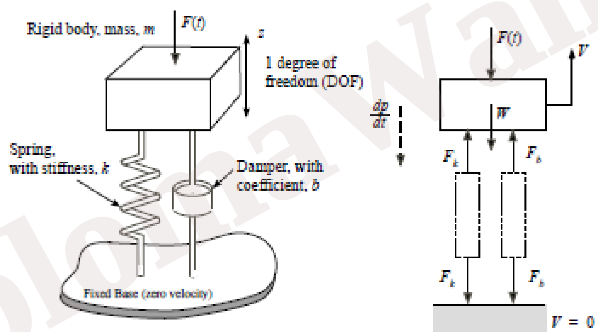
- Spring (elastic) element
- Damper (frictional) element
- Mass (inertia) element

When modeling translational and rotational systems it is common to break the system into parts. These parts are then described with Free Body Diagrams (FBDs). Spring, damper and mass are passive (non-energy producing) devices. Driving Inputs will force and motion sources which cause elements to respond. Each of the elements has one of two possible energy behaviors:

– stores all the energy supplied to it
– dissipates all energy into heat by some kind of “frictional” effect
  - Spring stores energy as potential energy
  - Mass stores energy as kinetic energy
  - Damper dissipates energy into heat

Dynamic Response of each element is important for

– step response
– frequency response



(a) (b)

Figure 1: Mass-Spring-Damper: A basic mechanical system that consists of a rigid body that can be translate in the z-direction is shown in Figure 3.57. The system is modeled using a mass, a spring and a damper and a force, *F(t)* is applied.

Table 3.5 Mechanical Dissipative Elements

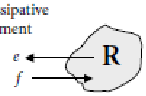
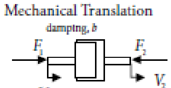
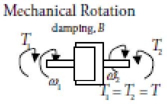
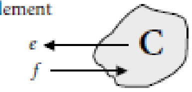
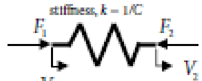
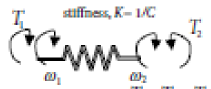
Physical System	Fundamental Relations
Generalized Dissipative Element  • Resistive element • Resistance, R	Dissipation: $\mathbf{e}\cdot\mathbf{f} = \sum_i e_i f_i = T\cdot$ Resistive law: $e = \Phi_R(f)$ Conductive law: $f = \Phi_R^{-1}(e)$ Content: $P_f = \int e \cdot df$ Co-content: $P_e = \int f \cdot de$
Mechanical Translation  $F_1 = F_2 = F$ $V_1 - V_2 = V$ • Damper • damping, b	Constitutive: $F = \Phi(V)$ Content: $P_V = \int F \cdot dV$ Co-energy: $P_F = \int V \cdot dF$ Dissipation: $P_d = P_V + P_F$
Mechanical Rotation  $T_1 = T_2 = T$ $\omega_1 - \omega_2 = \omega$ • Torsional damper • damping, B	Constitutive: $T = \Phi(\omega)$ Content: $P_\omega = \int T \cdot d\omega$ Co-energy: $P_T = \int \omega \cdot dT$ Dissipation: $P_d = P_\omega + P_T$

Table 3.6: Mechanical Potential Energy Storage Elements (Integral Form)

Physical System	Fundamental Relations
Generalized Potential Energy Storage Element  • Capacitive element • Capacitance, C	State: $\mathbf{q}=$ displacement Rate: $\dot{\mathbf{q}}=\mathbf{f}$ Constitutive: $\mathbf{e}=\Phi(\mathbf{q})$ Energy: $U_q=\int \mathbf{e}\cdot d\mathbf{q}$ Co-energy: $U_e=\int \mathbf{q}\cdot d\mathbf{e}$
Mechanical Translation  $F_1=F_2=F$ $V_1-V_2=V$ • spring • stiffness, k, compliance, C	State: $x=$ displacement Rate: $\dot{x}=V$ Constitutive: $F=F(x)$ Energy: $U_x=\int F\cdot dx$ Co-energy: $U_F=\int x\cdot dF$
Mechanical Rotation  $T_1=T_2=T$ $\omega_1-\omega_2=\omega$ • Torsional spring • stiffness, K, compliance, C	State: $\theta=$ angle Rate: $\dot{\theta}=\omega$ Constitutive: $T=T(\theta)$ Energy: $U_\theta=\int T\cdot d\theta$ Co-energy: $U_T=\int \theta\cdot dT$

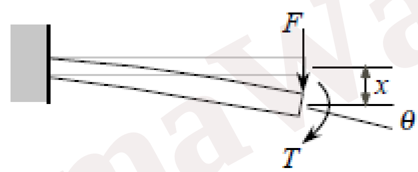
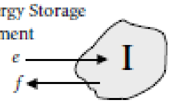
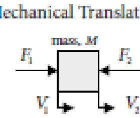
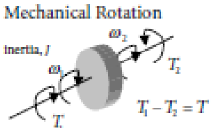


Figure 3.58: Example of potential energy storing element in a cantilevered beam with translation and rotational end connections.

Table 3.7: Mechanical Kinetic Energy Storage Elements (Integral Form)

Physical System	Fundamental Relations
Generalized Kinetic Energy Storage Element  • Inertive element • Inertance, I	State: $\mathbf{p}=$ momentum Rate: $\dot{\mathbf{p}}=\mathbf{e}$ Constitutive: $\mathbf{f}=\Phi(\mathbf{p})$ Energy: $T_p=\int \mathbf{f}\cdot d\mathbf{p}$ Co-energy: $T_f=\int \mathbf{p}\cdot d\mathbf{f}$
Mechanical Translation  $F_1-F_2=F$ $V_1=V_2=V$ • Mass • mass, m	State: $p=$ momentum Rate: $\dot{p}=F$ Constitutive: $V=V(p)$ Energy: $T_p=\int f\cdot dp$ Co-energy: $T_V=\int p\cdot dV$
Mechanical Rotation  $T_1-T_2=T$ $\omega_1=\omega_2=\omega$ • Rotational inertia • mass moment of inertia, J	State: $h=$ angular momentum Rate: $\dot{h}=T$ Constitutive: $\omega=\omega(h)$ Energy: $T_h=\int \omega\cdot dh$ Co-energy: $T_\omega=\int h\cdot d\omega$

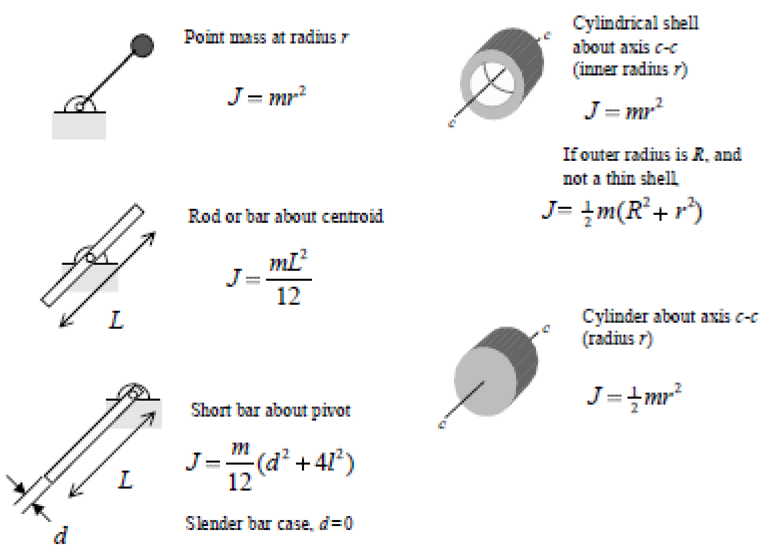


Figure 3.59: Mass moments of inertia for some common bodies

3.3.2 Translational Example

EXAMPLE 4-1-1 Consider the mass-spring-friction system shown in Fig. 4-5(a). The linear motion concerned is in the horizontal direction. The free-body diagram of the system is shown in Fig. 4-5(b). The force equation of the system is

$$f(t) - B\frac{dy(t)}{dt} - Ky(t) = M\frac{d^2y(t)}{dt^2} \qquad (4\text{-}9)$$

The last equation may be rearranged by equating the highest-order derivative term to the rest of the terms:

$$\frac{d^2y(t)}{dt^2} = -\frac{B}{M}\frac{dy(t)}{dt} - \frac{K}{M}y(t) + \frac{1}{M}f(t) \qquad (4\text{-}10)$$

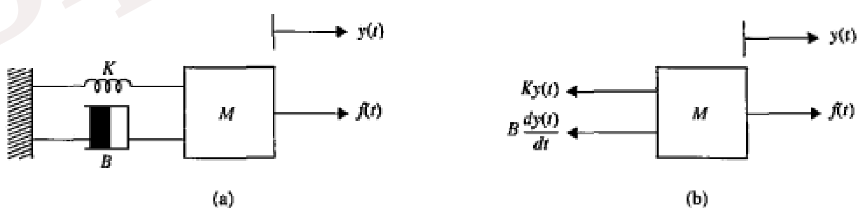


Figure 4-5 (a) Mass-spring-friction system. (b) Free-body diagram.

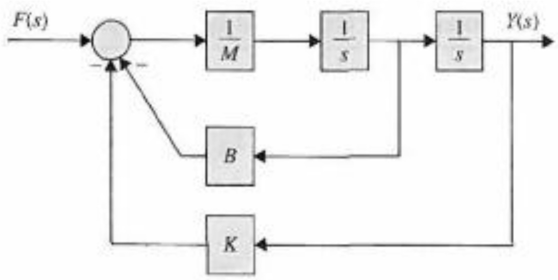


Figure 4-6 The mass-spring-friction system of Eq. (4-11) block diagram representation.

where $\dot{y}(t) = \left(\frac{dy(t)}{dt}\right)$ and $\ddot{y}(t) = \left(\frac{d^2y(t)}{dt^2}\right)$ represent velocity and acceleration, respectively. Or, alternatively, the former equation may be rewritten into an input–output form as

$$\ddot{y}(t) + \frac{B}{M}\dot{y}(t) + \frac{K}{M}y(t) = \frac{1}{M}f(t) \tag{4-11}$$

where $y(t)$ is the output and $\frac{f(t)}{M}$ is considered the input.

For zero initial conditions, the transfer function between $Y(s)$ and $F(s)$ is obtained by taking the Laplace transform on both sides of Eq. (4-11) with zero initial conditions:

$$\frac{Y(s)}{F(s)} = \frac{1}{Ms^2 + Bs + K} \tag{4-12}$$

The same result is obtained by applying the gain formula to the block diagram, which is shown in Fig. 4-6.

Eq. (4-10) may also be represented in the **space state form** using a state vector $\mathbf{x(t)}$ having n rows, where n is the number of state variables, so that

$$\dot{\mathbf{x}} = \mathbf{Ax} + \mathbf{Bu} \tag{4-13}$$

where

$$\mathbf{x(t)} = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} \tag{4-14}$$

$$y(t) = x_1(t) \quad \dot{y}(t) = x_2(t) \tag{4-15}$$

and

$$\mathbf{u(t)} = \frac{f(t)}{M} \tag{4-16}$$

So using Eqs. (4-13) through (4-16), Eq. (4-10) is rewritten in vectoral form as

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{pmatrix} 0 & 1 \\ -\frac{K}{M} & -\frac{B}{M} \end{pmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \frac{f(t)}{M} \tag{4-17}$$

The state Eq. (4-17) may also be written as a set of first-order differential equations:

$$\begin{aligned} \frac{dx_1(t)}{dt} &= x_2(t) \\ \frac{dx_2(t)}{dt} &= -\frac{K}{M}x_1(t) - \frac{B}{M}x_2(t) + \frac{1}{M}f(t) \\ y(t) &= x_1(t) \end{aligned} \tag{4-18}$$

For zero initial conditions, the transfer function between $Y(s)$ and $F(s)$ is obtained by taking the Laplace transform on both sides of Eq. (4-18):

$$\begin{aligned} sX_1(s) &= X_2(s) \\ sX_2(s) &= -\frac{B}{M}X_2(s) - \frac{K}{M}X_1(s) + \frac{1}{M}F(s) \\ Y(s) &= X_1(s) \\ \frac{Y(s)}{F(s)} &= \frac{1}{Ms^2 + Bs + K} \end{aligned} \tag{4-19}$$

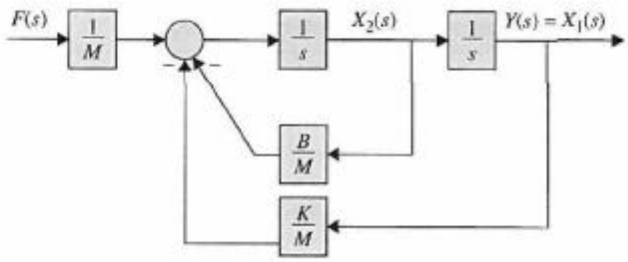


Figure 4-7 Block diagram representation of mass-spring-friction system of Eq. (4-19).

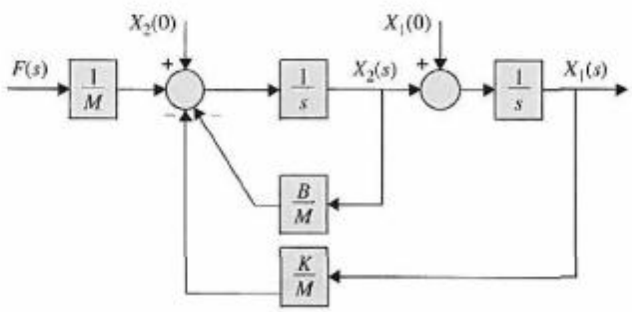


Figure 4-8 Block diagram representation of mass-spring-friction system of Eq. (4-20) with initial conditions $x_1(0)$ and $x_2(0)$.

The same result is obtained by applying the gain formula to the block diagram representation of the system in Eq. (4-19), which is shown in Fig. 4-7.

For nonzero initial conditions, Eq. (4-18) has a different Laplace transform representation that may be written as:

$$
\begin{aligned}
sX_1(s) - x_1(0) &= X_2(s) \\
sX_2(s) - x_2(0) &= -\frac{B}{M}X_2(s) - \frac{K}{M}X_1(s) + \frac{1}{M}F(s) \qquad (4\text{-}20) \\
Y(s) &= X_1(s)
\end{aligned}
$$

Upon simplifying Eq. (4-20) or by applying the gain formula to the block diagram representation of the system, shown in Fig. 4-8, the output becomes

$$Y(s) = \frac{1}{Ms^2 + Bs + K}F(s) + \frac{Ms}{Ms^2 + Bs + K}x_1(0) + \frac{M}{Ms^2 + Bs + K}x_2(0) \qquad (4\text{-}21)$$

EXAMPLE 4-1-2 As another example of writing the dynamic equations of a mechanical system with translational motion, consider the system shown in Fig. 4-9(a). Because the spring is deformed when it is subject to a force $f(t)$, two displacements, y_1 and y_2, must be assigned to the end points of the spring. The free-body diagrams of the system are shown in Fig. 4-9(b). The force equations are

$$f(t) = K[y_1(t) - y_2(t)] \tag{4-22}$$

$$-K[y_2(t) - y_1(t)] - B\frac{dy_2(t)}{dt} = M\frac{d^2y_2(t)}{dt^2} \tag{4-23}$$

These equations are rearranged in input–output form as

$$\frac{d^2y_2(t)}{dt^2} + \frac{B}{M}\frac{dy_2(t)}{dt} + \frac{K}{M}y_2(t) = \frac{K}{M}y_1(t) \tag{4-24}$$

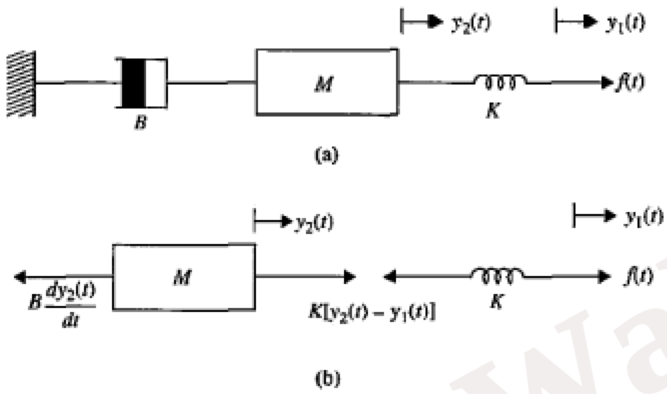


Figure 4-9 Mechanical system for Example 4-1-2. (a) Mass-spring-damper system. (b) Free-body diagram.

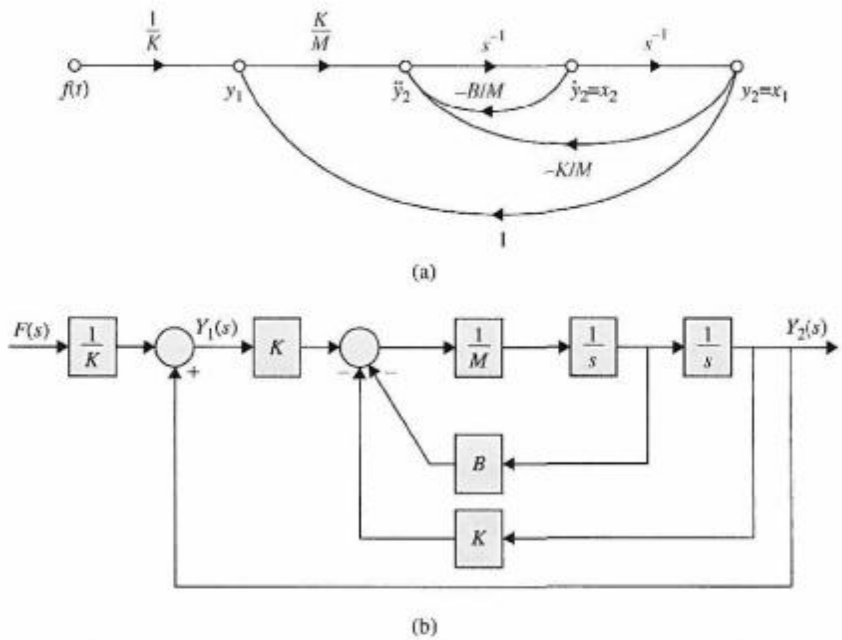


Figure 4-10 Mass-spring-friction system of Eq. (4-25) using Eq. (4-22). (a) The signal-flow graph representation. (b) Block diagram representation.

For zero initial conditions, the transfer function between $Y_1(s)$ and $Y_2(s)$ is obtained by taking the Laplace transform on both sides of Eq. (4-24):

$$\frac{Y_2(s)}{Y_1(s)} = \frac{K}{Ms^2 + Bs + K} \tag{4-25}$$

The same result is obtained by applying the gain formula to the block diagram representation of the system, which is shown in Fig. 4-10. Note that in Fig. 4-10, Eq. (4-22) was also used.

For state representation, these equations may be rearranged as

$$\begin{aligned} y_1(t) &= y_2(t) + \frac{1}{K} f(t) \\ \frac{d^2 y_2(t)}{dt^2} &= -\frac{B}{M}\frac{dy_2(t)}{dt} + \frac{K}{M}[y_1(t) - y_2(t)] \end{aligned} \tag{4-26}$$

For zero initial conditions, the transfer function of Eq. (4-26) is the same as Eq. (4-25). By using the last two equations, the state variables are defined as $x_1(t) = y_2(t)$ and $x_2(t) = dy_2(t)/dt$. The state equations are therefore written as

$$\begin{aligned} \frac{dx_1(t)}{dt} &= x_2(t) \\ \frac{dx_2(t)}{dt} &= -\frac{B}{M}x_2(t) + \frac{1}{M} f(t) \end{aligned} \tag{4-27}$$

The same result is obtained after taking the Laplace transform of Eq. (4-27) and applying the gain formula to the block diagram representation of the system, which is shown in Fig. 4-11. Note that in Fig. 4-11, $F(s)$, $Y_1(s)$, $X_1(s)$, $Y_2(s)$, and $X_2(s)$ are Laplace transforms of $f(t)$, $y_1(t)$, $x_1(t)$, $y_2(t)$, and $x_2(t)$, respectively.

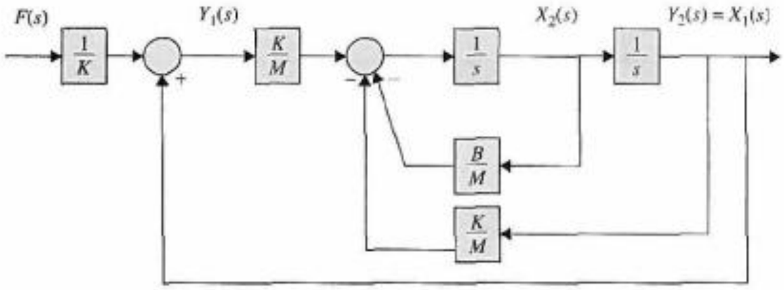


Figure 4-11 Block diagram representation of mass-spring-friction system of Eq. (4-27).

EXAMPLE 4-1-3 Consider the two degrees of freedom (2-DOF) spring-mass system, with two masses m_1 and m_2, two springs k_1 and k_2, and two forces f_1 and f_2, as shown in Fig. 4-12. Find the equations of motion.

SOLUTION To avoid any confusion, we first draw the free-body diagram (FBD) of the system by assuming the masses are displaced in the positive direction, so that $y_1 > y_2 > 0$ (i.e., springs are both in tension). The FBD of the system is shown in Fig. 4-13. Applying Newton's second law to the masses M_1 and M_2, we have

$$\begin{aligned} f_1(t) - K_1 y_1 + K_2(y_1 - y_2) &= M_1 \ddot{y}_1 \\ f_2(t) - K_2(y_1 - y_2) &= M_2 \ddot{y}_2 \end{aligned} \tag{4-28}$$

Rearranging the equations into the standard input–output form, we have

$$\begin{aligned} M_1\ddot{y}_1 + (K_1 + K_2)y_1 - K_2 y_2 &= f_1(t) \\ M_2\ddot{y}_1 - K_2 y_1 + K_2 y_2 &= f_2(t) \end{aligned} \tag{4-29}$$

Alternatively, Eq. (4-29) may be represented in the standard second-order matrix form, as

$$\begin{bmatrix} M_1 & 0 \\ 0 & M_2 \end{bmatrix}\begin{bmatrix} \ddot{y}_1 \\ \ddot{y}_2 \end{bmatrix} + \begin{bmatrix} K_1 + K_2 & -K_2 \\ -K_2 & K_2 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \tag{4-30}$$

In state space form, assuming the following state vector $\mathbf{x(t)}$, the inputs $u_1(t)$ and $u_2(t)$, and the output $y(t)$, we get

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ x_4(t) \end{bmatrix} = \begin{bmatrix} y_1(t) \\ y_2(t) \\ \dot{y}_1(t) \\ \dot{y}_2(t) \end{bmatrix}, \; u_1 = f_1(t),\; u_2 = f_2(t),\; y(t) = x_1(t) \tag{4-31}$$

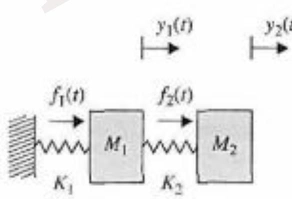


Figure 4-12 A 2-DOF spring-mass system.

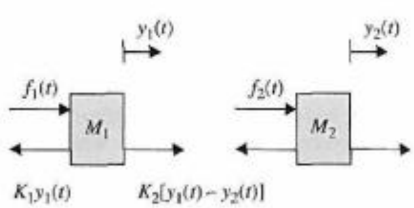


Figure 4-13 FBD of the 2-DOF spring-mass system.

Then, using $\dot{x}_3 = \ddot{y}_1$ and $\dot{x}_4 = \ddot{y}_2$, we get the state-space representation as

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -K_1/M_1 & K_1/M_1 & 0 & 0 \\ K_2/M_1 & -K_2/M_1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1/M_1 \\ 0 \end{bmatrix} u_1 + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1/M_2 \end{bmatrix} u_2 \quad \text{(state equation)}$$

$$y = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + 0 \cdot u_1 + 0 \cdot u_2 \quad \text{(output equation)}$$

$$(4\text{-}32)$$

where the state equation is a set of four first-order differential equations.

3.3.3 Rotational Example

EXAMPLE 4-1-4 The rotational system shown in Fig. 4-16(a) consists of a disk mounted on a shaft that is fixed at one end. The moment of inertia of the disk about the axis of rotation is J. The edge of the disk is riding on the surface, and the viscous friction coefficient between the two surfaces is B. The inertia of the shaft is negligible, but the torsional spring constant is K.

Assume that a torque is applied to the disk, as shown; then the torque or moment equation about the axis of the shaft is written from the free-body diagram of Fig. 4-16(b):

$$T(t) = J\frac{d^2\theta(t)}{dt^2} + B\frac{d\theta(t)}{dt} + K\theta(t) \tag{4-41}$$

Notice that this system is analogous to the translational system in Fig. 4-5. The state equations may be written by defining the state variables as $x_1(t) = \theta(t)$ and $x_2(t) = dx_1(t)/dt$.

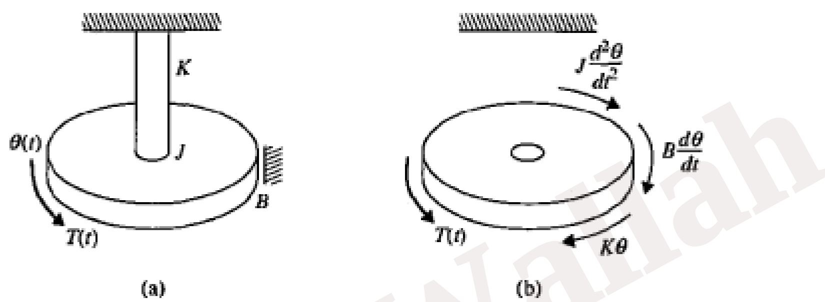


Figure 4-16 Rotational system for Example 4-1-4.

TABLE 4-2 Basic Rotational Mechanical System Properties and Their Units

Parameter	Symbol Used	SI Units	Other Units	Conversion Factors
Inertia	J	kg-m^2	slug-ft^2 lb-ft-sec^2 oz-in.-sec^2	1 g-cm $=$ 1.417×10^{-5} oz-in.-sec^2 1 lb-ft-sec^2 $=$ 192 oz-in.-sec^2 $=$ 32.2 lb-ft^2 1 oz-in.-sec^2 $=$ 386 oz-in^2 1 g-cm-sec^2 $=$ 980 g-cm^2
Angular Displacement	T	Radian	Radian	1 rad $= \frac{180}{\pi} = 57.3$ deg
Angular Velocity	O	radian/sec	radian/sec	1 rpm $= \frac{2\pi}{60}$ $=$ 0.1047 rad/sec 1 rpm $=$ 6 deg/sec
Angular Acceleration	A	radian/sec^2	radian/sec^2	
Torque	T	(N-m) dyne-cm	lb-ft oz-in.	1 g-cm $=$ 0.0139 oz-in. 1 lb-ft $=$ 192 oz-in. 1 oz-in. $=$ 0.00521 lb-ft
Spring Constant	K	N-m/rad	ft-lb/rad	
Viscous Friction Coefficient	B	N-m/rad/sec	ft-lb/rad/sec	
Energy	Q	J (joules)	Btu Calorie	1 J $=$ 1 N-m 1 Btu $=$ 1055 J 1 cal $=$ 4.184 J

EXAMPLE 4-1-5 Fig. 4-17(a) shows the diagram of a motor coupled to an inertial load through a shaft with a spring constant K. A non-rigid coupling between two mechanical components in a control system often causes torsional resonances that can be transmitted to all parts of the system. The system variables and parameters are defined as follows:

$T_m(t)$ = motor torque
B_m = motor viscous-friction coefficient
K = spring constant of the shaft
$\theta_m(t)$ = motor displacement
$\omega_m(t)$ = motor velocity

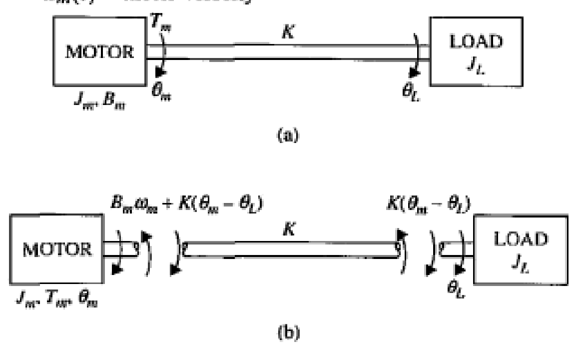


Figure 4-17 (a) Motor–load system. (b) Free-body diagram.

J_m = motor inertia
$\theta_L(t)$ = load displacement
$\omega_L(t)$ = load velocity
J_L = load inertia

The free-body diagrams of the system are shown in Fig. 4-17(b). The torque equations of the system are

$$\frac{d^2\theta_m(t)}{dt^2} = -\frac{B_m}{J_m}\frac{d\theta_m(t)}{dt} - \frac{K}{J_m}[\theta_m(t) - \theta_L(t)] + \frac{1}{J_m}T_m(t) \tag{4-42}$$

$$K[\theta_m(t) - \theta_L(t)] = J_L\frac{d^2\theta_L(t)}{dt^2} \tag{4-43}$$

In this case, the system contains three energy-storage elements in J_m, J_L, and K. Thus, there should be three state variables. Care should be taken in constructing the state diagram and assigning the state variables so that a minimum number of the latter are incorporated. Eqs. (4-42) and (4-43) are rearranged as

$$\frac{d^2\theta_m(t)}{dt^2} = -\frac{B_m}{J_m}\frac{d\theta_m(t)}{dt} - \frac{K}{J_m}[\theta_m(t) - \theta_L(t)] + \frac{1}{J_m}T_m(t) \tag{4-44}$$

$$\frac{d^2\theta_L(t)}{dt^2} = \frac{K}{J_L}[\theta_m(t) - \theta_L(t)] \tag{4-45}$$

The state variables in this case are defined as $x_1(t) = \theta_m(t) - \theta_L(t)$, $x_2(t) = d\theta_L(t)/dt$, and $x_3(t) = d\theta_m(t)/dt$. The state equations are

$$\begin{aligned}
\frac{dx_1(t)}{dt} &= x_3(t) - x_2(t)\\
\frac{dx_2(t)}{dt} &= \frac{K}{J_L}x_1(t)\\
\frac{dx_3(t)}{dt} &= -\frac{K}{J_m}x_1(t) - \frac{B_m}{J_m}x_3(t) + \frac{1}{J_m}T_m(t)
\end{aligned} \tag{4-46}$$

The SFG representation is shown in Fig. 4-18.

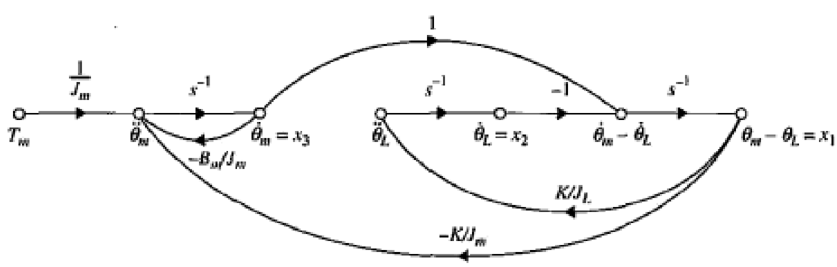


Figure 4-18 Rotational system of Eq. (4-46) signal-flow graph representation.

3.3.4 Electrical Analog

Systems that can be represented by the same mathematical model but that are different physically are called analogous systems. Thus analogous systems are described by the same differential or integral differential equations or set of equations. The concept of analogous systems is very useful in practice for the following reasons.

- The solution of the equation describing one physical system can be directly applied to analogous systems in any other field.
- Since one type of system may be easier to handle experimentally than another, instead of building and studying a mechanical system (or hydraulic system or pneumatic system), we can build and study its electrical analog, for electrical or electronic systems are, in general, much easier to deal with experimentally.

This section presents analogies between mechanical and electrical systems. The concept of analogous systems, however, is applicable to other kinds of systems, and analogies among mechanical, electrical, hydraulic, pneumatic, thermal, and other systems may be established.

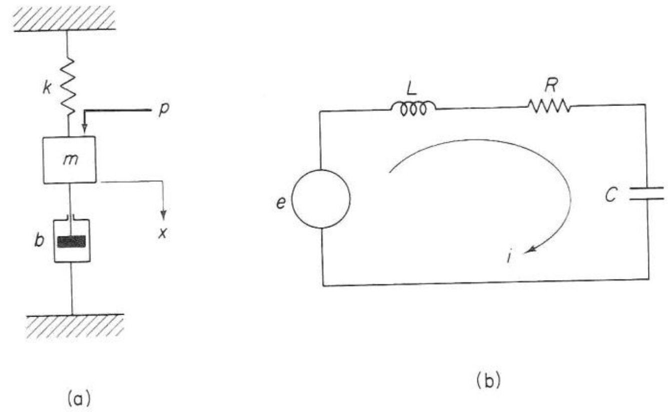


Figure 2–18
Analogous mechanical and electrical systems.

Mechanical–electrical analogies. Mechanical systems can be studied through the use of their electrical analogs, which may be more easily constructed than models of the corresponding mechanical systems. There are two electrical analogies for mechanical systems: the force–voltage analogy and the force–current analogy.

Force–voltage analogy. Consider the mechanical system of Figure 2–18(a) and the electrical system of Figure 2–18(b). The system equation for the former is

$$m\frac{d^2x}{dt^2} + b\frac{dx}{dt} + kx = p \tag{2–31}$$

whereas the system equation for the latter is

$$L\frac{di}{dt} + Ri + \frac{1}{C}\int i\,dt = e$$

In terms of electric charge q, this last equation becomes

$$L\frac{d^2q}{dt^2} + R\frac{dq}{dt} + \frac{1}{C}q = e \tag{2–32}$$

Comparing Equations (2–31) and (2–32), we see that the differential equations for the two systems are of identical form. Thus these two systems are analogous systems. The terms that occupy corresponding positions in the differential equations are called *analogous quantities*, a list of which appears in Table 2–2. The analogy here is called the *force–voltage analogy* (or mass–inductance analogy).

Table 2–2 Force–Voltage Analogy

Mechanical Systems	Electrical Systems
Force p (torque T)	Voltage e
Mass m (moment of inertia J)	Inductance L
Viscous-friction coefficient b	Resistance R
Spring constant k	Reciprocal of capacitance, $1/C$
Displacement x (angular displacement θ)	Charge q
Velocity $\dot{x}$ (angular velocity $\dot{\theta}$)	Current i

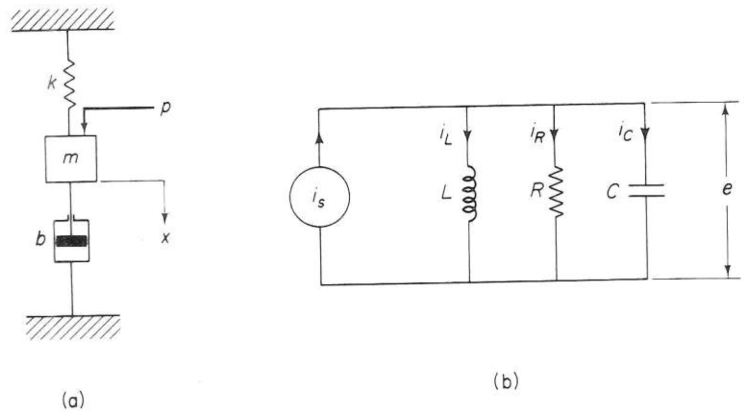


Figure 2–19
Analogous mechanical and electrical systems.

Force–current analogy. Another analogy between electrical and mechanical systems is based on the force–current analogy. Consider the mechanical system shown in Figure 2–19(a). The system equation can be obtained as

$$m\frac{dx^2}{dt^2} + b\frac{dx}{dt} + kx = p \tag{2–33}$$

Consider next the electrical system shown in Figure 2–19(b). Application of Kirchhoff's current law gives

$$i_L + i_R + i_C = i_s \tag{2–34}$$

where

$$i_L = \frac{1}{L}\int e\,dt, \qquad i_R = \frac{e}{R}, \qquad i_C = C\frac{de}{dt}$$

Equation (2–34) can be written as

$$\frac{1}{L}\int e\,dt + \frac{e}{R} + C\frac{de}{dt} = i_s \tag{2–35}$$

Since the magnetic flux linkage ψ is related to voltage e by the equation

$$\frac{d\psi}{dt} = e$$

in terms of ψ, Equation (2–35) can be written as

$$C\frac{d^2\psi}{dt^2} + \frac{1}{R}\frac{d\psi}{dt} + \frac{1}{L}\psi = i_s \tag{2–36}$$

Comparing Equations (2–33) and (2–36), we find that the two systems are analogous. The analogous quantities are listed in Table 2–3. The analogy here is called the *force–current analogy* (or mass–capacitance analogy).

It should be remembered that analogies between two systems may break down if the regions of operation are extended too far. In other words, since the differential equations

Table 2–3 Force–Current Analogy

Mechanical Systems	Electrical Systems
Force p (torque T)	Current i
Mass m (moment of inertia J)	Capacitance C
Viscous-friction coefficient b	Reciprocal of resistance, $1/R$
Spring constant k	Reciprocal of inductance, $1/L$
Displacement x (angular displacement θ)	Magnetic flux linkage ψ
Velocity $\dot{x}$ (angular velocity $\dot{\theta}$)	Voltage e

upon which the analogies are based are only approximations to the dynamic characteristics of physical systems in a certain operating region, the analogy may break down if the operating region of one system is very wide. If the operating region of a given mechanical system is wide, however, it may be divided into two or more subregions, and analogous electrical systems may be built for each subregion. As a matter of fact, analogies are not limited to electrical systems and mechanical systems; they are applicable to any systems as long as their differential equations, or transfer functions, are of identical form.

Comparing Eqs. (4-11), (4-41), and (4-65), it is not difficult to see that the mechanical systems in Eqs. (4-11) and (4-41) are analogous to a series ***RLC*** electric network shown in Example 4-2-1. As a result, with this analogy, mass M and inertia J are analogous to inductance L, the spring constant K is analogous to the inverse of capacitance $1/C$, and the viscous-friction coefficient B is analogous to resistance R.

➢ **EXAMPLE 4-10-1** It is logical, in Example 4-1-1, to assign $v(t)$, the velocity, and $f_k(t)$, the force acting on the spring, as state variables, since the former is analogous to the current in L and the latter is analogous to the voltage across C. Writing the force on M and the velocity of the spring as functions of the state variables and the input force $f(t)$, we have

Force on mass:

$$M\frac{dv(t)}{dt} = -f_k(t) - Bv(t) + f(t) \qquad (4\text{-}287)$$

Velocity of spring:

$$\frac{1}{K}\frac{df_k(t)}{dt} = v(t) \qquad (4\text{-}288)$$

The final equation of motion Eq. (4-11) may be obtained by dividing both sides of Eq. (4-287) by M and multiplying Eq. (4-288) by K. Hence, in terms of displacement $y(t)$,

$$\frac{d^2y(t)}{dt^2} + \frac{B}{M}\frac{dy(t)}{dt} + \frac{K}{M}y(t) = \frac{f(t)}{M} \qquad (4\text{-}289)$$

Considering Example 4-2-1, after rewriting Eq. (4-67) as

$$L\frac{di(t)}{dt} = -e_c(t) - Ri(t) + e(t) \qquad (4\text{-}290)$$

and using the current relation Eq. (4-66):

$$C\frac{de_c(t)}{dt} = i(t) \qquad (4\text{-}291)$$

the comparison of Eq. (4-287) with Eq. (4-290) and Eq. (4-288) with Eq. (4-291) clearly shows the analogies among the mechanical and electrical components.

EXAMPLE 4-10-2 As another example of writing the dynamic equations of a mechanical system with translational motion, consider the system shown in Fig. 4-9(a). Because the spring is deformed when it is subject to a force $f(t)$, two displacements, y_1 and y_2, must be assigned to the end points of the spring. The

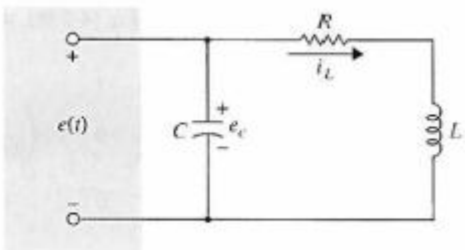


Figure 4-76 Electric network analogous to the mechanical system in Fig. 4-10.

free-body diagram of the system is shown in Fig. 4-9(b). The force equations are

$$f(t) = K[y_1(t) - y_2(t)] \tag{4-292}$$

$$K[y_1(t) - y_2(t)] = M\frac{d^2y_2(t)}{dt^2} + B\frac{dy_2(t)}{dt} \tag{4-293}$$

These equations are rearranged as

$$y_1(t) = y_2(t) + \frac{1}{K}f(t) \tag{4-294}$$

$$\frac{d^2y_2(t)}{dt^2} = -\frac{B}{M}\frac{dy_2(t)}{dt} + \frac{K}{M}[y_1(t) - y_2(t)] \tag{4-295}$$

By using the last two equations, the SFG diagram of the system is drawn in Fig. 4-10(a). The state variables are defined as $x_1(t) = y_2(t)$ and $x_2(t) = dy_2(t)/dt$. The state equations are written directly from the state diagram:

$$\frac{dx_1(t)}{dt} = x_2(t) \tag{4-296}$$

$$\frac{dx_2(t)}{dt} = -\frac{B}{M}x_2(t) + \frac{1}{M}f(t) \tag{4-297}$$

As an alternative, we can assign the velocity $v(t)$ of the mass M as one state variable and the force $f_k(t)$ on the spring as the other state variable. We have

$$\frac{dv(t)}{dt} = -\frac{B}{M}v(t) + \frac{1}{M}f_k(t) \tag{4-298}$$

$$f_k(t) = f(t) \tag{4-299}$$

One may wonder why there is only one state equation in Eq. (4-287), whereas there are two state variables in $v(t)$ and $f_k(t)$. The two state equations of Eqs. (4-296) and (4-297) clearly show that the system is of the second order. The situation is better explained by referring to the analogous electric network of the system shown in Fig. 4-76. Although the network has two energy-storage elements in L and C, and thus there should be two state variables, the voltage across the capacitance $e_c(t)$ in this case is redundant, since it is equal to the applied voltage $e(t)$. Eqs. (4-298) and (4-297) can provide only the solutions to the velocity of M, $v(t)$, which is the same as $dy_2(t)/dt$, once $f(t)$ is specified. Then $y_2(t)$ is determined by integrating $v(t)$ with respect to t. The displacement $y_1(t)$ is then found using Eq. (4-292). On the other hand, Eqs. (4-296) and (4-297) give the solutions to $y_2(t)$ and $dy_2(t)/dt$ directly, and $y_1(t)$ is obtained from Eq. (4-292).

The transfer functions of the system are obtained by applying the gain formula to the state diagram.

$$\frac{Y_2(s)}{F(s)} = \frac{1}{s(Ms + B)} \tag{4-300}$$

$$\frac{Y_1(s)}{F(s)} = \frac{Ms^2 + Bs + K}{Ks(Ms + B)} \tag{4-301}$$

EXAMPLE 4-10-3 A Pneumatic System Dry air passes through a valve into a rigid 1 m^3 container, as shown in Fig. 4-77, at a constant temperature T = 25°C(= 298°K). The pressure at the left-hand side of the valve is p_i, which is higher than the pressure in the tank p. Assuming a laminar flow, the valve resistance becomes linear, $R = 200\,\text{sec/m}^2$. Find the time constant of the system.

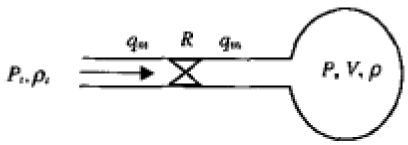


Figure 4-77 A pneumatic system with a valve and a spherical rigid tank.

SOLUTION Assuming air as an ideal gas, isothermal process, and low pressures, from Example 4-5-5, the equation of the system is

$$\frac{RV}{R_{\text{air}}T}\dot{p} + p = p_i \qquad (4\text{-}302)$$

where air at standard pressure and temperature is represented as an ideal gas,

$$pv = \frac{p}{\rho} = R_{\text{air}}T \quad \rho = \frac{1}{R_{\text{air}}T}p \qquad (4\text{-}303)$$

Thus, the time constant is

$$\tau = \frac{RV}{R_{air}T} = \frac{(200)(1)}{88.63(298)} = 7.5(10^{-3})\text{sec} \qquad (4\text{-}304)$$

where, from reference [1] at the end of this chapter,

$$R_{\text{air}} = 53.35\frac{\text{ft lb}_f}{\text{lb}_m\,°\text{R}}\,\frac{0.3048\text{ m}}{\text{ft}}\,\frac{4.45\text{ N}}{\text{lb}_f}\,\frac{\text{kg m/sec}^2}{\text{N}}\,\frac{\text{lb}_m}{0.4536\text{ kg}}\,\frac{°\text{R}}{°\text{K}(9/5)} = 88.63\frac{\text{m}^2}{\text{sec}^2\,°\text{K}}$$

EXAMPLE 4-10-4 A One-Tank Liquid-Level System For the liquid-level system shown in Fig. 4-45, $C = A/g$ is the capacitance and $\rho = R$ is the resistance. As a result, system time constant is $\tau = RC$. Comparing the thermal, fluid, and electrical systems, similar analogies may be obtained, as shown in Table 4-8.

TABLE 4-8 Mechanical, Thermal, and Fluid Systems and Their Electrical Equivalents

System	R, C, L	Analogy
Mechanical (translation)	$F = Bv(t)$ $R = B$ $F = K\int v(t)dt$ $C = \frac{1}{K}$ $v(t) = \frac{1}{M}\int F\,dt$ $L = M$	$e => F$ $i(t) => v(t)$ where e = voltage $i(t)$ = current F = force $v(t)$ = linear velocity
Mechanical (rotation)	$T = B\omega(t)$ $R = B$ $T = K\int \omega(t)dt$ $C = \frac{1}{K}$ $\omega = \frac{1}{J}T\,dt$ $L = J$	$e => T$ $i(t) => \omega(t)$ where e = voltage $i(t)$ = current T = torque $\omega(t)$ = angular velocity
Fluid (incompressible)	$\Delta P = Rq(t)$ (laminar flow) R depends on flow regime $q(t) = C\dot{P}$ C depends on flow regime $L = \frac{\rho \ell}{A}$ (flow in a pipe) where A = area of cross section l = length ρ = fluid density	$e => \Delta P$ $i(t) => q(t)$ where e = voltage $i(t)$ = current P = pressure $q(t)$ = volume flow rate
Thermal	$R = \frac{\Delta T}{q}$ $T = \frac{1}{C}\int q\,dt$	$e => T$ $i(t) => q(t)$ where e = voltage $i(t)$ = current T = temperature $q(t)$ = heat flow

3.4 Define the End Effectors

The Wikipedia states:
"In robotics, an end effector is the device at the end of a robotic arm, designed to interact with the environment."

ATI Industrial Automation says:
"A robotic end-effector is any object attached to the robot flange (wrist) that serves a function."

In robotics, an end effector is the device at the end of a robotic arm, designed to interact with the environment. The exact nature of this device depends on the application of the robot. In the strict definition, which originates from serial robotic manipulators, the end effector means the last link (or end) of the robot. At this endpoint the tools are attached.

In a wider sense, an end effector can be seen as the part of a robot that interacts with the work environment. This does not refer to the wheels of a mobile robot or the feet of a humanoid robot which are also not end effectors—they are part of the robot's mobility.

End effectors may *consist of* a gripper or a tool. The gripper can be of two fingers, three fingers or even five fingers.

The end effectors that can be used as tools serve various purposes. Such as, spot welding in an assembly, spray painting where uniformity of painting is necessary and for other purposes where the working conditions are dangerous for human beings. Surgical robots have end effectors that are specifically manufactured for performing surgeries. This would include *robotic grippers, robotic tool changers, robotic collision sensors, robotic rotary joints, robotic press tooling, compliance devices, robotic paint guns, robotic deburring tools, robotic arc welding guns, robotic transguns*, etc.

Grippers are active links between the handling equipment and the work piece or in a more general sense between the grasping organ (normally the gripper fingers) and the object to be acquired. Their functions depend on specific applications and include:

- Temporary maintenance of a definite position and orientation of the work piece relative to the gripper and the handling equipment.
- Retaining of static (weight), dynamic (motion, acceleration or deceleration) or process specific forces and moments.
- Determination and change of position and orientation of the object relative to the handling equipment by means of wrist axes.
- Specific technical operations performed with, or in conjunction with, the gripper.

Grippers are not only required for use with industrial robots: they are a universal component in automation. Grippers operate with:

- Industrial robots (handling and manipulation of objects).
- Hard automation (assembling, micro-assembling, machining, and packaging).
- NC machines (tool change) and special purpose machines.
- Hand-guided manipulators (remote prehension, medical, aerospace, nautical)
- Work piece turret devices in manufacturing technology.
- Rope and chain lifting tools (load-carrying equipment).
- Service robots (prehension tools potentially similar to prosthetic hands).

Figure 3.60: Range of grippers fingers

In robotics technology grippers belong to the functional units having the greatest variety of designs. This is due to the fact that, although the robot is a flexible machine, the gripper performs a much more specific task. Nevertheless, these tasks are not limited to prehension alone which is why the more generic term "end-effector" is often used.

The great number of different requirements, diverse work pieces and the desire for well adapted and reliable systems will continue to stimulate further developments in future gripper design. Many experts consider the capabilities of the gripper as an essential factor for the economic effectiveness of automatic assembly systems. Experience indicates that in the future it will only be possible to respond to practical demands if flexible designs for assembly equipment are available. Consequently, grippers must become ever more flexible. Assembly relates not only to prehension and manipulation of objects but also to pressing, fitting and joining operations. Many grippers are employed for the loading of manufacturing lines, in packaging and storage as well as the handling of objects in laboratory test and inspection systems.

More recently, miniaturized grippers have been developed in order to handle delicate components in micro-technology. This has gone hand in hand with the emergence of many novel prehension methods. The number of grippers used in nonindustrial areas, e.g. in civil engineering, space research, handicraft, medical and pharmaceutical engineering is steadily increasing. Hand-guided (tele-operation) or automatic manipulators are used in these areas primarily as handling machines. In addition to conventional grippers, for which the gripper jaws are shaped according to the work piece profile, there exist numerous application specific grippers. This explains why an overwhelming proportion of corresponding patent literature is devoted to prehension concepts of unconventional design. In general, end-effectors are not normally within the delivery remit of robot manufacturers. Depending on the specific requirements, they are selected as accessories from tooling manufacturers or specially designed for the given purpose.

Definitions and Conceptual Basics: Grasping organs or tools constitute the end of the kinematic chain in the joint system of an industrial robot and facilitate interaction with the work environment. Although universal grippers with wide clamping ranges can be used for diverse object shapes, in many cases they must be adapted to the specific work piece shape.

Grippers are subsystems of handling mechanisms which provide temporary contact with the object to be grasped. They ensure the position and orientation when carrying and mating the object to the handling equipment. Prehension is achieved by force producing and form matching elements. The term "gripper" is also used in cases where no actual grasping, but rather holding of the object as e.g. in vacuum suction where the retention force can act on a point, line or surface.

Three of the most usual forms (impactive, astrictive and contigutive) of object prehension are depicted in six different examples in Figure 3.61.

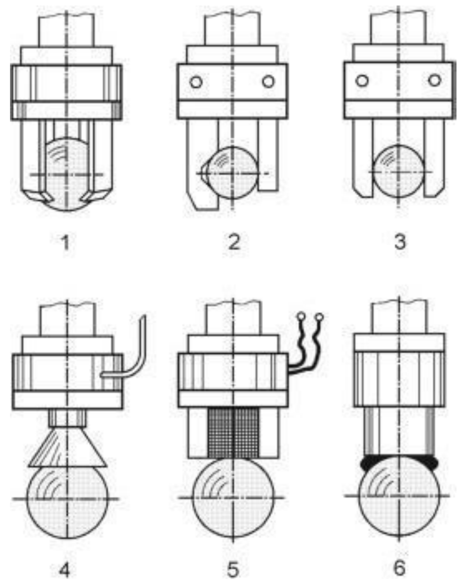


Figure 3.61: Possibilities for prehension of a spherical object. (1) pure enclosing without clamping, (2) partial form fit combine with clamping force, (3) pure force closure, (4) holding with vacuum air (pneumatic force closure), (5) retention using magnetic field (force field), (6) retention using adhensive media.

One should differentiate between grasping (prehension) and holding (retention) forces. While the grasping force is applied at the initial point of prehension (during the grasping process), the holding force maintains the grip thereafter (until object release). In the many cases the retention force may be weaker than the prehension force. The grasping force is determined by the energy required for the mechanical motion leading to a static prehension force. The functional chain drive - kinematics - holding system is given, however, only for mechanical grippers. Astrictive vacuum suction grippers require no such kinematics.

There are some characteristic terms that are often used in prehension technology. Grippers consist mostly of several modules and components. In the following, the most essential terms used will be explained considering as an example a mechanical gripper such as the one shown in Figure 3.62.

A short glossary of further important terms used in gripper technology is briefly explained below. Astrictive gripper: A binding force produced by a field is astrictive. This field may take the form of air movement (vacuum suction), magnetism or electrostatic charge displacement.

Basic jaw *(universaljaw):* The part of an impactive gripper subjected to movement. An integral part of the gripper mechanics, the basic jaw is not usually replaceable. However, the basic jaws may be fitted with additional fingers in accordance with specific requirements.

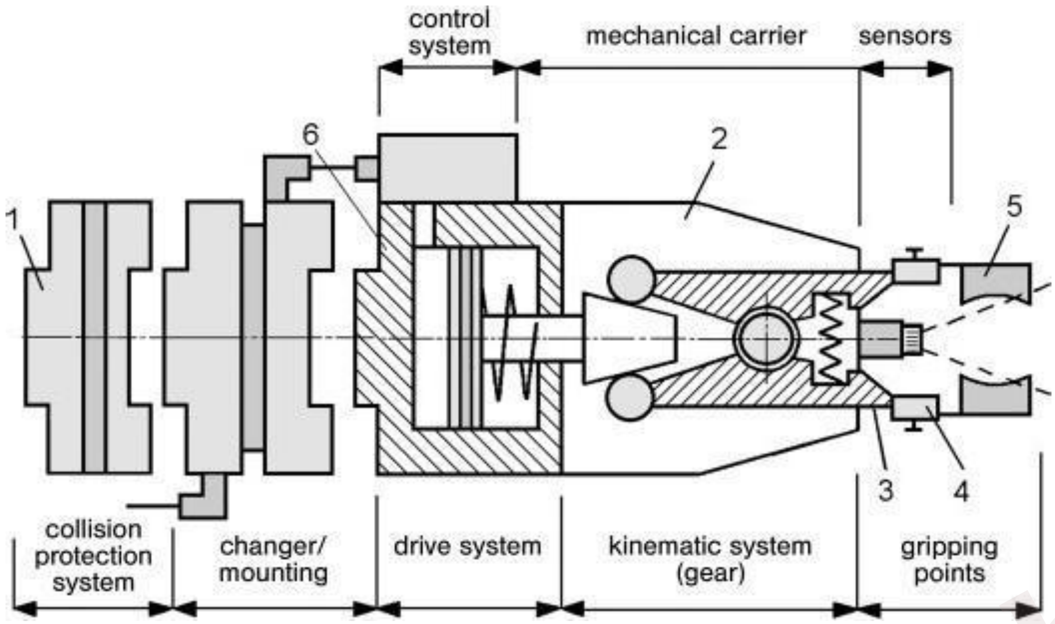


Figure 3.62: subsystems of a mechanical gripper. 1. remote center compliance, 2. carrier, 3. gripper finger, 4. basic jaw, 6. flange

Grippers are end effectors used to grasp and manipulate objects during the work cycle. The objects are usually workparts that are moved from one location to another in the cell. Machine loading and unloading applications fall into this category. Owing to the variety of part shapes, sizes, and weights, grippers must usually be custom designed.

Types of grippers used in industrial robot applications include the following:

- *mechanical grippers,* consisting of two or more fingers that can be actuated by the robot controller to open and dose to grasp the workpart; Figure 3.63 shows a two finger gripper
- *vacuum grippers,* in which suction cups are used to hold flat objects
- *magnetized devices,* for holding ferrous parts
- *adhesive devices,* where an adhesive substance is used to hold a flexible material such as a fabric
- *simple mechanical devices* such as hooks and scoops.

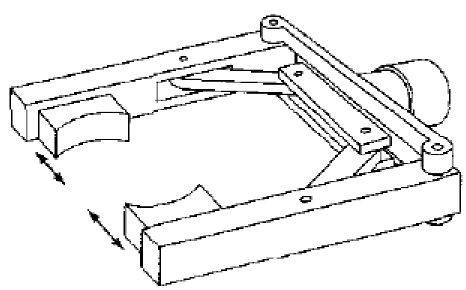


Figure 3.63: Robot mechanical gripper

Mechanical grippers are the most common gripper type. Some of the innovations and advances in mechanical gripper technology include:

- *Dual grippers,* consisting of two gripper devices in one end effector, which are useful for machine loading and unloading. With a single gripper, the robot must reach into the production machine twice, once to unload the finished part from the machine, and the second time to load the next part into the machine. With a dual gripper, the robot picks up the next workpart while the machine is still

processing the preceding part: when the machine finish, the robot reaches into the machine once to remove the finished part and load the next part. This reduces the cycle time per part.
- *interchangeable fingers* that can be used on one gripper mechanism. To accommodate different parts, different fingers are attached to the gripper.
- *Sensory feedback* in the fingers that provide the gripper with capabilities such as: (1) sensing the presence of the workpart or (2) applying a specified limited force to the workpart during gripping (for fragile workparts).
- *Multiple fingered grippers* that possess the general anatomy of a human hand.
- *Standard gripper products* that are commercially available, thus reducing the need to custom-design a gripper for each separate robot application.

Tools are used in applications where the robot must perform some processing operation on the work part. The robot therefore manipulates the tool relative to a stationary or slowly moving object (e.g., work part or subassembly). Examples of the tools used as end effectors by robots to perform processing applications include:
- spot welding gun
- arc welding tool
- spray painting gun
- rotating spindle for drilling, routing. grinding, and so forth
- assembly tool (e.g., automatic screwdriver)
- heating torch
- water jet cutting tool.

3.4.1 The Grapping problem

Robotic grasping is a complex field
- Hand design: high level (number of fingers, kinematic structure, etc.) and low-level (mechanism design, motors, materials, etc.);
- Hand control algorithms: high level (find an appropriate posture for a given task) and low-level (execute the desired posture);
- Information from sensors (tactile, vision, range sensing, etc.);
- Any pre-existing knowledge of objects shape, semantics and tasks (e.g. a cup is likely to be found on a table, should not be held upside-down, etc.);

Human Grasping vs. Robotic Grasping
- Human performance provides both a benchmark to compare against, and a working example that we can attempt to learn from. However, it has proven very elusive to replicate:
- The human hand is a very complex piece of equipment, with amazing capabilities;
- Humans benefit from an unmatched combination of visual and tactile sensing;
- Human continuously practice grasping and manipulation, the amount of data they are exposed to dwarfs anything tried so far in robotics;

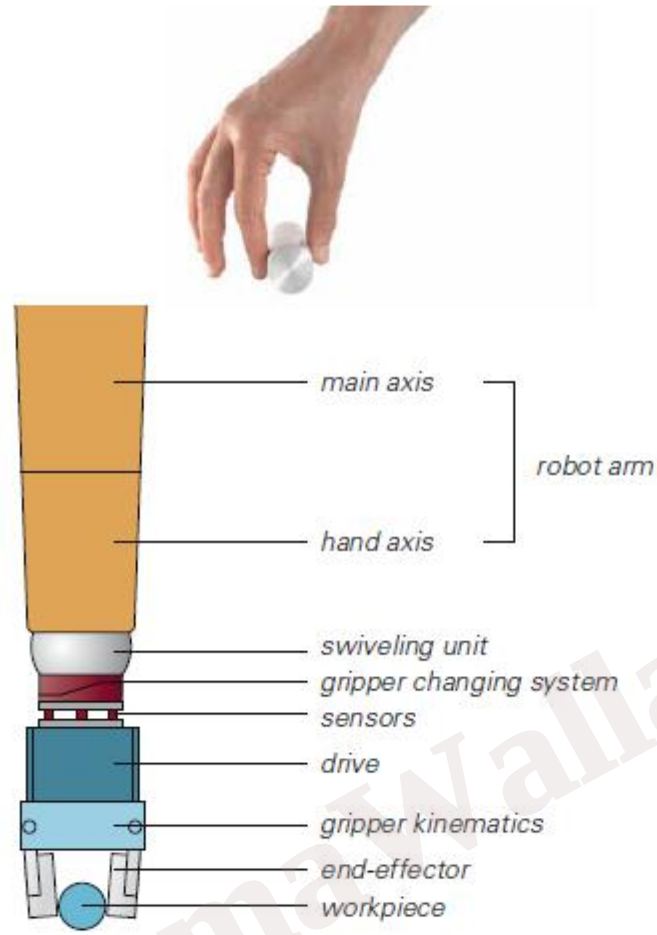


Figure 3.64: Human Grasping vs Robotic Grasping

3.4.1.1 Grasping Modes, Forces, and Stability

i. Grasping Stability

Robotic grasping has been the topic of numerous research efforts. This is best summarized by Bicchi and Kumar. Stable object grasping is the primary goal of gripper design. The most secure grasp is to enclose the gripper jaws or finger around the center of gravity of the object. Six basic grasping patterns for human hands have been identified by Taylor and Schwarz in the study of artificial limbs. Six grasping forms of (1) spherical, (2) cylindrical, (3) hook, (4) lateral, (5) palmar, and (6) tip are identified in Figure 3.65. For conventional two-dimensional grippers described in the previous sections, only the cylindrical, hook, and lateral grips apply. To securely grasp an object the cylindrical and lateral grips are effective for plane motion grippers. The two-jaw mechanisms of most end effectors/grippers most closely approximate the cylindrical grasp. Kaneko [9] also discusses stability of grasps for articulating multi-fingered robots. Envelop or spherical grasping is the most robust, as it has a greater number of contact points for an object that is enclosed by articulating fingers.

For plane-motion grippers using a cylindrical style of grasp, stability can be similarly defined. Stability increases as the object has more points of contact on the gripper jaws and when the object's center of gravity is most closely centered within the grasp. When using a cylindrical grip, it is desirable to grasp an object in a safe, non-slip manner if possible. For example, when grasping a cylindrical object with a flange, the jaws of the gripper should be just below the flange. This allows the flange to contact the top of the gripper jaws, to remove any possibility of slipping. If a vertical contact feature is not available, then a frictional grip must be used.

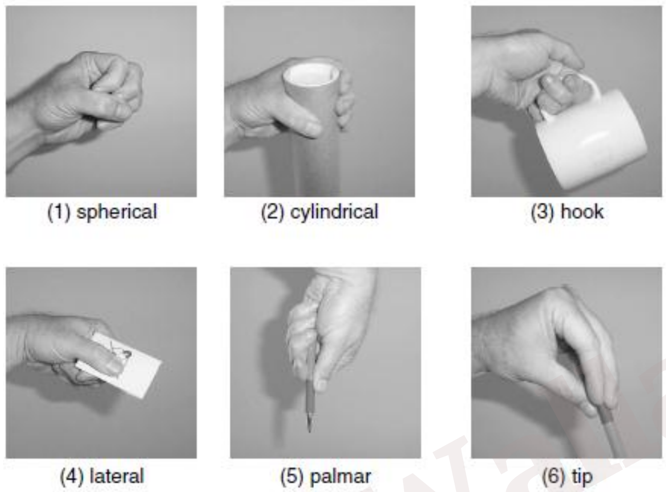


Figure 3.65: Grasp types for human hands

ii. Friction and Grasping Forces

While a slip-proof grasp previously described is preferred for end effectors, in the majority of cases a frictional grip is all that can be attained. When designed properly, the frictional grip is very successful. Friction forces can be visualized on a gripper jaw in Figure 3.66. It is desirable to have the center of gravity for grasped objects and end effector coincident in the *Z*-direction, to not have moments at the jaw surfaces. To successfully grasp an object the applied gripping frictional force on each jaw of a two-jaw gripper must be equal to or greater than the half the vertical weight and acceleration payload. From this the applied gripping normal force is found by dividing the required friction force by the static coefficient of friction. Typically a factor of safety is applied. Friction coefficients are a function of materials and surface geometries. Estimates can be found using standard references. Typically most surfaces will have a static coefficient of friction greater than 0.2. For metal to metal contacts, the static coefficient of friction is much higher (e.g., aluminum to mild steel has 0.6 and mild steel to hard steel has 0.78). In addition to an object's weight, surface texture, rigidity, and potential damage must also be considered in the selection or design of an end effector or the gripper. Pads are used on the jaws of the end effector to prevent surface damage to the object. Pads can also be used to increase the coefficient of friction between the object and the gripper jaws.

$$F_{friction} = \mu F_{grip} \geq \frac{(1 + A_{vert}/g)w}{2}$$

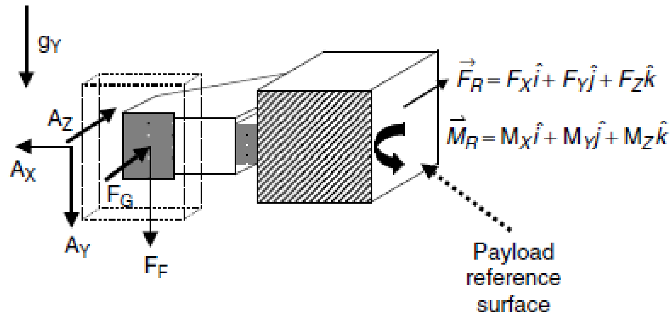


Figure 3.66: Gripper forces and moments

3.4.1.2 Types of End Effector

There are 2 types of end effector:

i. **Gripper** – use to grip, hold and move an object to another location.
ii. **Tools** @ *devices and end-of-arm tooling* @ equipment @ hardware
– use to do work such as drilling, grinding, spraying paint or coating.

3.4.1.3 Factors in Selecting End Effector

i. The position of the component to be held must be reached.
ii. Changes in the size of the components must be dealt with where it will affect the accuracy to put the components. The radius of the grip must be designed to follow the shape of the component.
iii. Appropriate power source is required to move the grip whether mechanical, electrical or pneumatic.
iv. Temperature and humidity working environment.
v. Safety during operation.
vi. Problems and possible crooked and distortion on the grip.
vii. The cost of construction or the purchase of grip.

3.4.1.4 Gripping Ability

i. Grip the object to be gripped.
ii. Adjustments can keep the object during the transportation stability.
iii. The effect of the object's position compared to the grip.

3.4.1.5 Types of Robotic Grippers

i. Impactive – jaws or claws which physically grasp by direct impact upon the object.
ii. Ingressive – pins, needles or hackles which physically penetrate the surface of the object (used in textile, carbon and glass fiber handling).
iii. Astrictive – suction forces (includes magnetic) applied to the objects surface.
iv. Contigutive – requiring direct contact for adhesion to take place (such as glue, surface tension or freezing)

Figure 3.67: Vacuum lifters, vacuum lifts & lifting devices

Additional characteristics such as low weight, fast reaction, and being fail-safe are required. There are different techniques for gripping parts. Grippers are typically classified into 5 groups by performing principle:

1. mechanical grippers
2. vacuum grippers
3. magnetic devices
4. flexible pneumatic devices
5. special purpose tools and special purpose devices

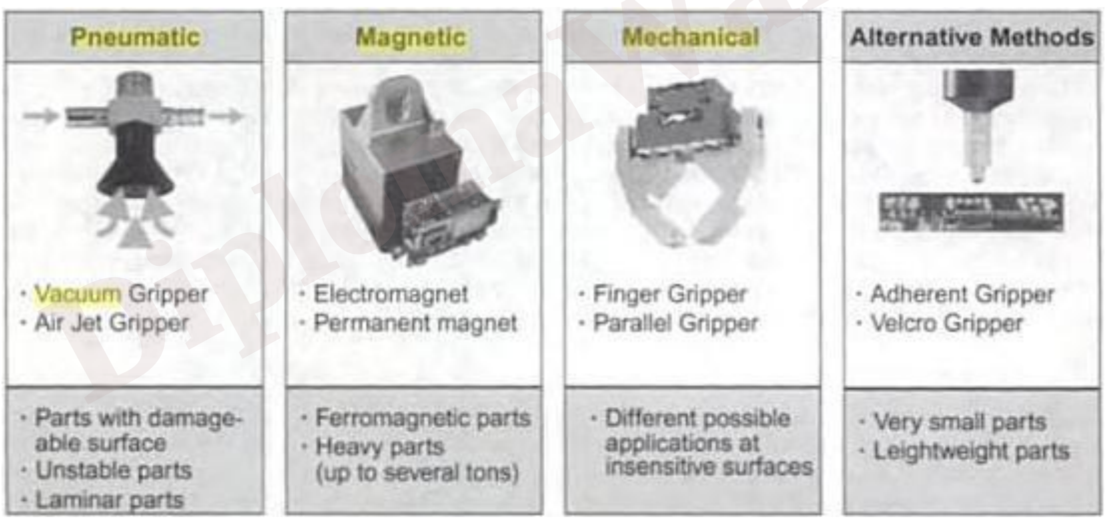


Figure 3.68: Different kinds of grippers

With a market share of 66%, mechanical grippers are the most commonly used. Mechanical grippers can be subdivided by their kind of closing motion and their number of fingers.

i. ***Three finger centric gripper*** are typically used for gripping round and spherical parts.
ii. ***Two finger parallel grippers*** perform the gripping movement with a parallel motion of their fingers, guaranteeing secure gripping because only forces parallel to the gripping motion occur.

Because these gripper models can cause harm to the component surface, **vacuum grippers** are used for handling damageable parts. Two dimensional parts such as sheet metal parts are also handled by vacuum grippers. Using the principle of the Venturi nozzle, an air jet builds up a vacuum in the suction cup that holds the parts. When the air jet is turned off, the parts are automatically released. Heavy parts such as shafts are lifted not by mechanical grippers but with electromagnetic grippers. However, secure handling, not exact positioning, is needed when using these grippers. To handle small and light parts, grippers using alternative physical principles, such as electrostatic and adherent grippers, are used because they do not exert any pressure that could cause damage to the part. Fields of application include micro assembly and electronics production.

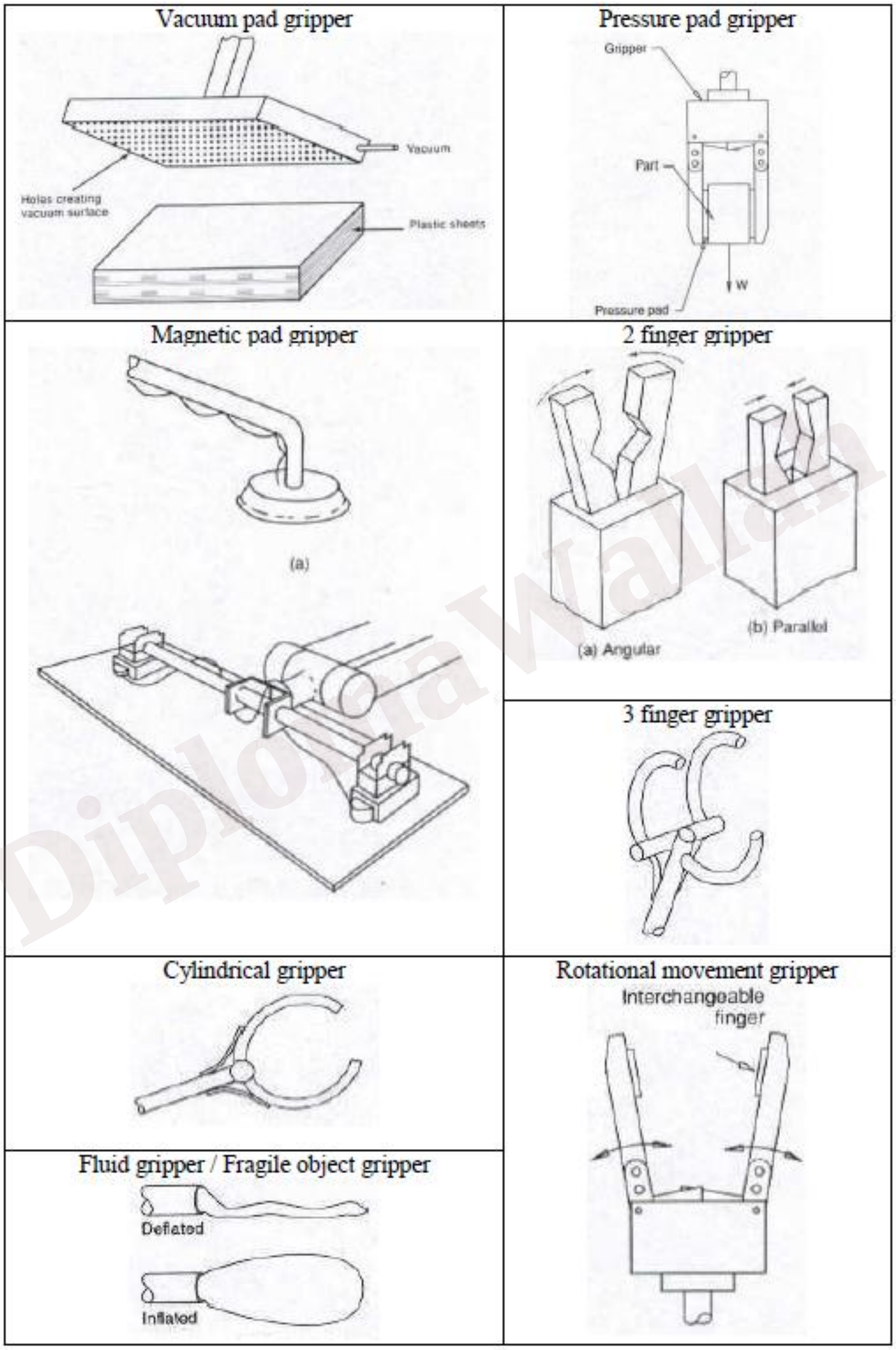


Figure 3.69: Types of Grippers

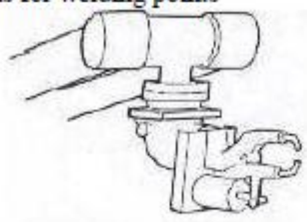
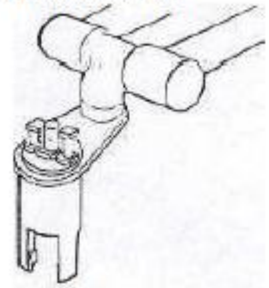
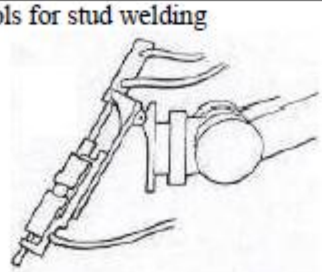
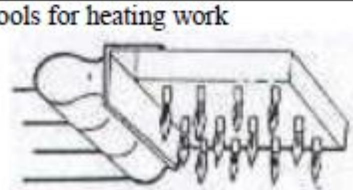
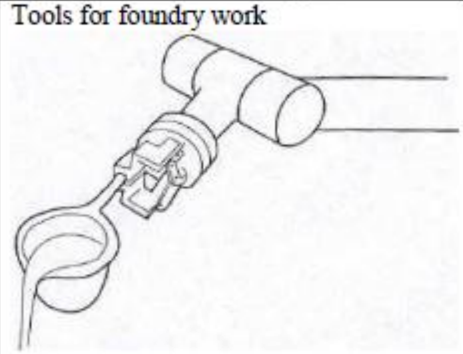
Type of end-of-arm tooling			
1.	Tools for welding points 	2.	Tools for installation nut 
3.	Tools for stud welding 	4.	Tools for arc welding 
5.	Tools for heating work 	6.	Tools for grinding work 
7.	Tools for spray paint or coating 	8.	Tools for foundry work 

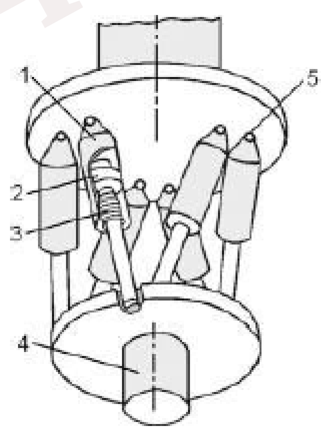
Figure 3.70: Special Purpose tools and Special purpose devices

3.4.2 Remote Centered Compliance Devices

The precise mating of tightly fitting parts often necessitates the physical modification of the components in question. Rather than making the robot more accurate, a reduction in the overall accuracy requirement often provides a more cost effective solution. Countersunk holes and chamfered shaft tips are good examples of "design for automation". A further requirement is that the gripper must be capable of accommodating these deliberately introduced tolerances. Consequently, the use of controlled, or at least predictable, compliance can be most helpful. One device which offers this possibility, deliberately introduced to robotics to avoid the need for extensive sensing and data processing, is the passive compliant mechanism known as the Remote Centre Compliance, or RCC . Nevertheless, the addition of instrumentation, such as force sensing, can also help facilitate error recovery strategies helpful in detecting mismatch and preventing jamming.

Figure 3.71 shows an early solution, developed at the Draper Laboratories, for an RCC which exhibits 6 degrees of freedom. It is suitable for the insertion of short bolts (diameter 12 to 58 mm, length 25 to 100 mm) where mechanical play may range from 12 to 24. It has been demonstrated that it is possible to compensate for lateral errors as large as 2 mm and angle errors along the axes of insertion as large as 2.5°. The RCC is constructed in such a way that there are separate structural elements for the position and angle deviations. In the case of angle error compensation the assembly part moves about a virtual center of rotation which is located in the middle of the lower front surface of the workpiece. Angle error compensation and position error compensation are decoupled and take place independently of one another.

Figure 3.72 shows the principle and function of each component in a simplified RCC. The compliance function is a typical mechanical second order model comprising a mass (gripper and workpiece), a spring force and a damper as illustrated in Figure 3.71. The springs may be of metal or elastomer construction. Elastomer links are designed as rubber-metal composites with a definite axial hardness and shear resistance. Many design examples for RCC units have been published in the literature and extensive information can be found elsewhere. The orientation motions enforced by the force field resulting from contact between the robot and gripper flanges are directly related to forces exerted vertically.



1. Damper cylinder
2. Damper piston
3. Return spring
4. Gripper flange
5. Ball joints

Fig. 3.71: Passively operating compensation system from *McCallion (1980)*

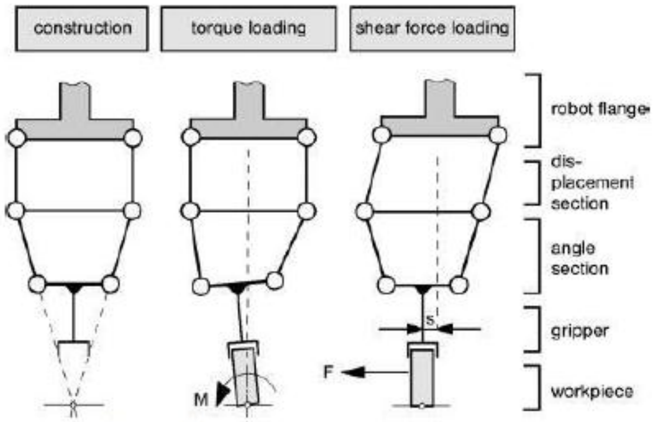


Figure 3.72: Principle and function of an RCC ***M*** torque, *s* displacement path, ***F*** force.

Figure 3.73 shows the effect of placing a peg directly into a chamfered hole. The peg initially makes a single point contact and the RCC takes up any horizontal movement as the peg is pushed vertically downwards into the hole. Figure 3.73b illustrates the effect of tilting the workpiece (or the gripper) to achieve a two point contact between the peg and the hole. The RCC in this case makes a slight angular displacement possible. For an extremely readable analysis of this, and related situations, the work by McKerrow is strongly recommended.

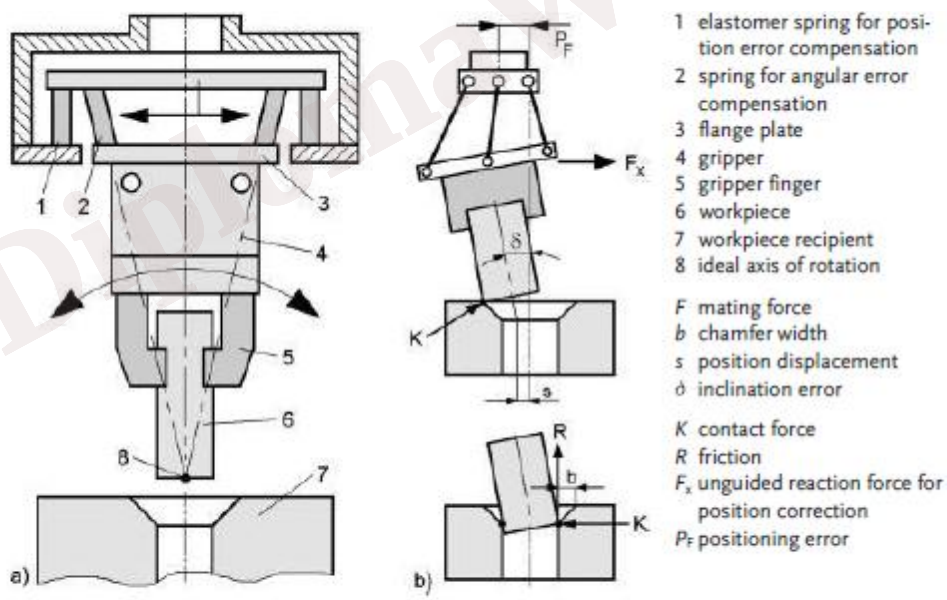


Figure 3.73 Operational principle of an RCC (a) task, (b) principle of passive correction

Compensation motions proceed in two phases:

- Compensation of the position deviation (s) during the sliding of the workpiece along the chamfer of the workpiece recipient (single point contact).
- Compensation of the position deviation (s) during insertion of the workpiece into the workpiece recipient (two point contact) through rotation about the ideal axis (8).

Mating mechanisms are specified according to the following important parameters:

- Overall size and mass.
- Maximum position and angle deviations between the assembly parts.

- Maximum applicable mating force.
- Overload protection.
- Complexity of fabrication and respective costs.

The art of designing RCC links consists in locating the apparent center of rotation at the end of the workpiece. This leads to an effective pulling of the workpiece into the hole. In alternative designs, the pneumatic damper may also be used as an additional actuator as shown schematically in Figure 3.74. Though in practical applications motion is the role of the robot, motion in the Z-axis (direction of insertion) may be augmented by a pneumatic cylinder.

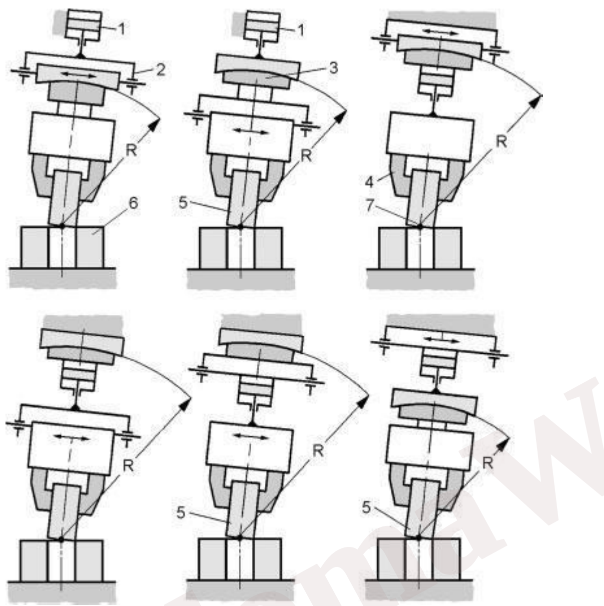


Figure 3.74: Operation of RCC employing two shear and one swivel joints

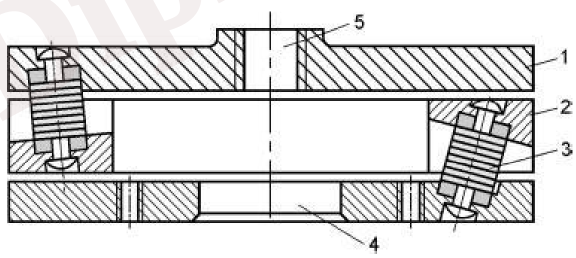


Figure 3.75: Design of a standard flat RCC mechanism

It is essential for the RCC link to be designed as flat as possible in order to save space, as shown in the example in Figure 3.75. The use of metal springs allows the transmission of larger forces in the Z-axis than is possible with elastomer springs.

The special design shown in Figure 3.76 uses two cones, one inside the other. Initially, the workpiece makes contact with the chamfer. However, since the joining head continues its downward motion, the inner cone which is subjected to spring retaining forces is allowed to move. This introduces a small degree of mechanical play allowing the inner cone to make contact with one of the inner surfaces of the hollow outer cone. This compensates for axial deviations X making mating of workpiece and recipient possible.

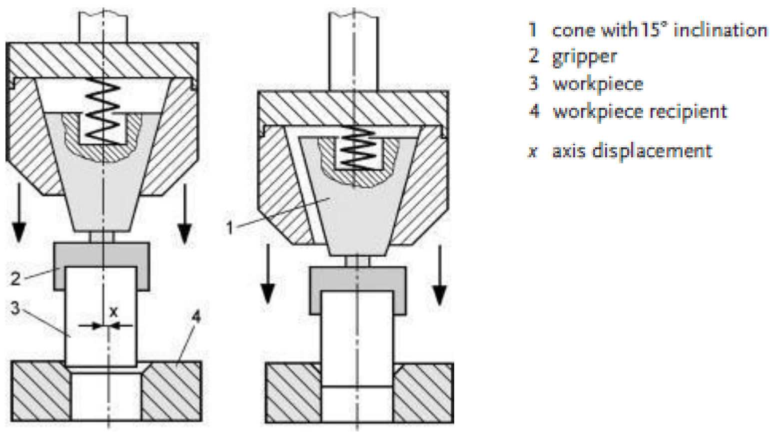


Figure 3.76: Double cone RCC

The designs of RCC hitherto considered are simple, light and effective as long as positional and angular deviations are not too great. However, unexpected forces such as jamming due to burrs in a hole are often beyond the abilities of an RCC. An extension to this simple passive compliance is the Instrumented RCC or IRCC.

3.4.2.1 Instrumented Remote Centre Compliance (IRCC)

Unlike the purely passive RCC systems, IRCC links actively compensate for angle and lateral displacements. The compensation error is measured, for example using a PSD, and corrections are performed actively. Figure3.77 shows one solution where such a device is inserted between the gripper and robot arm. Rough alignment is realized actively while fine positioning is ensured passively by means of the usual elastic elements. The PSD used in Figure 3.77 is a four quadrant photodetector. Gripper dislocations lead to displacement of the LED and hence to differential photocurrents between PSD segments. These differences are measured and the results used for robotic position correction.

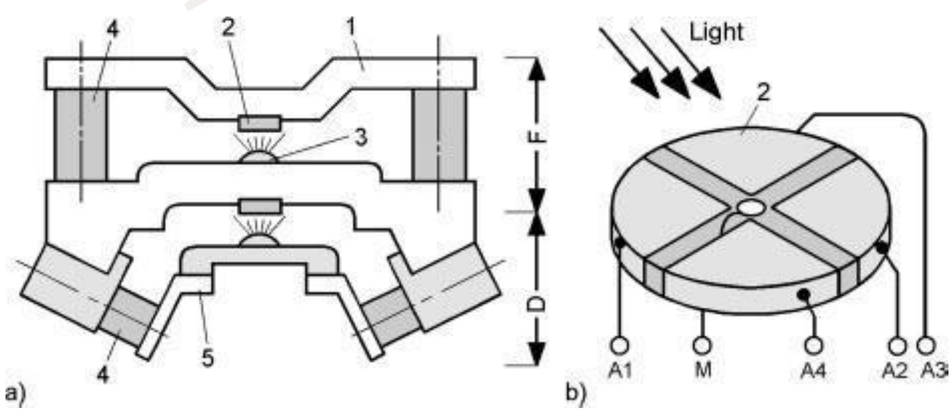


Fig. 3.77: Principles of an IRCC link (a) optical sensor compensation system, (b) PSD (position sensitive detector), 1 robot flange, 2 PSD, 3 LED, 4 elastometric element, 5 gripper flange
A photodetector, D moment compensation, F force compensation, M mass connection

The compensation mechanism schematically shown in Figure 3.78 contains leaf springs as elastic elements. Deformation of these springs can be detected using strain gauges and the information used for position correction of the gripper.

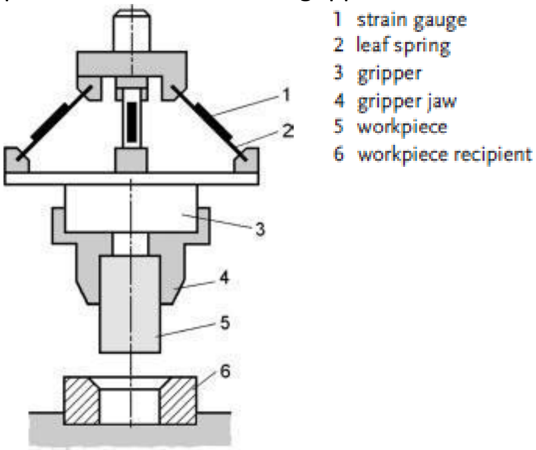


Figure 3.78: IRCC system using force measurement

Figure 3.79 shows a hybrid mating mechanism consisting of both passive and active systems. The actively controlled part operates pneumatically in the x and y directions. Two dynamic pressure nozzles (sensors) are used to measure distances from the base flange. The pneumatic control equipment, which is directly coupled with the sensors, consists of air suspended displaceable plates. The displacement continues until all sensor orifices attain the same position relative to the base hole sides and the measured pressure differences in the control chambers vanish. This signals the conclusion of the mating process.

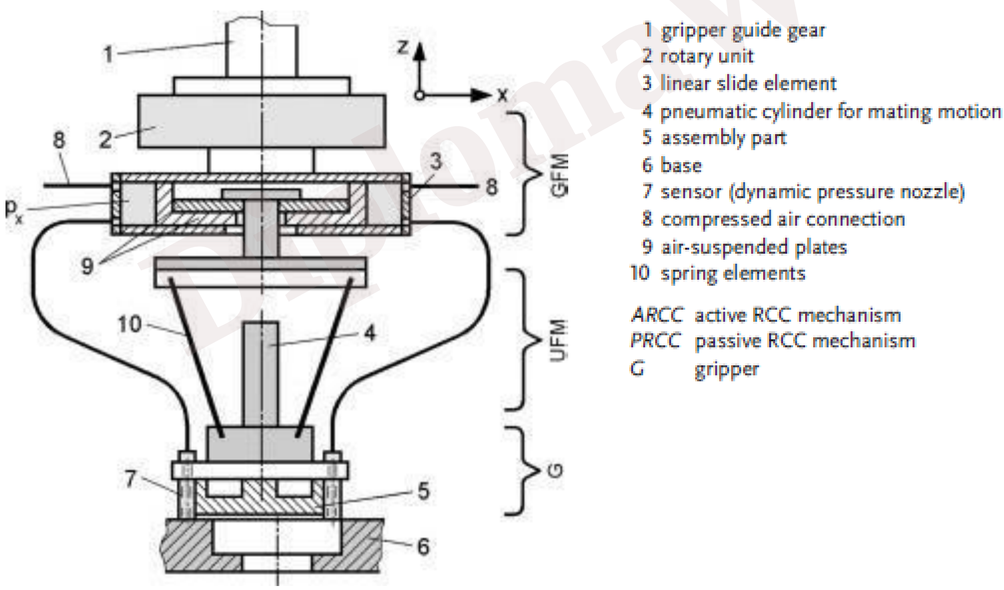
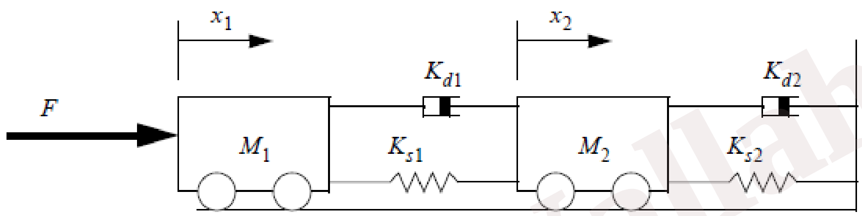


Fig. 3.79: Combined joining mechanism with pneumatic sensors and steel spring elements

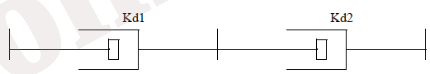
END OF CHAPTER 3

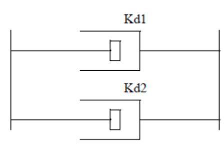
1. Differentiate between *Machines and Mechanism.*
2. Name two types of links.
3. Describe three types of mechanical motion.
4. What are the different between translation motion and rotational motion?
5. Discuss the concept of translation motion and an example of rotational motion.
6. Discuss the concept of rotational motion and an example of rotational motion
7. Name and draw three mechanical components that transform rotary to rotary motion.
8. Name and draw three mechanical components that transform rotary to translation motion.
9. What is the primary function of linkage?
10. Name five types of linkage.
11. Describe and sketches the Specialized Linkages.
12. What is the function of couplers?
13. Name two types of couplers.
14. Named two types of flexible coupling.
15. Discuss the mechanical flexible coupling.
16. What is the meaning of power transfer in mechanical concept?
17. Named five categories of power transfer mechanism.
18. Named three basic components in modeling of mechanical system.
19. Write the equation for torque function in modeling of a rotational spring.
20. Write the equation for torque function in modeling of a rotational damper.
21. Two springs are connected in series. One has a k of 34 and the other a k of 3.4. Calculate their effective spring constant. Which spring dominates? Repeat with the two springs in parallel. Which spring dominates? (Use any unit system)
22. Repeat question 2 with k_1 = 125 and k_2 = 25 (Use any unit system)
23. Repeat question 2 with k_1 = 125 and k_2 = 115 (Use any unit system)
24. Two dampers are connected in series. One has a damping factor c_1 =12.5 and the other, c_2 = 1.2. Calculate their effective damping constant. Which damper dominates? Repeat with the two dampers in parallel. Which damper dominates? (use any unit system).
25. Repeat question 5 with c_1 = 12.5 and c_2 = 2.5 (use any unit system)
26. Repeat question 5 with c_2 = 12.5 and c_2 = 10 (use any unit system)

27. A mass of m=2.5 and a spring with k = 42 are attached to one end of a lever at a radius of 4. Calculate the effective mass and effective spring constant at a radius of 12 on the same lever. (use any unit system)
28. A mass of m=1.5 and a spring with k = 24 are attached to one end of a lever at a radius of 3. Calculate the effective mass and effective spring constant at a radius of 10 on the same lever. (use any unit system)
29. A mass of m=4.5 and a spring with k = 15 are attached to one end of a lever at a radius of 12. Calculate the effective mass and effective spring constant at a radius of 3 on the same lever. (use any unit system)
30. Develop the equation relating the input force to the motion (in terms of x) of the left hand cart for the problem below.

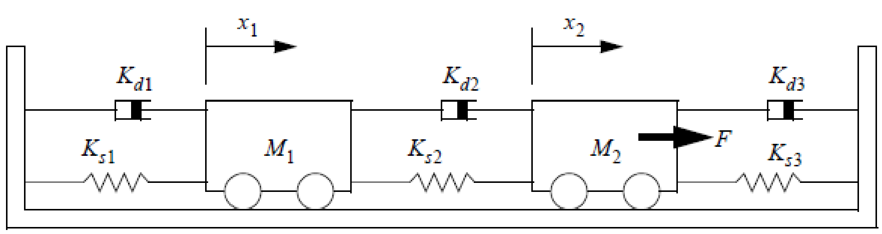


31. Find the effective coefficients for the pairs below

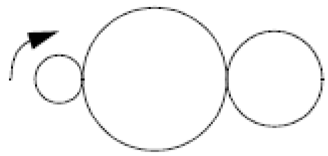




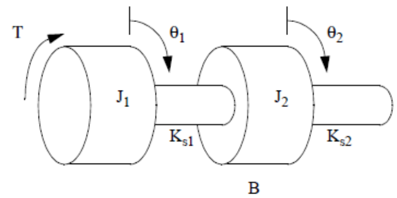
32. Write the differential equations for the system below:



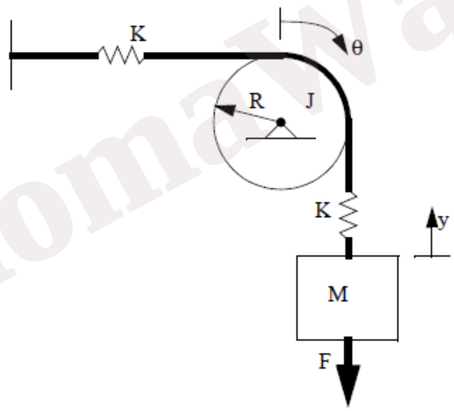
33. A gears train has an input gear with 20 teeth, a center gear that has 100 teeth, and an output gear that has 40 teeth. If the input shaft is rotating at 5 rad/sec what is the rotation speed of the output shaft?



34. Draw the FBDs and write the differential equations for the mechanism below. The right most shaft is fixed in a wall.



35. Write the differential equations for the mechanical system below, and then find the input-output equations. The input is force 'F', and the outputs are 'y' and the angle theta. Include the inertia of both masses, and gravity for mass 'M'.



36. What are Electrical Analogies?
37. Define the End Effectors
38. Grippers are not only required use with industrial robot; They are a universal component in automation system. State five applications in automation that needs a gripper.
39. Give FIVE (5) factors in selecting end effector
40. Explain the factor to be considered in Grasping problem.
41. Name SIX (6) types of grasping form.
42. Give THREE (3) types of robot end effector driven actuators.
43. Tooling's are fixed the end of the robots arm. Name five (5) types of tooling that fixed at the robot arm.
44. Explain the Remote Centered Compliance Devices

45. Named six types of forms of object prehension in robots.
46. Draw and named subsystem of a mechanical gripper.
47. What are Remote Centered Compliance Devices?
48. Named the two phases in compensation of motions?
49. A mechanism is used to open the door of a heat treating furnace and is shown in Figure 1. Draw a kinematic diagram and find their total degrees of freedom of the mechanism. The end of the handle should be identified as a point of interest.

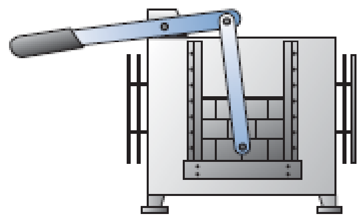


Figure 1

50. A pair of bolt cutters is shown in Figure 2. Draw a kinematic diagram and find their total degrees of freedom of the mechanism, selecting the lower handle as the frame. The end of the upper handle and the cutting surface of the jaws should be identified as points of interest.

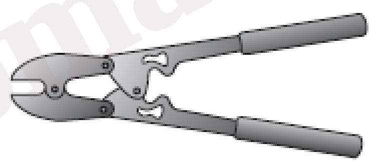


Figure 2

51. A folding chair that is commonly used in stadiums is shown in Figure 3. Draw a kinematic diagram and find their total degrees of freedom of the folding mechanism.

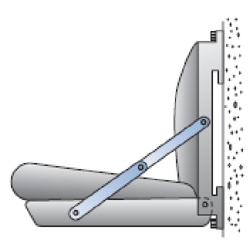


Figure 3

52. A foot pump that can be used to fill bike tires, toys, and so forth is shown in Figure 4. Draw a kinematic diagram and find their total degrees of freedom of the pump mechanism. The foot pad should be identified as a point of interest.

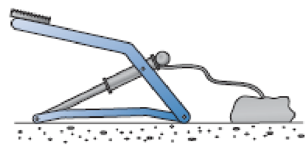


Figure 4

53. Another configuration for a pair of pliers is shown in Figure 5. Draw a kinematic diagram and find their total degrees of freedom of the mechanism.

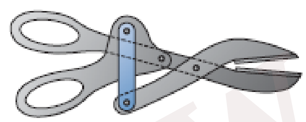


Figure 5

54. A sketch of a device to move packages from an assembly bench to a conveyor is shown in Figure 6. Draw a kinematic diagram and find their total degrees of freedom of the mechanism.

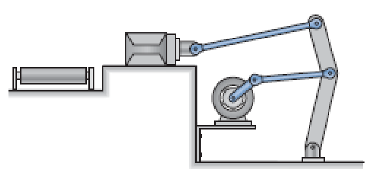


Figure 6

55. A sketch of a lift platform is shown in Figure 7. Draw a kinematic diagram and find their total degrees of freedom of the mechanism.

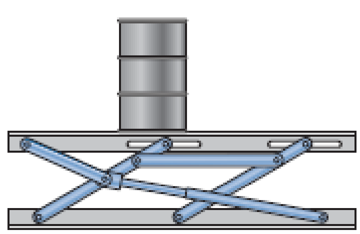


Figure 7

56. A sketch of a backhoe is shown in Figure 8. Draw a kinematic diagram and find their total degrees of freedom of the mechanism.

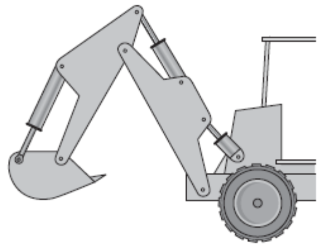


Figure 8

57. A sketch of a front loader is shown in Figure 9. Draw a kinematic diagram and find their total degrees of freedom of the mechanism.

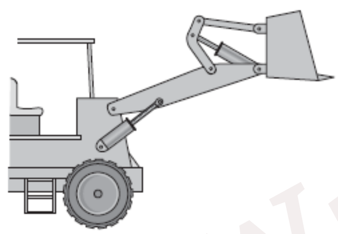


Figure 9

58. A sketch of an adjustable-height platform used to load and unload freight trucks is shown in Figure 10. Draw a kinematic diagram and find their total degrees of freedom of the mechanism.

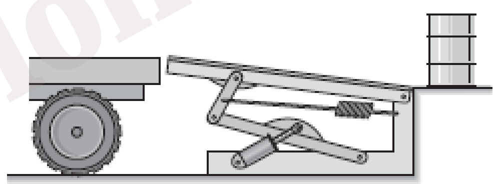


Figure 10

59. Problems in Classifying Four-Bar Mechanism.

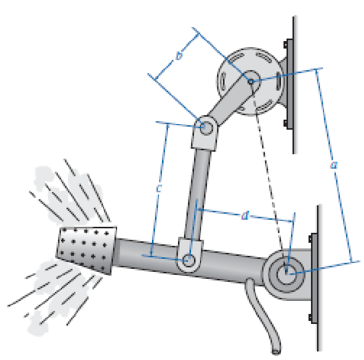


Figure 11

i. A mechanism to spray water onto vehicles at an automated car wash is shown in Figure 11. Classify the four-bar mechanism, based on its possible motion, when the lengths of the links are a = 12in, b =1.5in, c = 14in and d = 4in.

ii. For the water spray mechanism in Figure 11, classify the four-bar mechanism, based on its possible motion, when the lengths of the links are a = 12in, b = 5in, c = 12in and d = 4in.

iii. For the water spray mechanism in Figure 11, classify the four-bar mechanism, based on its possible motion, when the lengths of the links are a = 12in, b = 3in, c = 8in and d = 4in.

iv. For the water spray mechanism in Figure 11, classify the four-bar mechanism, based on its possible motion, when the lengths of the links are a = 12in, b = 3in, c = 12in and d = 5in.

60. CASE STUDY 1: The mechanism shown in Figure C1.1 has been taken from a feed device for an automated ball bearing assembly machine. An electric motor is attached to link A as shown. Carefully examine the configuration of the components in the mechanism. Then answer the following leading questions to gain insight into the operation of the mechanism.

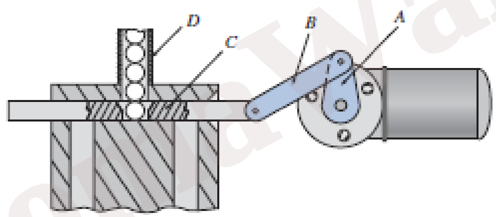


Figure C1.1

i. As link A rotates clockwise 90°, what will happen to slide *C*?
ii. What happens to the ball trapped in slide *C* when it is at this position?
iii. As link *A* continues another 90° clockwise, what action occurs?
iv. What is the purpose of this device?
v. Why are there chamfers at the entry of slide *C*?
vi. Why do you suppose there is a need for such a device?

61. CASE STUDY 2: Figure C1.2 shows a mechanism that is typical in the tank of a water closet. Note that flapper *C* is hollow and filled with trapped air. Carefully examine the configuration of the components in the mechanism. Then answer the following leading questions to gain insight into the operation of the mechanism.

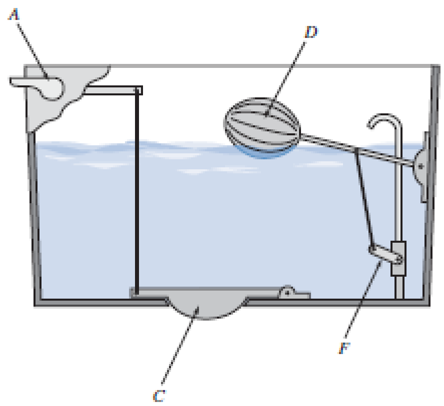


Figure C1.2

i. As the handle A is rotated counterclockwise, what is the motion of flapper *C*?
ii. When flapper *C* is raised, what effect is seen?
iii. When flapper *C* is lifted, it tends to remain in an upward position for a period of time. What causes this tendency to keep the flapper lifted?
iv. When will this tendency (to keep flapper *C* lifted) cease?
v. What effect will cause item *D* to move?
vi. As item *D* is moved in a counterclockwise direction, what happens to item *F*?
vii. What does item *F* control?
viii. What is the overall operation of these mechanisms?
ix. Why is there a need for this mechanism and a need to store water in this tank?

62. CASE STUDY 3: Figure C1.3 shows a mechanism that guides newly formed steel rods to a device that rolls them into reels. The rods are hot when formed, and water is used to assist in the cooling process. The rods can be up to several thousand feet long and slide at rates up to 25 miles per hour through channel *S*.

Once the reel is full, the reel with the coiled rod is then removed. In order to obtain high efficiency, the rods follow one another very closely. It is impossible to remove the reel in a short time interval; therefore, it is desirable to use two reels in alternation. This mechanism has been designed to feed the rods to the reels.

Buckets *B*1 and *B*2 have holes in the bottom. The water flow from the supply is greater than the amount that can escape from the holes. Carefully examine the configuration of the components in the mechanism, and then answer the following leading questions to gain insight into the operation of the mechanism.

i. In the shown configuration, what is happening to the level of water in bucket *B*1?
ii. In the shown configuration, what is happening to the level of water in bucket *B*2?
iii. What would happen to rocker arm *C* if bucket *B*2 were forced upward?
iv. What would happen to rocker arm *R* if bucket *B*2 were forced upward?
v. What does rocker arm *R* control?
vi. What is the continual motion of this device?
vii. How does this device allow two separate reels to be used for the situation described?
viii. Why do you suppose that water is used as the power source for the operation of this mechanism?

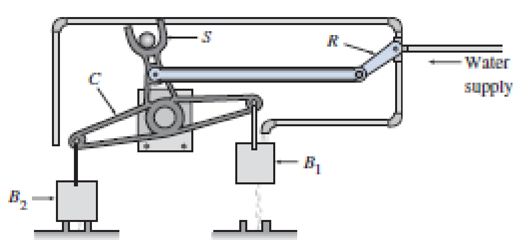
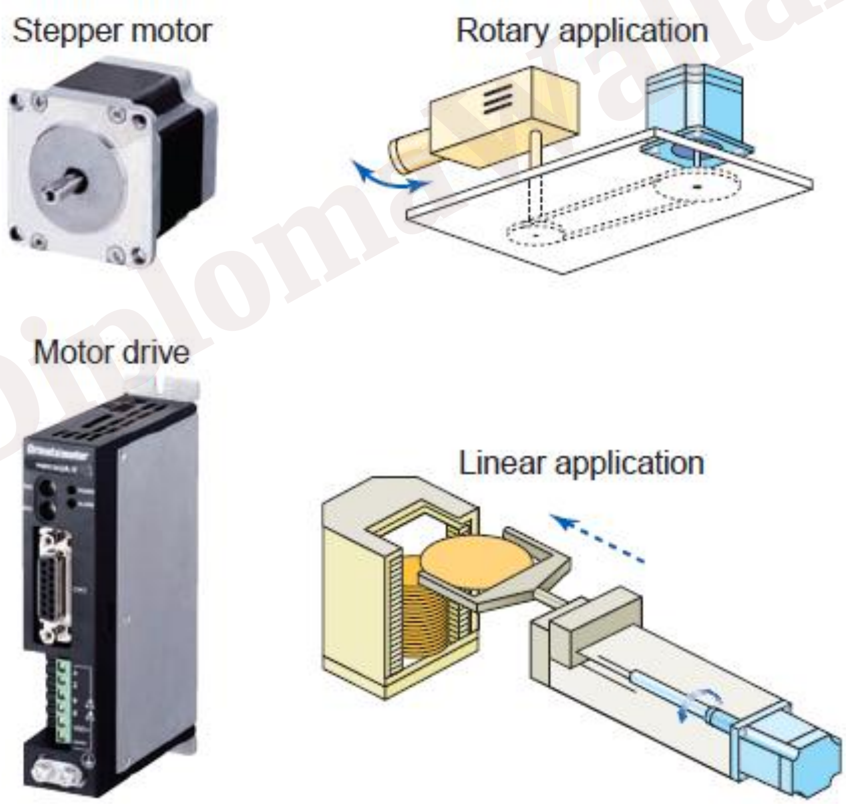


Figure C1.3

CHAPTER 4
CONTROL OF ACTUATORS IN AUTOMATION MECHANISMS

Upon completion of this course, students should be able to:-

- Identify Stepper Motors
- Apply control method of actuators

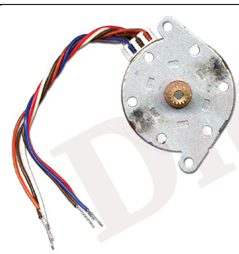


4.1 Identify Stepper Motors

If you are staring at a pile of stepper motors in a surplus shop, or have pulled one out of used equipment, here's how you can determine what you have. First, check for the number of wires coming out. If 5 or 6 or 8, that's good because you have a unipolar stepper. If 4, that's bad because you have a bipolar stepper and should put it back. If 2, you have a regular DC motor. Confirm you have a stepper motor by turning the shaft. You should feel the little detents indicating each step. Next, read the label on the side. If you are lucky, it will have the voltage and step size printed, or will be in a bin with the voltage marked. Look for 12V steppers. If you have a 5V stepper, and it is large, the currents will probably be too large for easy control. Small 5V steppers are OK. If you have no way of telling the voltage, it is probably best to look for another stepper.

Next, get out your digital ohmmeter and start reading resistances between the leads. You will get different values depending on which pair of leads you measure. The lowest resistance you find is the coil resistance. Use I=V/R to compute the coil current. If 250 mA or less, you are in good shape. Look at the output shaft and determine if it is something you can handle. Common steppers have plain shafts with 0.125, 0.196 or 0.250 diameter. Gears press fit onto the shaft may be useful to you or can be removed.

Consider the size and weight of the stepper. Very large or very heavy steppers will most likely require more current than you can control. Many steppers come in standard NEMA (National Electrical Manufacturer's Association) sizes. NEMA size 14, 15, or 16 are typically cubic in shape with the front mounting flange 1.38 to 1.65 inches on a size, and are great for robotics. NEMA sizes 23 are cylindrical with a square mounting flange 2.22 inches on a side. Size 23 motors may require too much current so check the specs carefully. Another common shape is stacked cans with a diamond-shaped mounting flange. Smaller sizes are also good for robotics.



Parallax PN 27964, 12V 90 ohm, 3.75 deg/step

NEMA size 17 stepper

NEMA size 23 stepper

Small, stacked can, diagonal flange stepper

NEMA size 15 stepper

Another stepper

Figure 4.1: Stepper motor with different shape

Stepper motors operate differently than standard types, which rotate continuously when voltage is applied to their terminals. The shaft of a stepper motor rotates in discrete increments when electrical command pulses are applied to it in the proper sequence. Every revolution is divided into a number of steps, and the motor must be sent a voltage pulse for each step. The amount of rotation is directly proportional to the number of pulses and the speed of rotation is relative to the frequency of those pulses. A 1-degree-per-step motor will require 360 pulses to move through one revolution; the degrees per step are known as the *resolution.* When stopped, a stepper motor inherently holds its position. Stepper systems are used most often in “open-loop” control systems, where the controller tells the motor only how many steps to move and how fast to move, but does not have any way of knowing what position the motor is at.

The movement created by each pulse is precise and repeatable, which is why stepper motors are so effective for load-positioning applications. Conversion of rotary to linear motion inside a linear actuator is accomplished through a threaded nut and lead screw. Generally, stepper motors produce less than 1hp and are therefore frequently used in low-power position control applications. The three basic kinds of stepper motors are *permanent magnet, variable reluctance,* and *hybrid.* The same controller circuit can drive both hybrid and PM stepping motors.

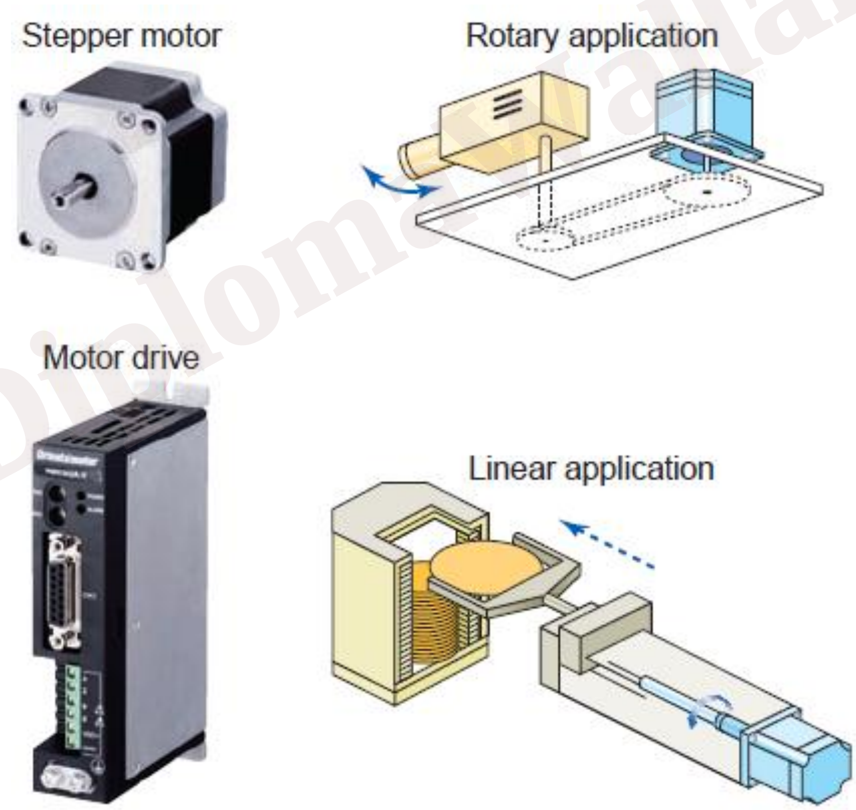


Figure 4.2: Shows a stepper motor/drive unit along with typical rotary and linear applications.

4.1.1 Principles of stepper motor operation

The essential property of the stepping motor is its ability to translate switched excitation changes into precisely defined increments of rotor position ('steps'). Stepping motors are categorized as doubly salient a machine, which means that they have teeth of magnetically permeable material on both the stationary part (the 'stator') and the rotating part (the 'rotor'). A cross-section of a small part of a stepping motor is shown schematically in Figure 4.3. Magnetic flux crosses the small air gap between teeth on the two parts of the motor. According to the type of motor, the source of flux may be a permanent-magnet or a current-carrying winding or a combination of the two.

However, the effect is the same: the teeth experience equal and opposite forces, which attempt to pull them together and minimize the air gap between them. As the diagram shows, the major component of these forces, the normal force (*n*), is attempting to close the air gap, but for electric motors the more useful force component is the smaller tangential force (*t*), which is attempting to move the teeth sideways with respect to each other. As soon as the flux passing between the teeth is removed, or diverted to other sets of teeth, the forces of attraction decrease to zero.

The following sections explain how this very simple principle is put to work in practical stepping motor devices. Most stepping motors can be identified as variations on the two basic types: variable-reluctance or hybrid. For the hybrid motor the main source of magnetic flux is a permanent magnet, and dc currents flowing in one or more windings direct the flux along alternative paths. There are two configurations for the variable-reluctance stepping motor, but in both cases the magnetic field is produced solely by the winding currents.

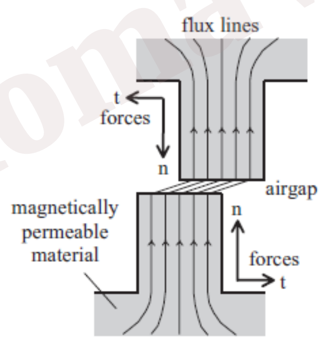


Figure 4.3: Force components between two magnetically permeable teeth

In the multi-stack variable-reluctance stepping motor the source of magnetic flux is current carrying windings placed on the stator teeth. These windings are excited in sequence to encourage alignment of successive sets of stator and rotor teeth, giving the motor its characteristic stepping action.

The multi-stack variable-reluctance stepping motor is divided along its axial length into magnetically isolated sections ('stacks'), each of which can be excited by a separate winding ('phase'). In the cutaway view of Figure 4.4, for example, the motor has three stacks and three phases, but motors with up to seven stacks and phases have been manufactured.

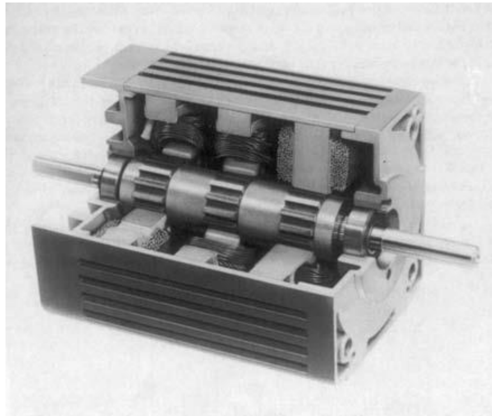


Figure 4.4: Cutaway view of a three-stack variable-reluctance stepping motor
(Photograph by Warner Electric Inc., USA)

Each stack includes a stator, held in position by the outer casing of the motor and carrying the motor windings, and a rotating element. The rotor elements are fabricated as a single unit, which is supported at each end of the machine by bearings and includes a projecting shaft for the connection of external loads, as shown in Fig. 4.5a. Both stator and rotor are constructed from electrical steel, which is usually laminated so that magnetic fields within the motor can change rapidly without causing excessive eddy current losses. The stator of each stack has a number of poles – Fig. 4.5b shows an example with four poles – and a part of the phase winding is wound around each pole to produce a radial magnetic field in the pole. Adjacent poles are wound in the opposite sense, so that the radial magnetic fields in adjacent poles are in opposite directions. The complete magnetic circuit for each stack 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 a closing section, called the 'back-iron'. This magnetic circuit is repeated for each pair of poles, and therefore in the example of Fig. 4.5b there are four main flux paths. The normal forces of attraction between the four sets of stator and rotor teeth cancel each other, so the resultant force between the rotor and stator arises only from the tangential forces.

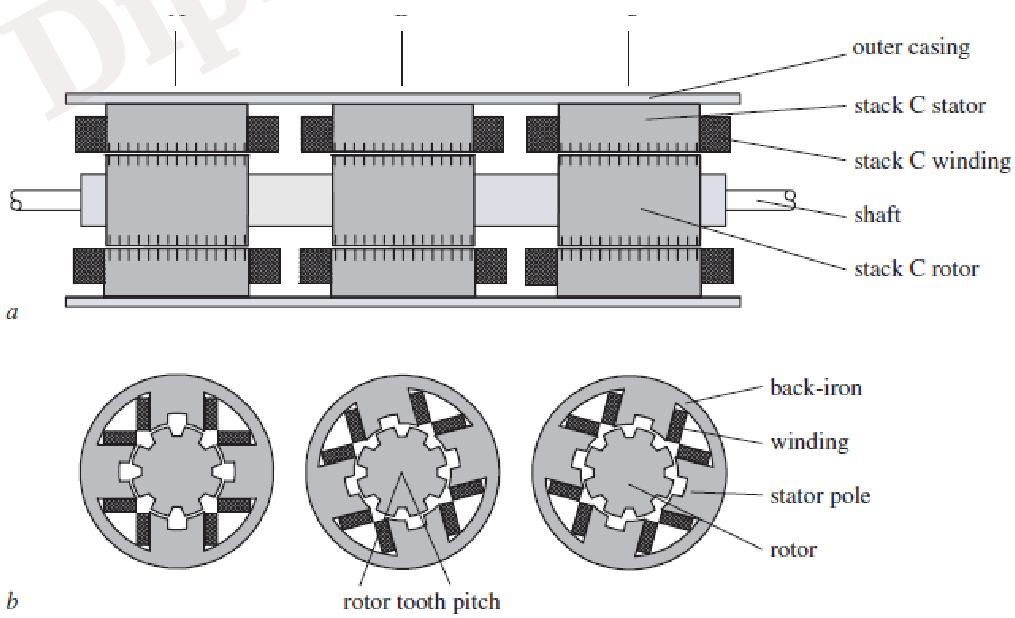


Figure 4.5 a. Cross-section of a three-stack variable-reluctance stepping motor parallel to the shaft, b. Cross-sections of a three-stack variable-reluctance stepping motor perpendicular to the shaft.

The position of the rotor relative to the stator in a particular stack is accurately defined whenever the phase winding is excited. Positional accuracy is achieved by means of the equal numbers of teeth on the stator and rotor, which tend to align so as to reduce the reluctance of the stack magnetic circuit. In the position where the stator and rotor teeth are fully aligned the circuit reluctance is minimized and the magnetic flux in the stack is at its maximum value.

The stepping motor shown in Fig. 4.5*b* has eight stator/rotor teeth and is in the position corresponding to excitation of stack *A*. Looking along the axial length of the motor the rotor teeth in each stack are aligned, whereas the stator teeth have different relative orientations between stacks, so in stacks *B* and *C* the stator and rotor teeth are partially misaligned. The effect of changing the excitation from stack *A* to stack *B* is to produce alignment of the stator and rotor teeth in stack *B*. This new alignment is made possible by a movement of the rotor in the clockwise direction; the motor moves one 'step' as a result of the excitation change. Another step in the clockwise direction can be produced by removing the excitation of stack *B* and exciting stack *C*. The final step of the sequence is to return the excitation to stack *A*. Again the stator and rotor teeth in stack *A* are fully aligned, except that the rotor has moved one rotor tooth pitch, which is the angle between adjacent rotor teeth defined in Fig. 4.5*b*. Therefore in this three-stack motor three changes of excitation cause a rotor movement of three steps or one rotor tooth pitch. Continuous clockwise rotation can be produced by repeating the excitation sequence: *A*, *B*, *C*, *A*, *B*, *C*, *A*, ... Alternatively anticlockwise rotation results from the sequence: *A*, *C*, *B*, *A*, *C*, *B*, *A*, ... If bidirectional operation is required from a multi-stack motor it must have at least three stacks so that two distinct excitation sequences are available.

There is a simple relationship between the numbers of stator/rotor teeth, number of stacks and the step length for a multi-stack variable-reluctance motor. If the motor has *N* stacks (and phases) the basic excitation sequence consists of each stack being excited in turn, producing a total rotor movement of *N* steps. The same stack is excited at the beginning and end of the sequence and if the stator and rotor teeth are aligned in this stack the rotor has moved one tooth pitch. Since one tooth pitch is equal to $(360/p)^\circ$, where *p* is the number of rotor teeth, the distance moved for one change of excitation is

$$\textbf{step length} = (360/Np)^\circ$$

The motor illustrated in Fig. 4.5 has three stacks and eight rotor teeth, so the step length is 15°. For the multi-stack variable-reluctance stepping motor typical step lengths are in the range 2–15°. Successful multi-stack designs are often produced with additional stacks, so that the user has a choice of step length; for example, a three-stack, 16 rotor tooth motor gives a step of 7.5° and by introducing an extra stack (together with reorientation of the other stacks) a 5.625° step is available. Although the use of higher stack numbers is a great convenience to the manufacturer, it must be remembered that more phase windings require more drive circuits, so the user has to pay a penalty in terms of drive circuit cost. Furthermore it can be shown that motors with higher stack numbers have no real performance advantages over a three-stack motor.

4.1.1.1 Aspects of Design

Each pole of the multi-stack stepping motor is provided with a winding which produces a radial magnetic field in the pole when excited by a dc current. The performance of the stepping motor depends on the strength of this magnetic field; a high value of flux leads to a high torque retaining the motor at its step position.

In the position where rotor and stator teeth are fully aligned, as in stack *A* of Fig. 4.5*b*, the reluctance of the main flux path is at its minimum value. For low values of current in the pole windings the flux density in the stator/rotor iron is small and the reluctance of these parts of the flux path is much less than the reluctance of the air gap between the stator and rotor teeth. As the winding current is increased, however, the flux density in the steel eventually reaches its saturation level. Further increases in winding current then produce a diminishing return in terms of improved flux level. Another limitation on pole field strength arises from the heating effect of the winding currents. The power dissipated in the windings is proportional to the square of the current, so the temperatures rise of the windings increases rapidly for higher currents. In most applications it is the ability of the winding insulation to withstand a given temperature rise which limits the current to what is termed its 'rated' value. For a well-designed variable-reluctance stepping motor the limitations on pole flux density and winding temperature rise are both effective (Harris *et al*., 1977): the stator/rotor iron reaches magnetic saturation at the rated winding current.

For the three-stack motor illustrated in Fig. 4.5 there are four poles, and hence four pole windings, per stack. Since all four windings in one stack must be excited concurrently it is common practice to interconnect the windings to form one phase. The three alternative methods of connecting four windings are shown in Fig. 4.6. Although the rated pole winding current depends only on the acceptable temperature rise, the corresponding rated phase current also depends on the interconnection, as shown in Table 4.1. The rated phase voltage is the voltage which must be applied at the phase terminals to circulate the rated current in the windings. For the series connection the phase current is low and the voltage high compared to the parallel connection, but there is no difference in the power supplied to the phase. Most manufacturers produce a given design of stepping motor with a range of winding interconnections, so the user can select a low-voltage, high-current drive with the parallel connection or a high-voltage, low-current drive with the series connection.

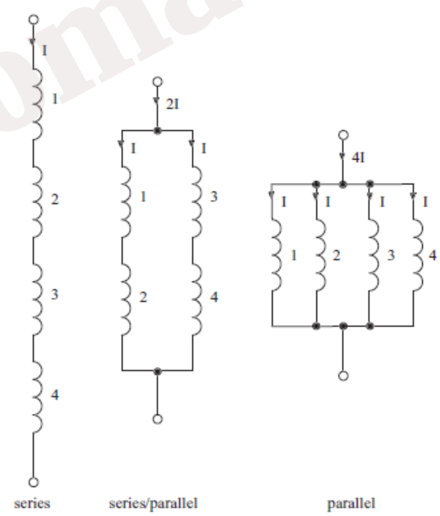


Figure 4.6: Interconnection of pole windings

Table 4.1: Effect of winding connection on ratings

Connection	Rated current	Resistance	Rated voltage	Power
Series	I	$4r$	$4rI$	$4rI^2$
Series/parallel	$2I$	r	$2rI$	$4rI^2$
Parallel	$4I$	$r/4$	rI	$4rI^2$

4.1.2 Half Step Mode Operation

We have already seen how to step the motor in 308 increments by energizing the phase's one at a time in the sequence ABCA, etc. Although this 'one-phase-on' mode is the simplest and most widely used, there are two other modes, which are also frequently employed. These are referred to as the 'two-phase-on' mode and the 'half-stepping' mode. The two phase- on can provide greater holding torque and a much better damped single-step response than the one-phase-on mode; and the half-stepping mode permits the effective step angle to be halved – thereby doubling the resolution – and produces a smoother shaft rotation.

In the two-phase-on mode, two phases are excited simultaneously. When phases A and B are energized, for example, the rotor experiences torques from both phases, and comes to rest at a point midway between the two adjacent full step positions. If the phases are switched in the sequence AB, BC, CA, AB, etc., the motor will take full (308) steps, as in the one-phase-on mode, but its equilibrium positions will be interleaved between the full step positions.

To obtain 'half stepping' the phases are excited in the sequence A, AB, B, BC, etc., i.e. alternately in the one-phase-on and two-phase-on modes. This is sometimes known as 'wave' excitation, and it causes the rotor to advance in steps of 158, or half the full step angle. As might be expected, continuous half stepping usually produces a smoother shaft rotation than full stepping, and it also doubles the resolution.

We can see what the static torque curve looks like when two phases are excited by superposition of the individual phase curves. An example is shown in Figure 4.7, from which it can be seen that for this machine, the holding torque (i.e. the peak static torque) is higher with two phases excited than with only one excited. The stable equilibrium (half-step) position is at 158, as expected. The price to be paid for the increased holding torque is the increased power dissipation in the windings, which is doubled as compared with the one-phase-on mode. The holding torque increases by a factor less than two, so the torque per watt (which is a useful figure of merit) is reduced.

A word of caution is needed in regard to the addition of the two separate one-phase-on torque curves to obtain the two-phase-on curve. Strictly, such a procedure is only valid where the two phases are magnetically independent, or the common parts of the magnetic circuits are unsaturated. This is not the case in most motors, in which the phases share a common magnetic circuit, which operates under highly saturated conditions. Direct addition of the one-phase-on curves cannot therefore be expected to give an accurate result for the two-phase-on curve, but it is easy to do, and provides a reasonable estimate.

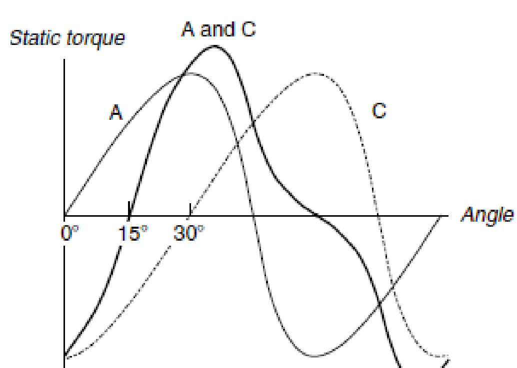


Figure 4.7: Static torque-angle curve (thick line) corresponding to two-phase-on excitation

Apart from the higher holding torque in the two-phase-on mode, there is another important difference which distinguishes the static behavior from that of the one-phase-on mode. In the one-phase-on mode, the equilibrium or step positions are determined solely by the geometry of the rotor and stator: they are the positions where the rotor and stator are in line. In the two-phase-on mode, however, the rotor is intended to come to rest at points where the rotor poles are lined-up midway between the stator poles. This position is not sharply defined by the 'edges' of opposing poles, as in the one-phase-on case; and the rest position will only be exactly midway if (a) there is exact geometrical symmetry and, more importantly (b) the two currents are identical. If one of the phase currents is larger than the other, the rotor will come to rest closer to the phase with the higher current, instead of halfway between the two. The need to balance the currents to obtain precise half stepping is clearly a drawback to this scheme. Paradoxically, however, the properties of the machine with unequal phase currents can sometimes be turned to good effect, as we now see.

4.1.3 Micro-step Mode

There are some applications (e.g. in printing and phototypesetting) where very fine resolution is called for, and a motor with a very small step angle – perhaps only a fraction of a degree – is required. We have already seen that the step angle can only be made small by increasing the number of rotor teeth and/or the number of phases, but in practice it is inconvenient to have more than four or five phases, and it is difficult to manufacture rotors with more than 50–100 teeth. This means it is rare for motors to have step angles below about 18. When a smaller step angle is required a technique known as micro-stepping (mini-stepping or step division) is used.

Micro-stepping is a technique based on two-phase-on operation which provides for the subdivision of each full motor step into a number of 'substeps' of equal size. In contrast with half stepping, where the two currents have to be kept equal, the currents are deliberately made unequal.
By correctly choosing and controlling the relative amplitudes of the currents, the rotor equilibrium position can be made to lie anywhere between the step positions for each of the two separate phases.

Closed-loop current control is needed to prevent the current from changing as a result of temperature changes in the windings, or variations in the supply voltage; and if it is necessary to ensure that the holding torque stays constant for each micro-step both currents must be changed according to a prescribed algorithm. Despite the difficulties referred to above, mini-stepping is used extensively, especially in photographic and printing applications where a high resolution is needed. Schemes involving between 3 and 10 micro-steps for a 1.88 step motor are numerous, and there are instances where up to 100 micro-steps (20 000 micro-steps/rev) have been successfully achieved.

So far, we have concentrated on those aspects of behavior, which depend only on the motor itself, i.e. the static performance. The shape of the static torque curve, the holding torque and the slope of the torque curve about the step position have all been shown to be important pointers to the way the motor can be expected to perform. All of these characteristics depend on the current(s) in the windings, however, and when the motor is running the instantaneous currents will depend on the type of drive circuit employed, as discussed in the next two sections.

What is commonly referred to as micro-stepping is often "sine cosine micro-stepping" in which the winding current approximates a sinusoidal AC waveform. Sine cosine micro-stepping is the most common form, but other waveforms can be used. Regardless of the waveform used, as the micro-steps become smaller, motor operation becomes more smooth, thereby greatly reducing resonance in any parts the motor may be connected to, as well as the motor itself. Resolution will be limited by the mechanical staction, backlash, and other sources of error between the motor and the end device. Gear reducers may be used to increase resolution of positioning.

Step size repeatability is an important step motor feature and a fundamental reason for their use in positioning. Example: many modern hybrid step motors are rated such that the travel of every full step (example 1.8 Degrees per full step or 200 full steps per revolution) will be within 3% or 5% of the travel of every other full step; as long as the motor is operated within its specified operating ranges. Several manufacturers show that their motors can easily maintain the 3% or 5% equality of step travel size as step size is reduced from full stepping down to 1/10 stepping. Then, as the micro-stepping divisor number grows, step size repeatability degrades. At large step size reductions it is possible to issue many micro-step commands before any motion occurs at all and then the motion can be a "jump" to a new position.

4.1.4 Additional Methods of Damping Rotor Oscillations

At very low stepping rates the motor comes to rest at the appropriate equilibrium position after each excitation change. The response of the system to each excitation change – known as the single-step response – is generally very oscillatory (Russell and Pickup, 1996); a typical response is shown in Fig. 4.8. In applications requiring frequent accurate positioning this poorly damped response can be a great disadvantage. For example, if a stepping motor is used to drive a printer carriage then the system must come to rest for the printing of each letter. The operating speed of the printer is limited by the time taken for the system to settle to within the required accuracy at each letter position. The frequency of oscillation can be predicted for any motor/load combination from the static torque/rotor position characteristic, provided the system is lightly damped. At a rotor position θ from the equilibrium position the motor torque is – $T'\theta$, where T' is the stiffness of the torque/position characteristic. If there is no load torque then this motor torque is used to accelerate the motor/load inertia (J); therefore:

$$-T'\theta = J\left(\frac{d^2\theta}{dt^2}\right) \qquad (4.1)$$

$$J\left(\frac{d^2\theta}{dt^2}\right) + T'\theta = 0$$

This is an equation of simple harmonic motion for the rotor position and so the natural frequency f_n of rotor oscillation about the equilibrium position is

$$f_n = \frac{1}{2\pi}\sqrt{\frac{T'}{J}} \qquad (4.2)$$

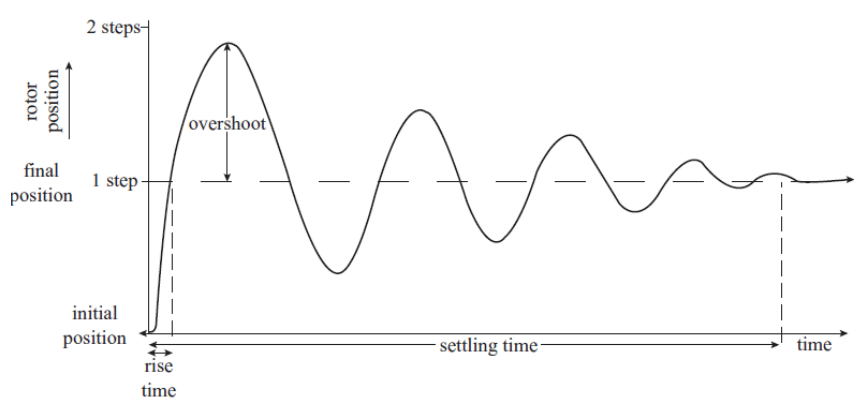


Figure 4.8: Typical single-step response

The simple analysis of oscillation frequency assumes that the system is undamped. In practice there is a small amount of viscous friction present in the system so that the oscillations are lightly damped and the rotor eventually settles at the equilibrium position, as illustrated in Fig. 4.8. Friction effects in an electromechanical system are generally undesirable, since they lead to wear in the moving parts, and are variable, because they are a function of this wear. The designer attempts to reduce friction as far as possible, so most stepping motor systems have very little inherent damping and consequently a poorly damped single-step response.

The parameters of the single-step response are defined in Fig. 4.8. Rise time is the time taken for the motor to first reach the demanded step position, which is attained with maximum velocity. The system therefore overshoots the target and the amplitude of this first overshoot is expressed as a percentage of the total step, giving the percentage overshoot. Finally the settling time is the time taken for oscillation to decay so that the system is within 5% of the target.

One consequence of the highly oscillatory single-step response is the existence of resonance effects at stepping rates up to the natural frequency of rotor oscillation. Figure 4.9 shows two responses of a motor to a series of steps at different rates. In the first response the stepping rate is about 0.6 times the natural frequency and therefore the rotor is behind the equilibrium position and has a low velocity when the next excitation change occurs. The rotor quickly settles into a uniform response to each step. In the other response the stepping rate is approximately equal to the natural frequency and so the rotor is at the equilibrium position with a positive velocity at the end of the first step. As a result of this initial velocity the response to the second step is more oscillatory; the rotor swings still further from the equilibrium position.

The rotor oscillations increase in amplitude as successive steps are executed until the rotor lags or leads the demanded step position by more than half a rotor tooth pitch. Once this oscillation amplitude is exceeded the motor torque causes the rotor to move towards an alternative step position which is a complete rotor tooth pitch from the expected position. The correspondence between rotor position and the number of excitation changes is now lost and the subsequent rotor movement is erratic. Note that motors with a large number of phases have an advantage here, since a step length is a small proportion of the rotor tooth pitch and therefore in these motors the rotor can be several steps from the demanded position without losing synchronism.

This resonant behavior of the system leads to a loss of motor torque at well-defined stepping rates, as illustrated by the dips in the pull-out torque/speed characteristic. The location of these dips can be predicted if the natural frequency is known from direct measurement of the single-step response. Resonance is likely to occur if, at the end of the excitation interval, the rotor is in advance of the equilibrium position and has a positive velocity. These regions are indicated in Fig. 4.9. The rotor has to pass through these regions after times which are a multiple of the rotor oscillation period ($1/f_n$) and therefore

Resonant stepping rates $= \frac{f_n}{k} = \frac{1}{2\pi k}\sqrt{\frac{T\prime}{J}}\ , k = 1, 2, 3 \ldots.$ (4.3)

A motor with a natural frequency, from eqn (4.4), of 100 Hz can be expected to have dips in the torque/speed characteristics at 100, 50, 33, 25, 20, . . . steps per second.

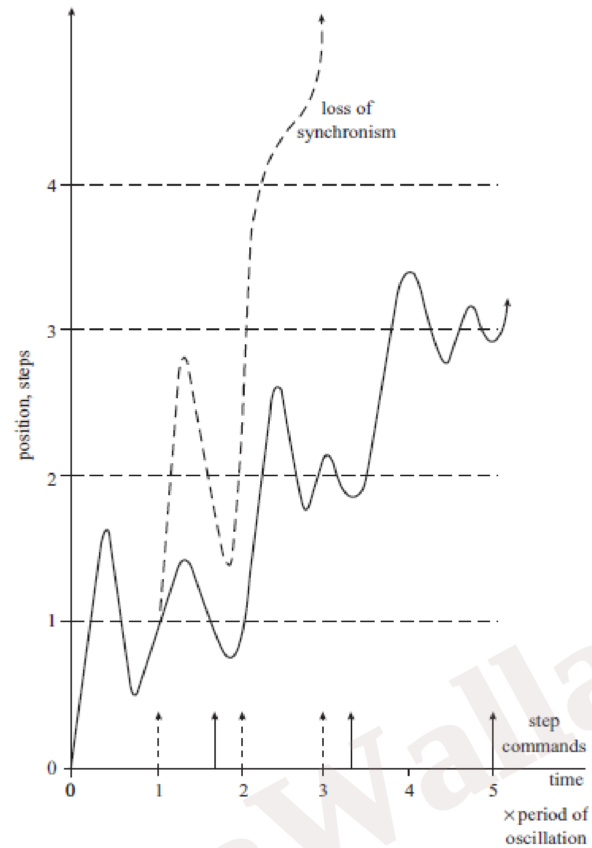


Figure 4.9: Responses to stepping rates near the natural frequency

- - - - Stepping rate = natural frequency

——— Stepping rate = 0.6 x natural frequency

This result is not precise because the oscillation frequency depends on the amount of damping, but it is sufficiently accurate for most purposes. The additional complication of damping-dependent oscillation frequency is included in the analysis by Lawrenson and Kingham (1977). For applications requiring repeated fast positioning over a single step, it is possible to utilise the high overshoot of the system. If the step corresponds to a change of excitation from phase *A* to phase *B*, for example, the half-step with both phases *A* and *B* excited is first taken. Figure 4.10 shows that the system overshoots the demanded position for *A* and *B* excited, coming to rest near the phase *B* equilibrium position. At this time the excitation is switched to phase *B* only and the transition to the final step is accomplished from a small initial error with consequently small overshoot. The contrast between this response and the effect of changing directly from single-phase excitation of *A* to *B* is shown in Fig. 4.11. Unfortunately the timing of the excitation changes in this intermediate half-step control is quite critical and is heavily dependent on load conditions (Miura and Taniguchi, 1999). It is therefore restricted in application to situations where the load is constant, or to closed-loop position control systems.

The resonant tendencies of a stepping motor system can be reduced by introducing more damping and therefore limiting the amplitude of oscillation in the single-step response. There are two important techniques for improving the damping, using either mechanical or electrical methods, and these are discussed in the following sections.

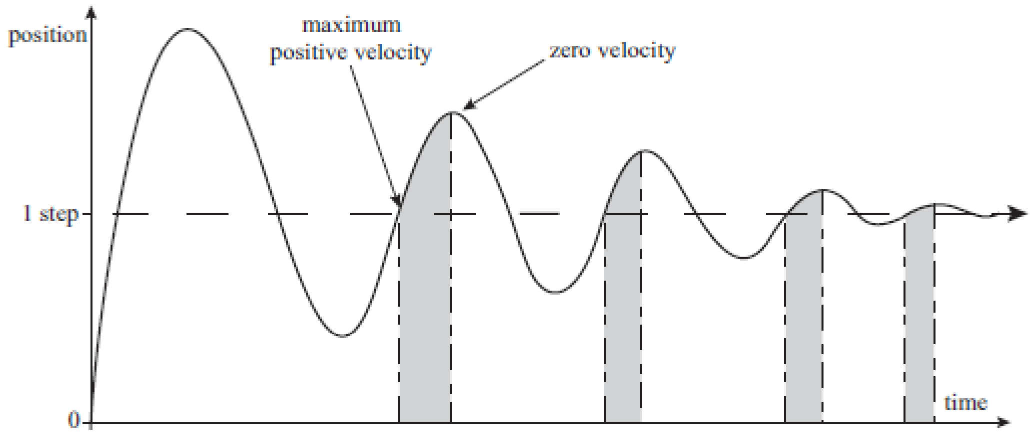


Figure 4.10 Regions of the single-step response in which phase switching leads to resonance

1. The viscously coupled inertia damper

One mechanical method of damping the single-step response is to introduce additional viscous friction (torque proportional to speed), so that the rotor oscillations decay at a faster rate (Kent, 1973). However, the use of straightforward viscous friction is undesirable because the operation of the motor at high speeds is severely limited by the friction torque. A solution to this problem is the viscously coupled inertia damper (VCID), sometimes known as the Lanchester damper. This device gives a viscous friction torque for rapid speed changes, such as occur in the single-step response, but does not interfere with operation at constant speeds.

The essential features of the VCID are illustrated in Fig. 4.12. Externally the damper appears as a cylindrical inertial load which can be clamped to the motor shaft so the damper housing rotates at the same speed as the motor. Internally the damper has a high inertia rotor which is separated from the housing by a viscous fluid. The housing and the inner rotor can therefore rotate relative to each other, but are loosely coupled by the viscous fluid. When relative motion occurs between the damper components there is a mutual drag torque.

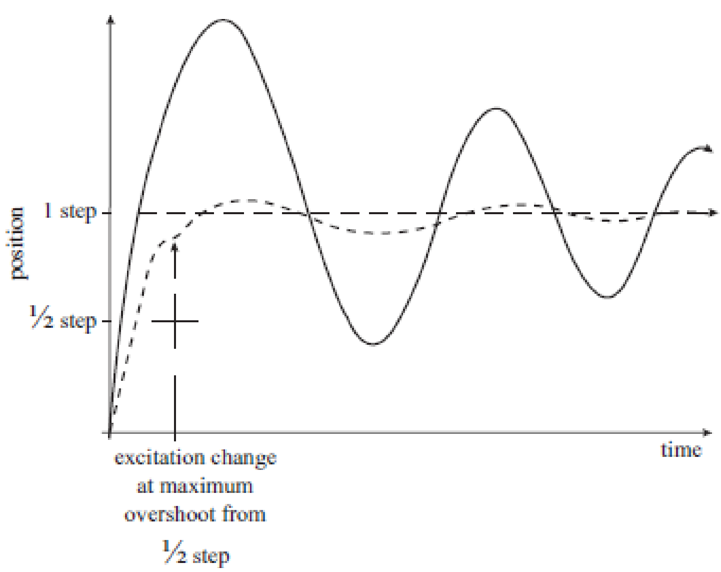


Figure 4.11: Intermediate half-step response

– – – Response to full-step excitation change

——— Response with intermediate half-step

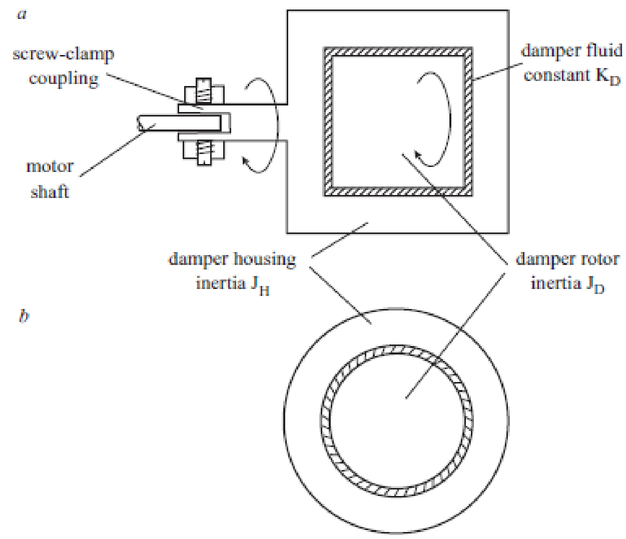


Figure 4.12: Cross-sections of the viscously coupled inertia damper

a. Parallel to the shaft

b. Perpendicular to the shaft

The basic parameters of the damper are its inertia (J_D) and the viscous f)luid constant k_D, but in addition there is the inertia of the damper housing (J_H), which increases the overall motor/load inertia. In terms of these parameters the drag torque (T_D) can be expressed in terms of the difference between housing and rotor velocities:

$$T_D = k_D\left[\left(\frac{d\theta}{dt}\right) - \left(\frac{d\theta_D}{dt}\right)\right] \tag{4.4}$$

where θ is the instantaneous position of the motor/load/housing and θ_D is the damper inertia position. Dampers are carefully designed so that this linear relationship is preserved over a wide range of speed difference. The damper torque acts as a drag torque on the motor shaft and also accelerates the damper rotor:

$$T_D = J_D\left(\frac{d^2\theta_D}{dt^2}\right) \tag{4.5}$$

If the motor is operating at constant speed the damper rotor must also be running at a constant speed, so the damper torque is zero. If the damper torque is zero, the damper rotor and housing must be operating at equal speeds. A well designed damper can produce a considerable improvement in the single step response. If the inherent friction torque of the system can be neglected the equation for the rotor position relative to equilibrium, eqn. (4.3), is modified by the damper torque:

$$T'\theta + J\left(\frac{d^2\theta}{dt^2}\right) + T_D = 0 \tag{4.6}$$

where J includes the damper housing inertia, J_H .

Substituting for *TD* from eqn. (4.4):

$$T'\theta + J\left(\frac{d^2\theta}{dt^2}\right) + k_D\left(\frac{d\theta}{dt}\right) - k_D\left(\frac{d\theta_D}{dt}\right) = 0 \tag{4.7}$$

θ_D can be eliminated by differentiating this expression and substituting from eqn. (4.7):

$$T'\left(\frac{d\theta}{dt}\right) + J\left(\frac{d^3\theta}{dt^3}\right) + k_D\left(\frac{d^2\theta}{dt^2}\right) - k_d\frac{T_D}{J_D} = 0 \tag{4.8}$$

Finally an expression wholly in terms of rotor position can be obtained by substituting for T_D from eqn. (4.8):

$$T'\left(\frac{d\theta}{dt}\right) + J\left(\frac{d^3}{dt^3}\right) + k_D\left(\frac{d^2\theta}{dt^2}\right) + \frac{k_D T'\theta}{J_d} + \frac{k_d J\left(\frac{d^2\theta}{dt^2}\right)}{J_D} = 0 \tag{4.9}$$

and therefore

$$\left(\frac{d^3\theta}{dt^3}\right) + k_D\left(\frac{1}{J} + \frac{1}{J_D}\right)\left(\frac{d^2\theta}{dt^2}\right) + \left(\frac{T'}{J}\right)\left(\frac{d\theta}{dt}\right) + \left(\frac{k_d T'}{JJ_D}\right) = 0 \tag{4.10}$$

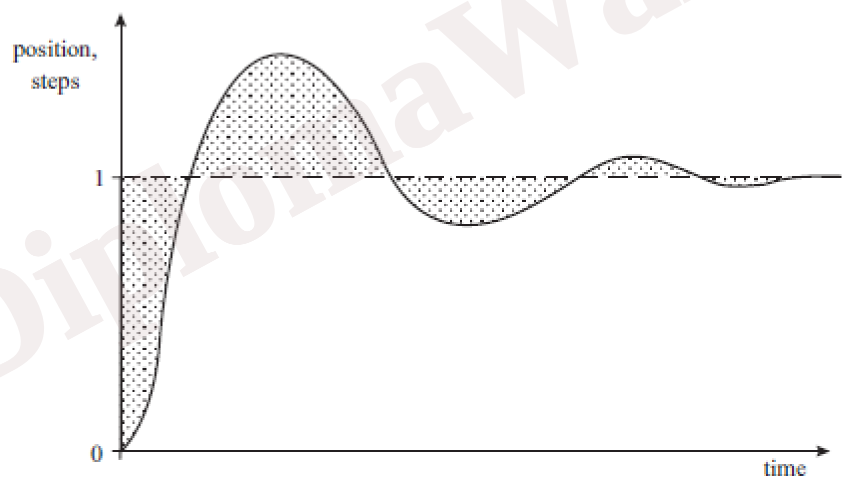


Figure 4.13: Single-step response showing the area corresponding to the integral of the absolute error (IAE)

So the single-step response of a system with a VCID is third-order, compared to the second-order eqn. (4.3) for the system without a damper. Turning now to design it is important to answer the question of which damper produces the 'best' single-step response from a given system. In terms of the parameters, we have to choose J_D and k_D for the damper, given J and T for the motor/load. One problem here is to establish a suitable criterion to judge the quality of a single-step response. Faced with this problem, Lawrenson and Kingham (1973) chose to minimize the integral-of-absolute-error (IAE), which corresponds to minimising the area shown in Fig. 4.13. Using this criterion it was possible to show that the damper inertia should be four times the total motor/load inertia (including the housing inertia):

$$J_D = 4 \times J \tag{4.11}$$

and the viscous fluid constant should then be related to the stiffness and damper inertia by

$$k_D = 0.53\sqrt{T'J_D} \qquad (4.12)$$

If a well matched damper is coupled into the system the improvement in the single step response is achieved with a shorter settling time and lower overshoot, as shown in Fig. 4.14. However, the damper does have the effect of increasing the rise time of the response, because the available motor torque has to accelerate both the load and damper inertias towards the step position.

When a VCID is used the total motor/load inertia includes the damper housing inertia and therefore it is important that this inertia be as small as possible. The penalty to be paid for the use of a VCID is that the system is slower to accelerate. Even if the viscous coupling is low (so that the damper housing and rotor operate almost independently) the system inertia is increased by the housing inertia and the acceleration is correspondingly reduced. With a high value of k_D the damper housing and rotor are closely coupled, so the effective system inertia is increased by both the housing and damper rotor inertia.

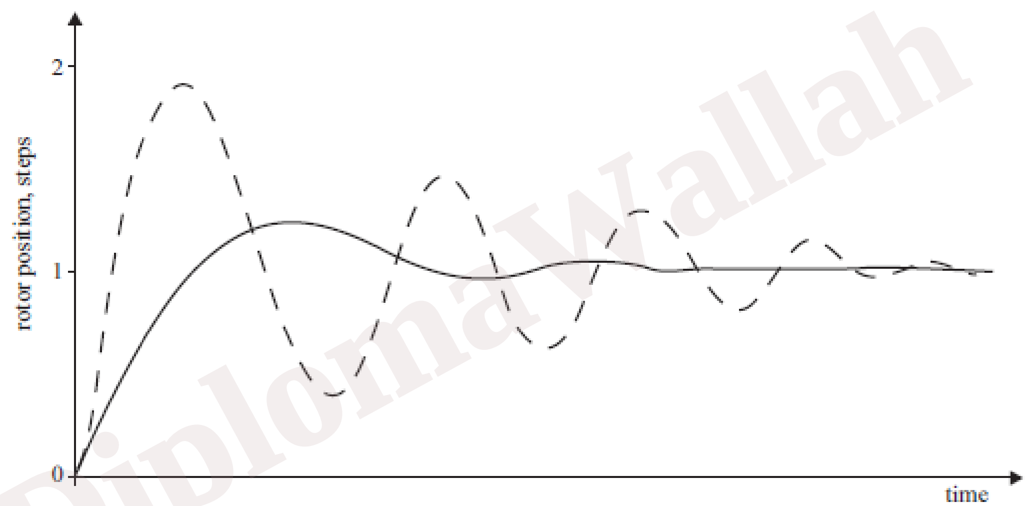


Figure 4.14: Effect of VCID on the ssingle-step response

———— with VCID

– – – – without VCID

2. Electromagnetic damping

The basic aim of any damping scheme is to extract stored mechanical energy, which is in the form of rotational kinetic energy when the system inertia is moving. Damping with the VCID is achieved by transferring the system's mechanical energy to and fro between the damper housing and rotor using an inefficient method of coupling (the viscous fluid), so that some of the energy is dissipated with each transfer. Therefore in the VCID the mechanical energy is used to heat the coupling fluid. In electromagnetic damping schemes the mechanical energy acquired by the system in moving between the step positions is transferred to the motor's electrical circuit and dissipated in the motor winding and forcing resistances.

The transfer of energy to the electrical circuit is accomplished by means of the voltages induced in the phase windings when the rotor oscillates, so these voltages are considered first. Figure 4.15*a* shows the variation of magnet flux linked with the two phase windings of a hybrid motor as the rotor position varies over a rotor tooth pitch. This characteristic is approximately sinusoidal with a wavelength equal to the rotor tooth pitch, and the sinusoids for the two phases are displaced by $\pi/2p$. The rate of change of flux linkages with rotor position is shown in Fig. 4.15*b*, which has the correct phase relationship to Fig. 4.14*a*. When the flux linkages in phase *A* are at a maximum, for example, the rate of change of flux linked with phase *A* is zero.

For a given phase current the torque produced by one phase is proportional to the rate of change of flux linkages with rotor position and therefore the static torque/rotor position characteristics for positive excitation of the two phases can be deduced, as in Fig. 4.14*c*.

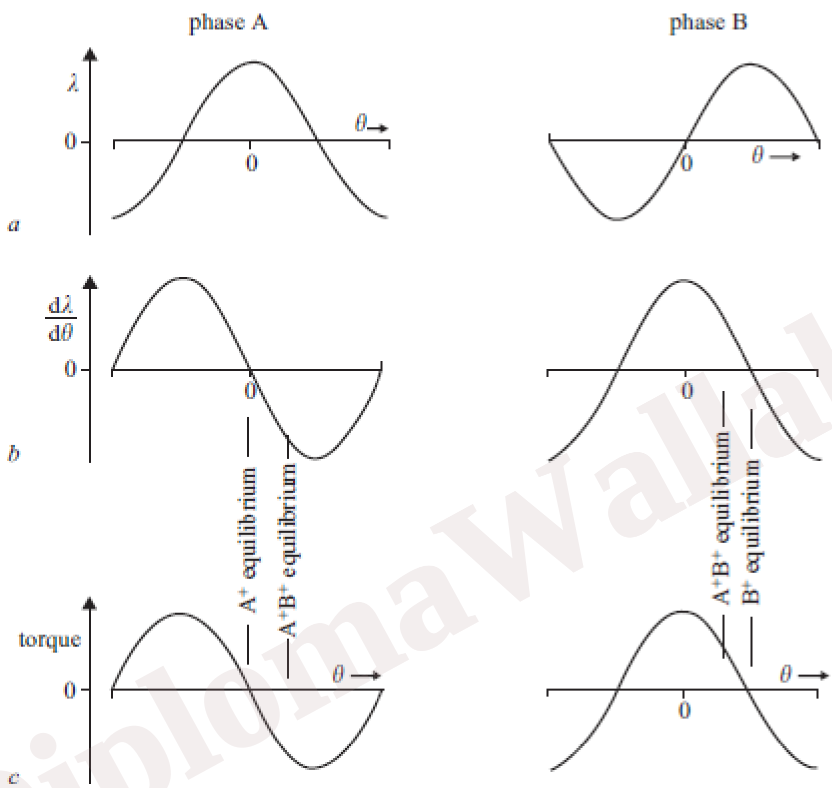


Figure 4.15: Flux linkage/rotor position characteristics

a. **Flux linkages against rotor position**
b. **Rate of change of flux linkages against rotor position**
c. **Static torque against rotor position**

With only one of the phases excited the rotor moves to the phase equilibrium position, where the torque is zero and, from Fig. 4.15*b*, the rate of change of flux linkages is also zero. If the rotor oscillates about this one-phase-on equilibrium position the flux linked with the phase winding undergoes only small changes and the voltage induced by the magnet flux is insignificant. Now consider the situation when two phases are excited. The equilibrium position is between the two separate phase equilibrium positions and so the rate of change of flux linkages with rotor position is relatively large. Therefore if the rotor is oscillating about the two-phases-on equilibrium position a voltage is induced in each phase by the magnet flux with a frequency equal to the frequency of rotor oscillation. It is these induced voltages which are used to extract energy from the mechanical system and provide electromagnetic damping.

A rigorous analysis of the mechanisms involved in electromagnetic damping has been undertaken by Hughes and Lawrenson (1975), who demonstrated that the single step response is third-order when the electrical circuit is taken into account. The results of this analysis show that for a well damped response the phase resistance (winding forcing) must be set at an optimum value which depends on several parameters of the motor and load. With two-phases-on excitation, damping occurs because the induced voltages produce additional ac components of phase current, which are superimposed on the steady dc phase current. These ac current components give extra power losses in the phase resistance when the rotor is oscillating, so mechanical energy is extracted from the system to supply this extra power. If the phase resistance is set too high the ac current component is low and the power losses ($i^2 r$) are small. Conversely if the phase resistance is below the optimum value the ac current is high, but there is very little resistance in which the current can dissipate power. For a hybrid stepping motor the optimum value of phase resistance for maximum electromagnetic damping is

$$\boldsymbol{R} = \sqrt{\frac{\boldsymbol{T'}}{\boldsymbol{J}}}\ \boldsymbol{x}\ \boldsymbol{L}\ \boldsymbol{x}\ (\mathbf{1} + \frac{\boldsymbol{k}}{\mathbf{2}}) \qquad (4.13)$$

where *L* is the inductance of the phase winding. The factor *k* is a parameter of the motor and depends on the ratio of the magnet flux linking the phase winding to the flux linkages brought about by the winding current. Typical values of *k* are in the range 0.25–1.0. A similar result to eqn (4.11) also applies to variable-reluctance motors, except that the parameter *k* has a different definition.

Although the optimum phase resistance can be calculated, in practice it is a fairly simple matter to determine the optimum experimentally. The single-step response can be examined over a range of forcing resistance values (with appropriate changes of supply voltage to maintain constant phase current) until a suitable response is obtained. The discussion has centered on the two-phase hybrid motor, but electromagnetic damping can be produced in all types of motor, provided more than one phase is excited when the rotor is settling to the equilibrium position. In some cases the electromagnetic damping effect can be enhanced by introducing a dc bias to all phases of the motor (Tal and Konecny, 1980). In addition, Jones and Finch (1983) have shown that the single-step response can be optimized by allowing the phase winding currents to change gradually.

As with the VCID, the design of a system for good damping using electromagnetic methods is often in direct conflict with the demands of high-speed operation. In the next chapter it is shown that the system requires a large forcing resistance to operate at the highest speeds and in most cases the total phase resistance is then much greater than the optimum for electromagnetic damping. The system designer is therefore left to make a compromise choice of forcing resistance according to the application.

4.1.5 Permanent Magnet Stepper Motors

Permanent-magnet stepper motors have smooth armatures and include a permanent magnet core that is magnetized widthwise or perpendicular to its rotation axis. These motors usually have two independent windings, with or without center taps. The most common step angles for PM motors are 45° and 90°, but motors with step angles as fine as 1.8° per step as well as 7.5, 15, and 30° per step are generally available.

Armature rotation occurs when the stator poles are alternately energized and de-energized to create torque. A 90° stepper has four poles and a 45° stepper has eight poles, and these poles must be energized in sequence. Permanent-magnet steppers step at relatively low rates, but they can produce high torques and they offer very good damping characteristics.

Permanent magnet (PM) stepper motors are similar in construction to that of single stack, variable reluctance stepper motors except that the rotor is made of permanent magnet. Fig. 4.16 presents the circuit configuration and different operation modes for a 2-phase, permanent magnet stepper motor that rotate in an anticlockwise direction with a 90° step. Table 2 and Fig. 4.16 present each phase PM switching sequence. PM stepper motors offer many features compared to variable reluctance type such as:

- Higher inertia and consequently lower acceleration (deceleration) rates.
- Maximum step pulse rate is 300 pulses per second compared to 1200 pulses per second for variable reluctance stepper motors.
- Larger step sizes, ranging from 30° to 90° compared to step sizes as low as 1.8° for variable reluctance stepper motors.
- Generate higher torque per ampere of stator currents than variable reluctance stepper motors.

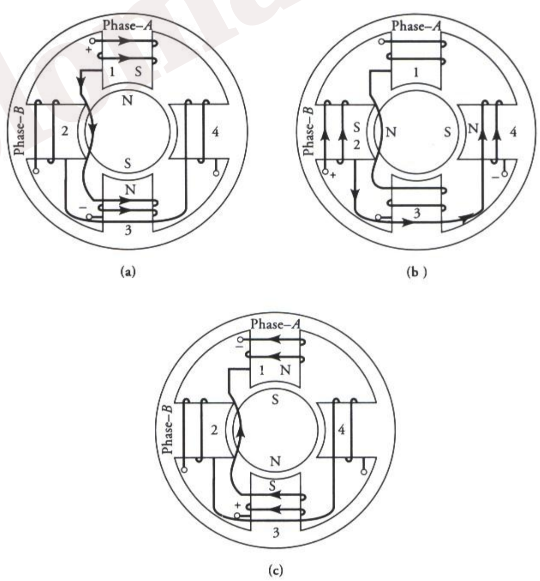


Fig. 4.16: Construction and operation of 2-phase, permanent magnet stepper motor

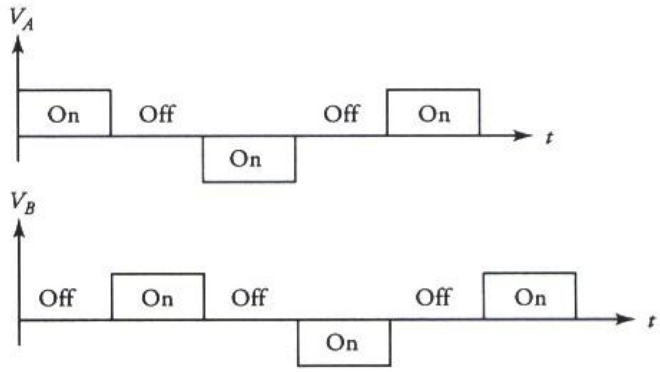


Figure 4.17: Phase switching sequence

Table 2: Phase switching sequence: “1”, “-1” and “0” corresponds to positive, negative, and zero phase voltage (currents), respectively

Cycle	Phase A	Phase B	Position δ°
+	1	0	0
	0	1	90
−	−1	0	180
	0	−1	270
+	1	0	360

4.1.6 Stepper motor drives circuit

In this section, we will look at how the motor would perform if it were supplied by an ideal drive circuit, which turns out to be one that is capable of supplying rectangular pulses of current to each winding when required, and regardless of the stepping rate. Because of the inductance of the windings, no real drive circuit will be able to achieve this, but the most sophisticated (and expensive) ones achieve near-ideal operation up to very high stepping rates.

The basic function of the complete drive is to convert the step command input signals into appropriate patterns of currents in the motor windings. This is achieved in two distinct stages, as shown in Figure 4.18, which relates to a 3-phase motor.

The ‘translator’ stage converts the incoming train of step command pulses into a sequence of on/off commands to each of the three power stages. In the one-phase-on mode, for example, the first step command pulse will be routed to turn on phase A, the second will turn on phase B and so on. In a very simple drive, the translator will probably provide for only one mode of operation (e.g. one-phase-on), but most commercial drives provide the option of one-phase-on, two-phase-on and half stepping. Single-chip ICs with these three operating modes and with both three-phase and four-phase outputs are readily available. The power stages (one per phase) supply the current to the windings. An enormous diversity of types is in use, ranging from simple ones with one switching transistor per phase, to elaborate chopper-type circuits with four transistors per phase. At this point, however, it is helpful to list the functions required of the ‘ideal’ power stage. These are firstly that when the translator calls for a phase to be energized, the full (rated) current should be established immediately; secondly, the current should be maintained constant (at its rated value) for the duration of the ‘on’ period and finally, when the translator calls for the current to be turned off, it should be reduced to zero immediately.

The ideal current waveforms for continuous stepping with one phase-on operation are shown in the lower part of Figure 4.18. The currents have a square profile because this leads to the optimum value of running torque from the motor. But because of the inductance of the windings, no real drive will achieve the ideal current waveforms, though many drives come close to the ideal, even at quite high stepping rates. Drives which produce such rectangular current waveforms are (not surprisingly) called constant-current drives. We now look at the running torque produced by a motor when operated from an ideal constant current drive. This will act as a yardstick for assessing the performance of other drives, all of which will be seen to have inferior performance.

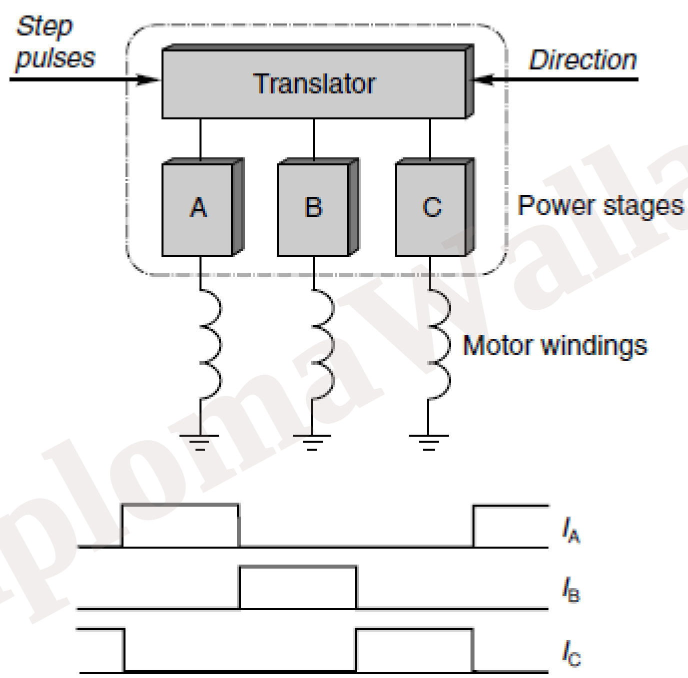


Figure 4.18: General arrangement of drive system for 3-phase motor and winding currents corresponding to an “ideal” drive

Pull-out torque under constant-current conditions

If the phase currents are taken to be ideal, i.e. they are switched-on and switched-off instantaneously, and remain at their full-rated value during each ‘on’ period, we can picture the axis of the magnetic Weld to be advancing around the machine in a series of steps, the rotor being urged to follow it by the reluctance torque. If we assume that the inertia is high enough for fluctuations in rotor velocity to be very small, the rotor will be rotating at a constant rate, which corresponds exactly to the stepping rate.

Now if we consider a situation where the position of the rotor axis is, on average, lagging behind the advancing Weld axis, it should be clear that, on average, the rotor will experience a driving torque. The more it lags behind, the higher will be the average forward torque acting on it, but only up to a point. We already know that if the rotor axis is displaced too far from the Weld axis, the torque will begin to diminish, and eventually reverse, so we conclude that although more torque will be developed by increasing the rotor lag angle, there will be a limit to how far this can be taken.

Turning now to a quantitative examination of the torque on the rotor, we will make use of the static torque–displacement curves discussed earlier, and look at what happens when the load on the shaft is varied, the stepping rate being kept constant. Clockwise rotation will be studied, so the phases will be energized in the sequence ABC. The instantaneous torque on the rotor can be arrived at by recognizing (a) that the rotor speed is constant, and it covers one-step angle (308) between step command pulses, and (b) the rotor will be 'acted on' sequentially by each of the set of torque curves.

When the load torque is zero, the net torque developed by the rotor must be zero (apart from a very small torque required to overcome friction). This condition is shown in Figure 4.19(a). The instantaneous torque is shown by the thick line, and it is clear that each phase in turn exerts first a clockwise torque, then an anticlockwise torque while the rotor angle turns through 308. The average torque is zero, the same as the load torque, because the average rotor lag angle is zero.

When the load torque on the shaft is increased, the immediate effect is to cause the rotor to fall back in relation to the Weld. This causes the clockwise torque to increase, and the anticlockwise torque to decrease. Equilibrium is reached when the lag angle has increased sufficiently for the motor torque to equal the load torque. The torque developed at an intermediate load condition like this is shown by the thick line in Figure 4.19(b). The highest average torque that can possibly be developed is shown by the thick line in Figure 4.19(c): if the load torque exceeds this value (which is known as the pull-out torque) the motor loses synchronism and stalls, and the vital one-to-one correspondences between pulses and steps are lost.

Since we have assumed an ideal constant-current drive, the pull-out torque will be independent of the stepping rate, and the pull-out torque– speed curve under ideal conditions is therefore as shown in Figure 4.20. The shaded area represents the permissible operating region: at any particular speed (stepping rate) the load torque can have any value up to the pull-out torque, and the motor will continue to run at the same speed. But if the load torque exceeds the pull-out torque, the motor will suddenly pull out of synchronism and stall.

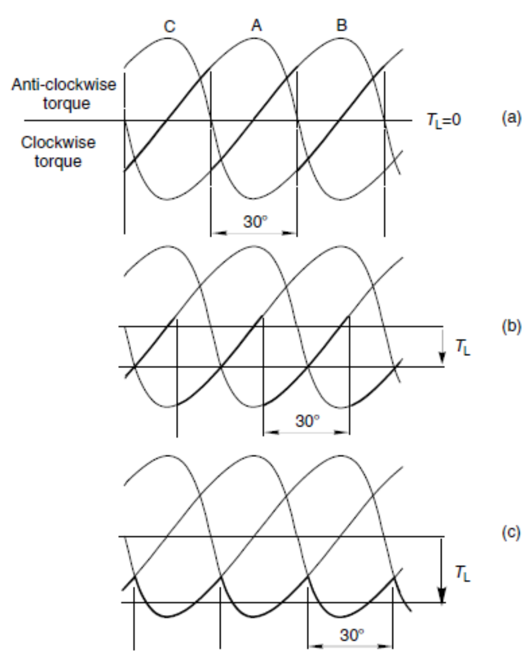


Figure 4.19: Static torque curves indicating how the average steady-state torque (T_L) is developed during constant-frequency operation

As mentioned earlier, no real drive will be able to provide the ideal current waveforms, so we now turn to look briefly at the types of drives in common use, and at their pull-out torque–speed characteristics.

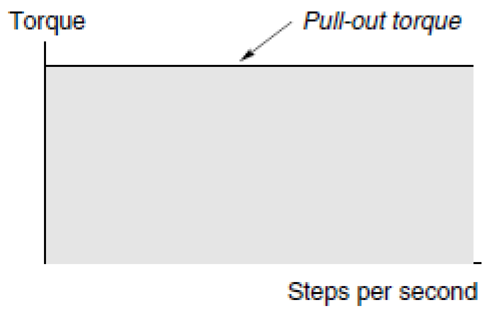


Figure 4.20: Steady-state operating region with ideal constant-current drive. In such idealized circumstances there would be no limit to the stepping rate, but as shown in Figure 9.14 a real drive circuit imposes an upper limit)

4.1.6.1 Drive Circuit

There are two main drive circuits for stepper motors, namely:

1. *Uni-polar*
2. *Bi-polar* drive circuits.

4.1.6.2 Uni-polar Drive Circuit

Fig. 4.21 presents a schematic diagram for a uni-polar drive circuit. This circuit is suitable for three phase variable reluctance stepper motors. Each phase winding of the motor is controlled by a separate drive circuit with a transistor as its controllable power switch. All drive circuits are energized by the same DC source. The transistor (power switch) of each winding has two modes of operation as follows:

On Mode: When sufficiently high base current flow through the transistor base it turn ON and acts ideally like a short circuit. Consequently, the supply voltage will be applied across the phase winding and the external resistor (R_{ext}) connected in series with the phase winding. The DC source magnitude is adjusted to produces the rated phase current when the switch is turned ON. Therefore,

$$V_s = I\,(R_{ph} + R_{ext})$$

where V_s is the DC source voltage in V, I is the phase winding rated current in A, R_{ph} is the phase winding resistance in Ω, and R_{ext} is the external resistance connected in series to the phase winding in Ω. The phase winding inductance is very large and consequently results in slow rate of building the phase winding current that might result in unsatisfactory operation of the stepper motor at high stepping rates. Therefore, the external resistance is connected in series with the phase winding to reduce the time constant. The net *ON Mode* circuit time constant will be very large and can be expressed by,

$$\tau\tau_{ON} = \frac{L_{ph}}{(R_{ph} + R_{ext})}$$

where L_{ph} is the phase winding average inductance in H

OFF Mode: In this mode, the base drive current of the transistor is removed and the switch is turned *OFF* and acts as an open circuit. The phase winding current will continue to flow through the freewheeling path formed by the freewheeling diode (D_f) and the freewheeling resistance (R_f). The maximum *OFF state* voltage appears across the transistor (switch) (*VCE (max))* can be expressed by,

$$VV_{CE(max)} = V_s + IR_f$$

During this mode of operation, phase current decays in the *OFF mode* circuit with a net *OFF Mode* circuit time constant that can be expressed by,

$$\tau_{OFF} = \frac{L_{ph}}{(R_{ph} + R_{ext} + R_f)}$$

The energy stored in the phase inductance during the *ON mode* is dissipated in the OFF mode circuit resistances during the switch turn OFF period.

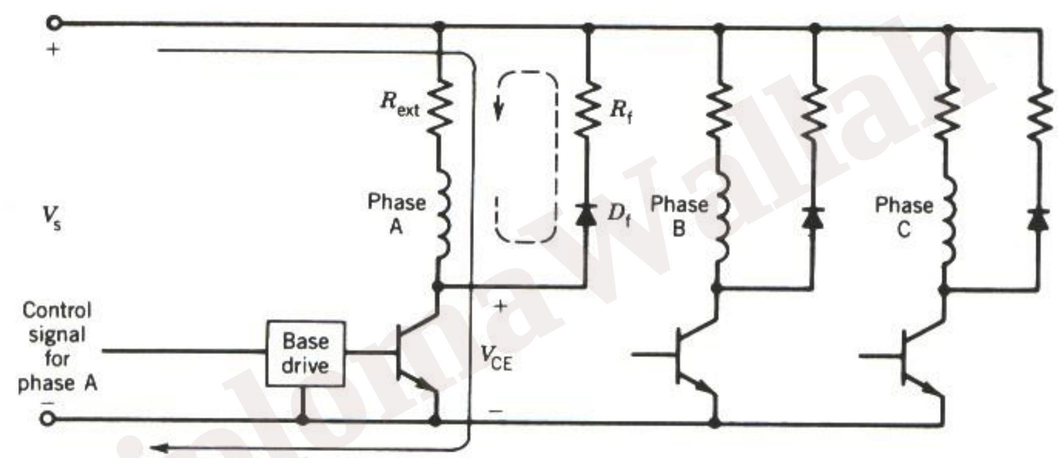


Fig. 4.21: Uni-polar drive circuit for three-phase variable reluctance stepper motor

4.1.6.3 Bi-polar Drive Circuit

Fig. 4.22 presents a schematic diagram for one phase of a bi-polar drive circuit. This circuit is suitable for permanent magnet or hybrid stepper motors. Each phase winding of the motor is controlled by a separate drive circuit with a transistor as its controllable power switch. All drive circuits are energized by the same DC source. Each two transistors (power switches) of each phase winding are turned ON simultaneously. Two modes of operation occur as follows:

T1 and T2 are in the On Mode: This is done by injecting sufficiently high base current through their bases simultaneously. Each transistor acts ideally like a short circuit. Consequently, the current will flow as indicated by the solid line in Fig. 18. The inductor is then energized.

D3 and D4 are in the On Mode: This mode follows the switching *OFF* of *T1* and *T2*. In this mode, the phase winding current cannot change its direction or decay to zero instantaneously after turning *OFF* of *T1* and *T2* because of the phase winding inductances. Thus the current continue to flow through of *D3* and *D4* as indicated by the dotted line in Fig. 18. The inductor discharges and the energy is returned back to the DC source.

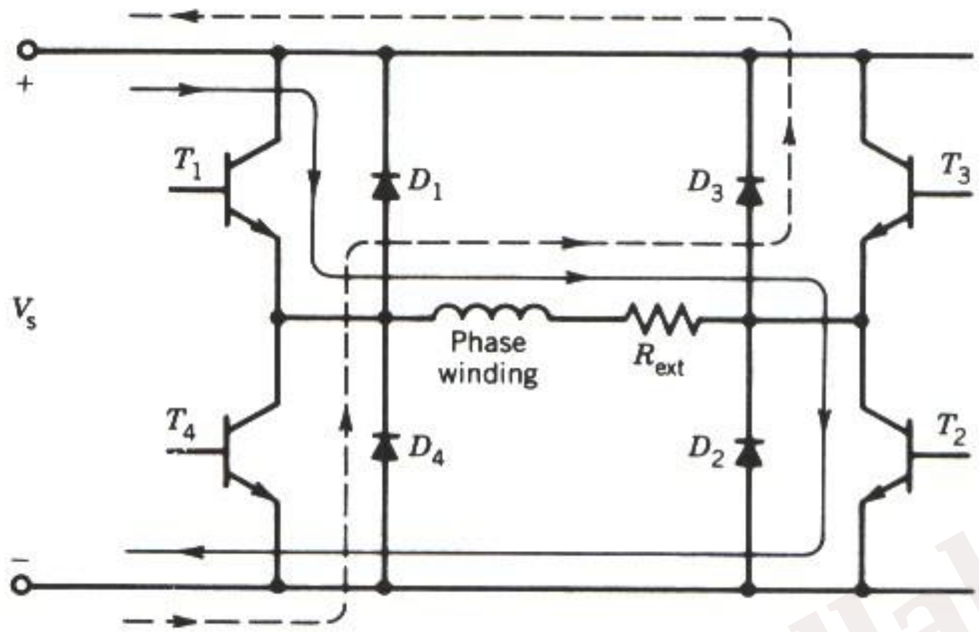


Figure 4.22: Bi-polar Drive Circuit for three-phase variable reluctance stepper motor

A reverse flow of current in the phase windings and hence a reverse direction of rotation of the motor can be achieved by activating *T3* and *T4*. When *T3* and *T4* are turned *OFF* the freewheeling path will provided through *D1* and *D2*. The bi-polar circuit is characterized by,

- Higher efficiency than the uni-polar drive circuit as part of the stored energy in the phase winding returns back to the DC source during the power switches turn OFF mode.
- Fast decaying of the freewheeling current as the inductor discharge through the phase winding resistance, phase external resistance and the DC source.
- No freewheeling resistance is required.
- More power switches (devices) than the uni-polar drive circuit.
- More expensive than the uni-polar drive circuit.
- Most of the large stepper motors types (> 1 kW) are driven by the bi-polar drive circuit including variable reluctance types.

4.1.7 Linear stepper motors

Linear actuators are available with axial integral threaded shafts and bolt nuts that convert rotary motion to linear motion. Powered by fractional horsepower permanent-magnet stepper motors, these linear actuators are capable of positioning light loads. Digital pulses fed to the actuator cause the threaded shaft to rotate, advancing or retracting it so that a load coupled to the shaft can be moved backward or forward. The bidirectional digital linear actuator shown in Figure 4.23 can provide linear res- solution as fine as 0.001 in. per pulse. Travel per step is determined by the pitch of the lead screw and step angle of the motor. The maximum linear force for the model shown is 75 oz.

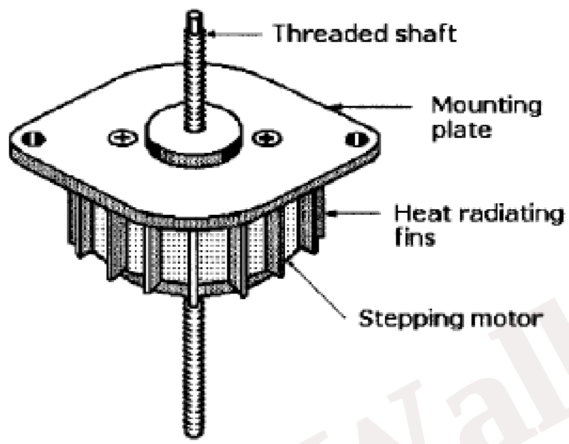


Figure 4.23: This light-duty linear actuator based on a permanent-magnet stepping motor has a shaft that advances or retracts

Calculation for Stepper motor pulse frequency, f_p

The step angle of stepper motors is related to the number of steps for the motor according to the relationship:

$$\alpha = \frac{360}{n_s}$$

where α is the step angle in degrees; and n_s is the number of steps for the stepper motor, which must be an integer value. The total angle through which the motor rotates (A_m) is given by:

$$A_m = n_p \alpha$$

where A_m is the total angle through which the motor rotates in degrees; n_p is the number of pulses received by the motor; and α is the step angle in degrees. Angular velocity is given by:

$$\omega = \frac{2\pi f_p}{n_s}$$

where ω is angular velocity; f_p is the pulse frequency; and n_s is the number of steps for the stepper motor. The speed of rotation is given by:

$$N = \frac{60 f_p}{n_s}$$

where N is the rotational speed; f_p is the pulse frequency; and n_s is the number of steps for the stepper motor.

EXAMPLE 4.2
A stepper motor has a step angle = 3.6°. (1) How many pulses are required for the motor to rotate through ten complete revolutions? (2) What pulse frequency is required for the motor to rotate at a speed of 100 rev/min?

(1) We know that the step angle is given by:

$$\alpha = \frac{360}{n_s}$$

where α is the step angle in degrees; and n_s is the number of steps for the stepper motor.

=> $3.6^\circ = 360 / n_s$

=> $3.6^\circ (n_s) = 360$

=> $n_s = 360 / 3.6 = 100$ step angles

The total angle through which the motor rotates (A_m) is given by:

$$A_m = n_p \alpha$$

where A_m is the total angle through which the motor rotates in degrees; n_p is the number of pulses received by the motor; and α is the step angle in degrees.

Now, to rotate through ten complete revolutions: $10(360^\circ) = 3600^\circ = A_m$

$n_p = 3600 / 3.6 = 1000$ pulses

(2) We know that:

$$N = \frac{60 f_p}{n_s}$$

where N is the rotational speed; f_p is the pulse frequency; and n_s is the number of steps for the stepper motor.

Thus, from the information derived from (1) above, and where N = 100 rev/min:

$100 = 60\, f_p / 100$

=> $10{,}000 = 60\, f_p$

=> $f_p = 10{,}000 / 60 = 166.667 = 167$ Hz

4.2 Apply control method of actuators

Continuous actuators allow a system to position or adjust outputs over a wide range of values. Even in their simplest form, continuous actuators tend to be mechanically complex devices. For example, a linear slide system might be composed of a motor with an electronic controller driving a mechanical slide with a ball screw. The cost for such actuators can easily be higher than for the control system itself. These actuators also require sophisticated control techniques that will be discussed in later chapters. In general, when there is a choice, it is better to use discrete actuators to reduce costs and complexity.

4.2.1 Apply control method for Brushless DC Motors

DC brushless motors are another type of servomotor in that feedback is required for stable operation. ADC brushless motor is like a DC brush motor turned inside out because the rotor contains a permanent magnet and the stator contains windings. The windings are electronically commutated so that the mechanical commutator and brushes are no longer required as compared with a DC brush motor. DC brushless motors are commonly used in robotics applications because of their high speed capability, improved efficiency, and low maintenance in comparison with DC brush motors. They are capable of higher speeds because of the elimination of the mechanical commutator. They are more efficient because heat from the windings in the stator can be dissipated more quickly through the motor case. Finally, they require less maintenance because they do not have brushes that require periodic replacement. However, the total system cost for brushless motors is higher than that for DC brush motors due to the complexity of electronic commutation.

The position of the rotor must be known so that the polarity of current in the windings of the stator can be switched at the correct time. Two types of commutation are used with brushless motors. With trapezoidal commutation, the rotor position must only be known to within 60° so that only three digital Hall Effect sensors are typically used. Sinusoidal commutation is employed instead of trapezoidal commutation when torque ripple must be reduced for the motor. In this case, the rotor position must be determined more accurately so that a resolver is used or an encoder is used in addition to Hall Effect sensors. The Hall Effect sensors are needed with the encoder to provide rotor shaft position at startup. The resolver provides absolute position information of the rotor shaft, so that the Hall Effect sensors are not required for startup.

Brushless motors use a permanent magnet on the rotor, and use windings on the stator. Therefore there is no need to use brushes and a commutator to switch the polarity of the voltage on the coil. The lack of brushes means that these motors require less maintenance than the brushed DC motors. A typical Brushless DC motor could have three poles, each corresponding to one power input, as shown in Figure 4.25. Each of coils is separately controlled. The coils are switched on to attract or repel the permanent magnet rotor. To continuously rotate these motors the current in the stator coils must alternate continuously. If the power supplied to the coils was a 3-phase AC sinusoidal waveform, the motor will rotate continuously. The applied voltage can also be trapezoidal, which will give a similar effect. The changing waveforms are controller using position feedback from the motor to select switching times. The speed of the motor is proportional to the frequency of the signal.

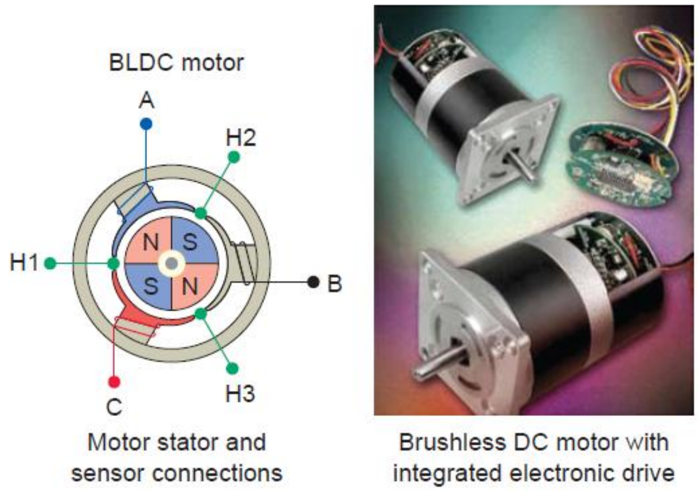


Figure 4.24: Brushless DC motor with integrated drive

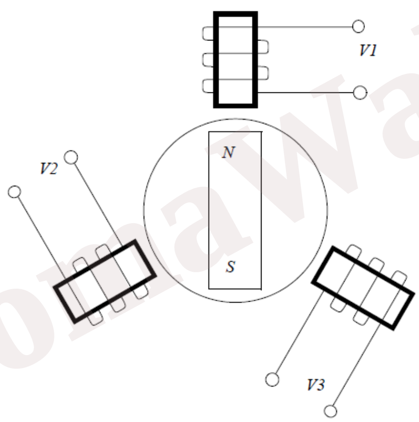


Figure 4.25: Brushless DC Motor

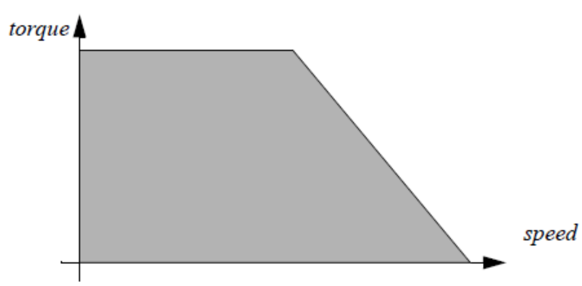


Figure 4.26: Torque speed curve for a Brushless DC Motor

In a DC motor there is normally a set of coils on the rotor that turn inside a stator populated with permanent magnets. Figure 4.27 shows a simplified model of a motor. The magnetics provide a permanent magnetic field for the rotor to push against. When current is run through the wire loop it creates a magnetic field.

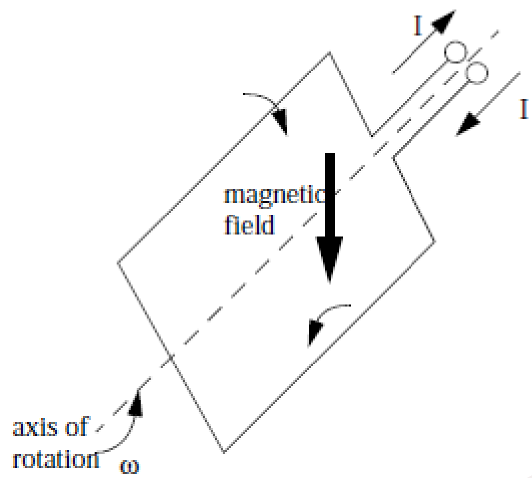


Figure 4.27: A simplified rotor

The power is delivered to the rotor using a commutator and brushes, as shown in Figure 4.28. In the figure the power is supplied to the rotor through graphite brushes rubbing against the commutator. The commutator is split so that every half revolution the polarity of the voltage on the rotor, and the induced magnetic field reverses to push against the permanent magnets.

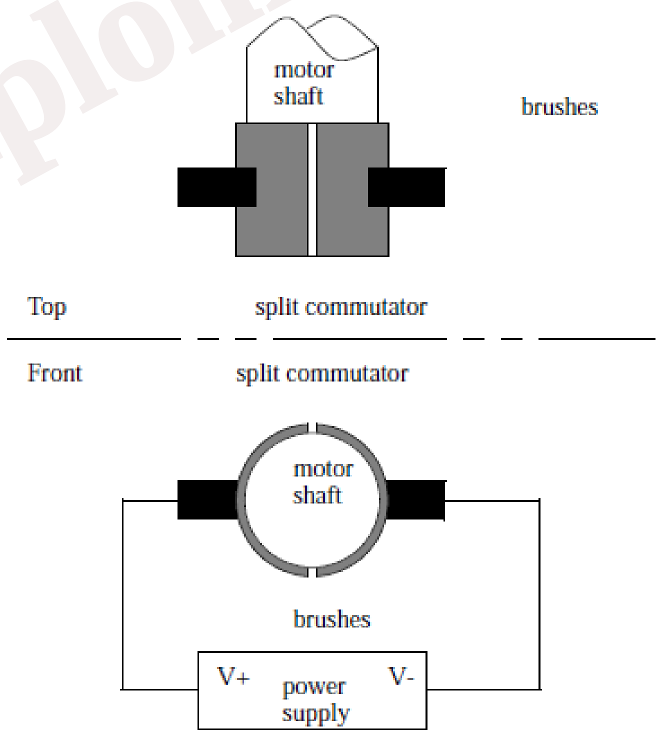


Figure 4.28: A split ring commutator

The direction of rotation will be determined by the polarity of the applied voltage, and the speed is proportional to the voltage. A feedback controller is used with these motors to provide motor positioning and velocity control. These motors are losing popularity to brushless motors. The brushes are subject to wear, which increases maintenance costs. In addition, the use of brushes increases resistance, and lowers the motors efficiency.

ASIDE: The controller to drive a servo motor normally uses a Pulse Width Modulated (PWM) signal. As shown below the signal produces an effective voltage that is relative to the time that the signal is on. The percentage of time that the signal is on is called the *duty cycle*. When the voltage is on all the time the effective voltage delivered is the maximum voltage. So, if the voltage is only on half the time, the effective voltage is half the maximum voltage. This method is popular because it con produce a variable effective voltage efficiently. The frequency of these waves is normally above 20KHz, above the range of human hearing.

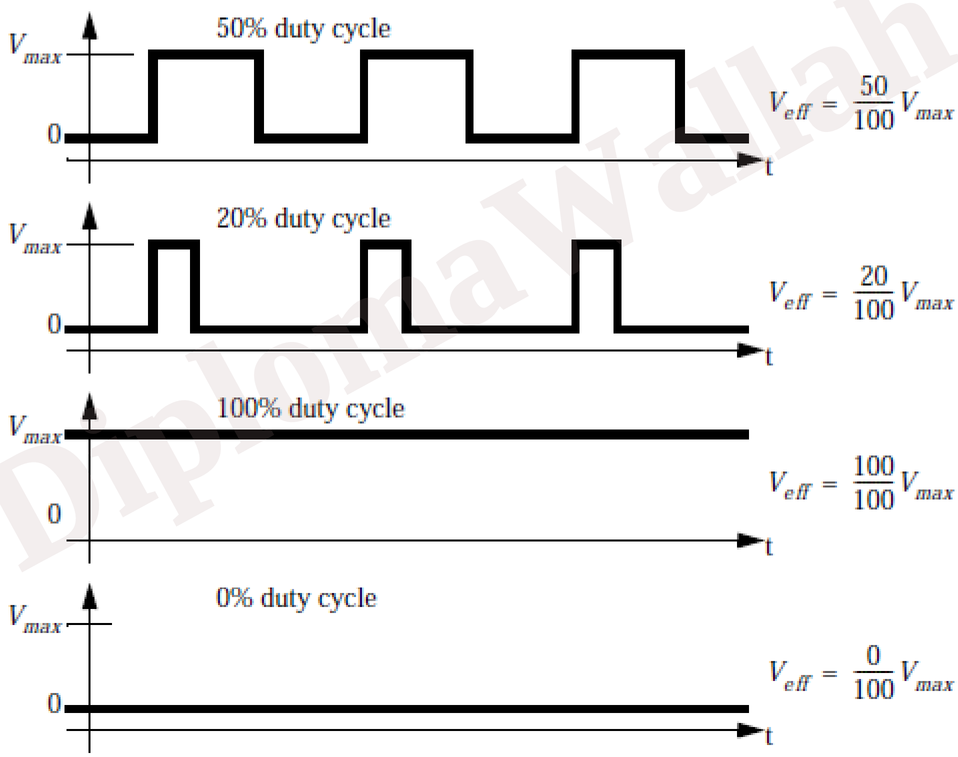


Figure 4.29: Pulse width modulation (PWM) for control

ASIDE: A PWM signal can be used to drive a motor with the circuit shown below. The PWM signal switches the NPN transistor, thus switching power to the motor. In this case the voltage polarity on the motor will always be the same direction, so the motor may only turn in one direction.

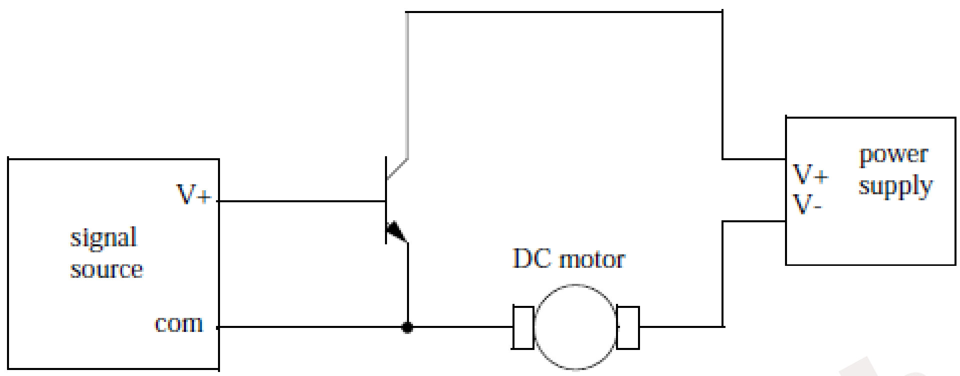


Figure 4.30: PWM unidirectional motor control circuit

ASIDE: When a motor is to be controlled with PWM in two directions the H-bridge circuit (shown below) is a popular choice. These can be built with individual components, or purchased as integrated circuits for smaller motors. To turn the motor in one direction the PWM signal is applied to the Va inputs, while the Vb inputs are held low. In this arrangement the positive voltage is at the left side of the motor. To reverse the direction the PWM signal is applied to the Vb inputs, while the Va inputs are held low. This applies the positive voltage to the right side of the motor.

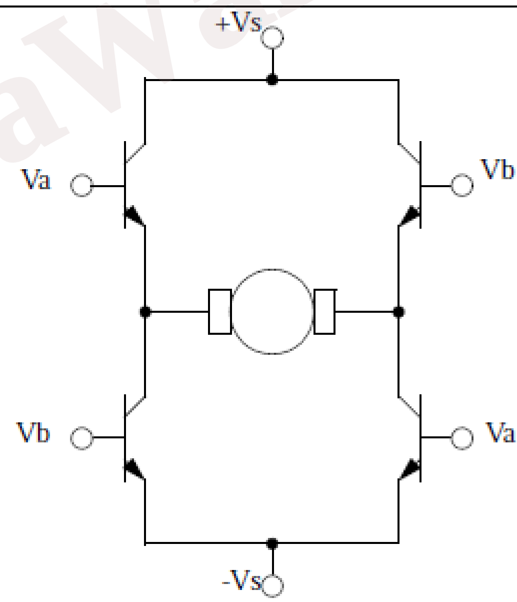


Figure 4.31: PWM Bidirectional motor control circuit

4.2.2 Apply control method for Direct Drives Actuator

1. Solenoid Type Devices

Solenoids, see Fig. 4.32, is the simplest electromagnetic actuators that are used in linear as well as rotary actuations for valves, switches, and relays. As the name indicates, a solenoid consists of a stationary iron frame (stator), a coil (solenoid), and a ferromagnetic plunger (armature) in the center of the coil, see Fig. 4.33.

As the coil is energized, a magnetic field is induced inside the coil. The movable plunger moves to increase the flux linkage by closing the air gap between the plunger and the stationary frame. The magnetic force generated is approximately proportional to the square of the applied current *I* and is inverse proportional to the square of the air gap δ, which is the stroke of the solenoid, i.e.,

$$F \propto \frac{i^2}{\delta^2}$$


Figure 4.32: Assorts of solenoid actuators

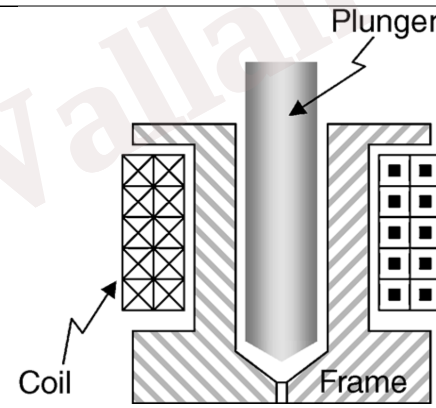


Figure: 4.33: A typical solenoid

As shown in Fig. 4.34, for strokes less than 0.060 in., the flat face plunger is recommended with a pull or push force three to five times greater than 60° plungers. For longer strokes up to 0.750 in., the 60° plunger offers the greatest advantage over the flat face plunger. When the coil is de-energized, the field decreases and the plunger will return to the original location either by the load itself or through a return spring.

All linear solenoids basically pull the plunger into the coil when energized. Push-type solenoids are implemented by extending the plunger through a hole in the back-stop, see Fig. 4.35. Therefore, when energized, the plunger is still pulled into the coil, but the extended producing a pushing motion from the back end of the solenoid. Return motion, upon de-energizing the coil, is provided by the load itself (i.e., the weight of the load) and/or by a return spring, which can be provided as an integral part of the solenoid assembly.

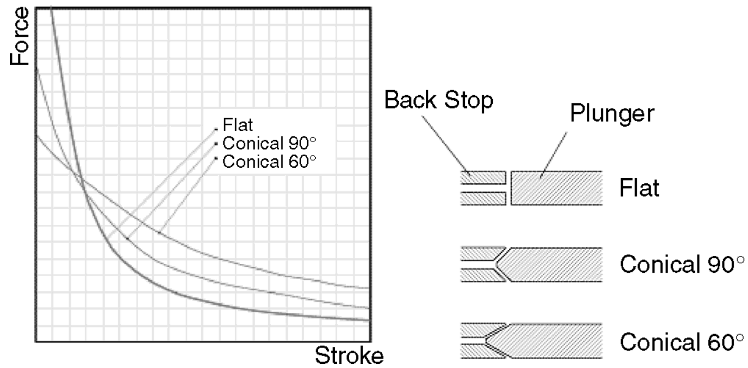


FIGURE 4.34: Typical force-stroke curve of solenoids. (Courtesy of Magnetic Sensor Systems.)

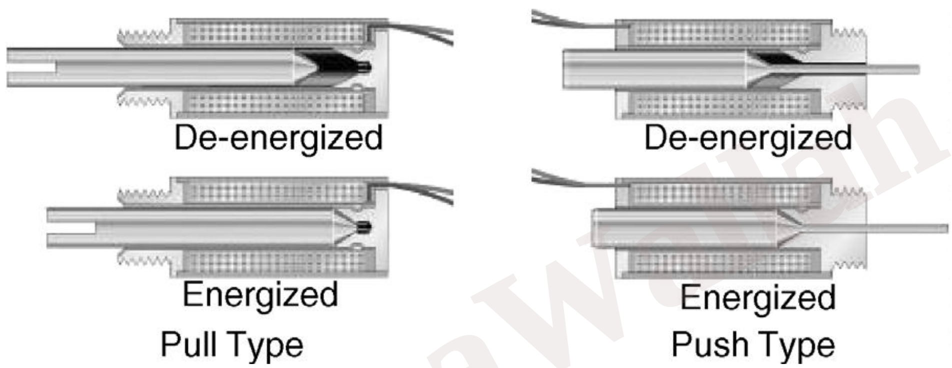


FIGURE 4.35: Push and pull type solenoids. (Courtesy of Ledex® & Dormeyer® Products.)

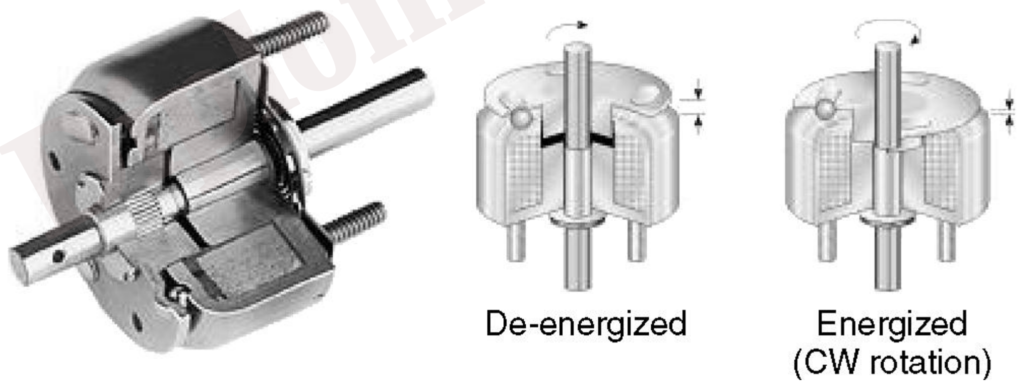


FIGURE 4.36: Rotary solenoid. (Courtesy of Ledex® & Dormeyer® Products)

Table 4.5: Temperature Rating for Electrical Insulations

Insulation Classification		Temperature Rating	
Class A	Class 105	105°C	221°F
Class E	Class 120	120°C	248°F
Class B	Class 130	130°C	266°F
Class F	Class 155	155°C	311°F
Class H	Class 180	180°C	356°F
Class N	Class 200	200°C	392°F

Rotary solenoids utilize ball bearings that travel down inclined raceways to convert linear motion to rotary motion. When the coil is energized, the plunger assembly is pulled towards the stator and rotated through an arc determined by the coining of the raceways, see Fig. 4.36. An *electromechanical relay* (*EMR*) is a device that utilizes a solenoid to close or open a mechanical contact (switch) between high power electrical leads. A relay performs the same function as a power transistor in that relatively small electrical energy is used to switch a large amount of currents. The difference is that a relay has the capability of controlling much larger current level. Variations on this mechanism are possible: some relays have multiple contacts, some are encapsulated, some have built-in circuits that delay contact closure after actuation, and some, as in early telephone circuits, advance through a series of positions step by step, as they are energized and de-energized.

Design/Selection Considerations. Force, stroke, temperature, and duty cycle are the four major design/ selection considerations for solenoids. A linear solenoid can provide up to 30 lb of force from a unit less than in. long. A rotary solenoid can provide well over 100 lb of torque from a unit also less than in. long. As shown in Fig. 4.34, the relationship between force and stroke can be modified by changing the design of some internal components. Higher performance, e.g., force output, can be achieved by increasing the current to the coil winding. However, higher current tends to increase the winding temperature. As the winding temperature increases, the wire resistance increases. This will reduce the output force level. Solenoids are often rated as operating under continuous duty cycle or intermittent duty cycle. A solenoid rated for 100% duty cycle may be energized at its rated voltage continuously because its total coil temperature will not exceed maximum allowable ratings, while an intermittent duty cycle solenoid has an associated allowable “on” time which must not be exceeded. Intermittent duty coils provide considerably higher forces than continuous duty solenoids. The maximum operating temperature for a solenoid is determined by the rated temperature of the insulation material used in the winding (see Table 4.5).

2. Voice-Coil Motors (VCMs)

As the name indicates, the voice-coil motor was originally developed for loudspeakers. It is now extensively used in moving read/write heads in hard disk drives. Since the coil is in motion, VCM is also referred to as a *moving-coil* actuator. The VCM consists of a moving coil (armature) in a gap and a permanent magnet (stator) that provides the magnetic field in the gap, see Fig. 20.14.

$$\vec{F} = \vec{i} \times \vec{B}$$

Since most voice-coils are designed so that the flux is perpendicular to the current direction, the resultant Lorentz force can be written as

$$F_{\mathrm{VCM}} = \gamma BNl \cdot i = K_F \cdot i \Rightarrow F_{\mathrm{VCM}} \propto i$$

where l is the coil length per turn, B is the flux density, N is the number of turns in the coil, i is the current, and γ is a coil utilization factor. It is important to know that the force is proportional to the applied current amplitude and the proportional constant K_F is often called the *force constant*.

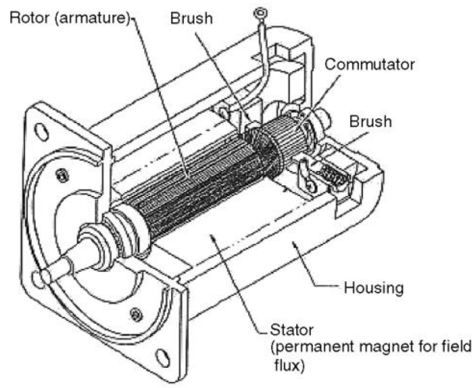


FIGURE 4.37: Permanent magnet DC motor. (T. Keujo and S. Nagamori, *Permanent-Magnet and Brushless DC Motors,* 1985, by permission of Oxford University Press.)

The coil is usually suspended in the gap by springs and attached to the load such as the diaphragm of an audio speaker, the spool of a hydraulic valve, or the read/write head of the disk drive. The linear relationship between the output force and the applied current and the bidirectional capability makes the voice coil more attractive than solenoids. However, since the controlled output of the voice coil is force, some type of closed loop control or some type of spring suspension is needed.

Design/Selection Consideration. From Eq. (20.10) we see that the force constant depends on the flux density and the amount of wires that can be packed into the gap. There are two options to increase the force constant. One is to increase the flux density, which can be achieved by using stronger magnetic material and the other is to increase either *N* or *l*, i.e., to pack more turns and/or make a larger diameter coil. Given a fixed gap volume, using higher gauge (thinner) wires is the only way to increase the number of turns. However, higher gauge wires have larger resistance, which will increase the resistive heating of the winding and limit the allowable current. In addition, the additional insulation will also occupy more volume and tends to reduce the effect of increasing *N*. In summary, to improve the performance of the voice coil, a designer can either choose a better magnetic material or to make the motor bigger by either making the coil wider (increase *D*) or longer (increase *N*).

3. Electric Motors

Electric motors are the most widely used electromechanical actuators. They can either be classified based on functionality or electromagnetic characteristics. The differences in electric motors are mainly in the rotor design and the method of generating the magnetic field. Figure 20.17 shows the composition of a permanent magnet DC motor. Some common terminologies for electric motors are:

Stator is the stationary outer or inner housing of the motor that supports the material that generates the appropriate stator magnetic field. It can be made of permanent magnet or coil windings.
Field coil (*system*) is the portion of the stator that is responsible for generating the stator (field) magnetic flux.
Rotor is the rotating part of the motor. Depending on the construction, it can be a permanent magnet or a ferromagnetic core with coil windings (armature) to provide the appropriate armature field to interact with the stator field to create the torque.
Armature is the rotor winding that carries current and induces a rotor magnetic field.
Air gap is the small gap between the rotor and the stator, where the two magnetic fields interact and generate the output torque.
Brush is the part of a DC motor through which the current is supplied to the armature (rotor). For synchronous AC motors, this is done by *slip rings*.

Commutator is the part of the DC motor rotor that is in contact with the brushes and is used for controlling the armature current direction. Commutation can be interpreted as the method to control the current directions in the stator and/or the armature coils so that a desired relative stator and rotor magnetic flux direction is maintained. For AC motors, commutation is done by the AC applied current as well as the design of the winding geometry. For stepping motors and brushless DC (BLDC) motors, commutations are done in the drive electronics and/or motor commands.
Torque generation in an electric motor is either through the interaction of the armature current and the stator magnetic field (Lorentz Law) or through the interaction of the stator field and the armature field. Table 4.7 summarizes the common classification of electric motors. The next chapter will give a detailed discussion of the operation of various electric motors and the associated design considerations.

Table 4.7: Electric Motor Classification

Classification			
Command Input	Magnetic Field		Description
DC motors	Permanent magnet		Permanent magnets are used to generate the stator magnetic field. Electrical current is supplied directly into the armature winding of the rotor through the brushes and commutators.
	Electro-magnets	Shunt wound	A stator (field) winding is used as electromagnet. Stator winding is connected in parallel with the armature winding.
		Series wound	A stator (field) winding is used as electromagnet. Stator winding is connected in series with the armature winding.
		Compound wound	Two stator (field) windings are used as electromagnet. The stator windings are connected, one in series and one in parallel, with the armature winding.
		Separate wound	A stator (field) winding is used as electromagnet. Both the stator and armature fields are individually energized.
AC motors	Single-phase	Induction	Single stator winding with squirrel-cage rotor. No external connection to the rotor. Torque generation is based on the electromagnetic induction between the stator and rotor. AC current provides the commutation of the fields. Rotor speed is slightly slower than the rotating stator field (slip).
		Synchronous	Permanent magnet rotor or rotor winding with slip ring commutation. Rotating speed is synchronized with the frequency of the AC source.
	Poly-phase	Induction	Similar to single-phase induction motor but with multiple stator windings. Self-starting.
		Synchronous	Similar to single-phase synchronous motor but with multiple stator windings for smoother operation.
	Universal		Essentially a single-phase AC induction motor with similar electrical connection as a *series wound* DC motor. Can be driven by either AC or DC source.
Stepper Motors	Permanent magnet		Permanent magnet rotor with stator windings to provide matching magnetic field. By applying different sequence (polarity) of coil current, the rotor PM field will align to match induced stator field.
	Variable reluctance		Teethed ferromagnetic rotor with stator windings. Rotor motion is the result of the minimization of the magnetic reluctance between the rotor and stator poles.
	Hybrid		Multi-toothed rotor with stator winding. The rotor consists of two identical teethed ferromagnetic armatures sandwiching a permanent magnetic.
Brushless DC motors	Poly-phase	Synchronous	Essentially a poly-phased AC synchronous motor but using electronic commutation to match rotor and stator magnetic fields. Electronic commutation enables using a DC source to drive the synchronous motor.

4. Electrostatics—Electrical Field

Since electrical fields have lower energy density than magnetic fields, typical applications of electrical field forces are limited to measurement devices and accelerating charge particles, where the required energy density is small. Recently, with the proliferation of micro-fabrication technology, it is possible to apply the small electrostatic forces to microelectromechanical actuators, such as comb actuators (see Fig. 4.38).

The advantage of electrostatic actuation is the higher switching rate and less energy loss as compared to the electromagnetic actuation. However, the limitation in force, travel, and high operating voltage still needs to be addressed. Electrostatic actuation is the main actuation for moving charged toner particles in electrophotographic (xerographic) processes, e.g., laser printers.

FIGURE 4.38: MEMS comb actuator uses electrostatic actuation. (Courtesy of Sandia National Laboratories)

5. Piezoelectric

Piezoelectric is the property of certain crystals that produces a voltage when subjected to mechanical deformation, or undergoes mechanical deformation when subjected to a voltage. When a piezoelectric material is under mechanical stress, it produces an asymmetric displacement in the crystal structure and in the charge center of the affected crystal ions. The result is charge separation. An electric potential proportional to the mechanical strain can be measured. This is called the *direct piezoelectric effect.* Conversely, the material will have deformation without volume change when electric potential is applied. This *reciprocal piezoelectric effect* can be used to produce mechanical actuation. There are two categories of piezoelectric materials: sintered ceramics, such as lead-zirconate-titinate (PZT), and polymers, such as polyvinylidence fluoride (PVDF). Piezoceramics have a larger force output and are used more as actuators. PVDFs tend to generate larger deformation and are used more for sensor applications.

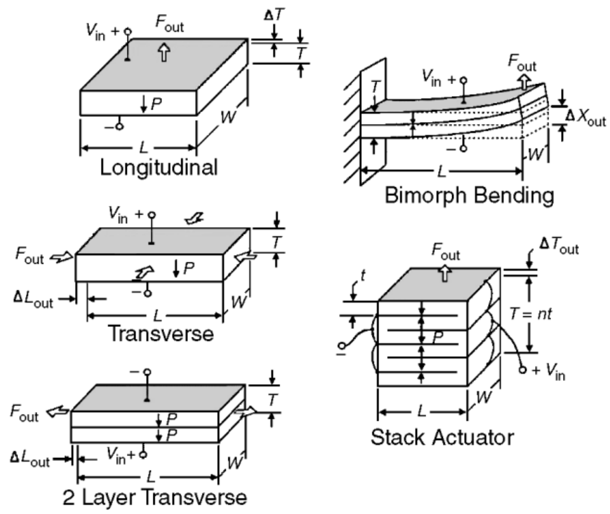


FIGURE 4.39: Common piezoelectric actuation geometries. (Courtesy of Piezo Systems, Inc.)

Design/Selection Considerations

Figure 4.39 shows the common orientation for piezoelectric actuation. With a typical strain of less than 0.3%, the amount of deflection or deformation is usually the limited factor for piezoelectric actuators. The most common architectures are the stacked and bending actuation. Piezoelectric actuators are most suited for high bandwidth, large force, and small stroke/deflection applications. They are widely used in noise and acoustical applications, as well as optical applications, where precision motion is critical.

Piezoelectric actuators are usually specified in terms of their free deflection and blocked force. Free deflection (X_f) refers to displacement attained at the maximum recommended voltage level when the actuator is completely free to move and is not asked to exert any force. Blocked force (F_b) refers to the force exerted at the maximum recommended voltage level when the actuator is totally blocked and not allowed to move.

Figure 4.40 shows the static performance curve of a typical piezoelectric actuator (force vs. deflection). Generally, a piezo actuator must deform a specified amount and exert a specified force, which determines its operating point on the force vs. deflection line. An actuator is considered optimized for a particular application if it delivers the required force at one half its free deflections. High operating voltage, hysteresis, creep, and fatigue are the main mechanical design considerations.

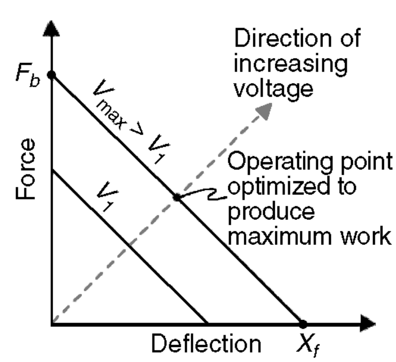


FIGURE 4.40: Static performance curve of a typical piezoelectric actuator. (Courtesy of Piezo Systems, Inc.)

4.2.3 Apply control method for Hydraulic Actuators

An actuation system, which is part of an automatic machine, consists of a power part and a control part as illustrated in Fig. 4.41. The power part comprises all the devices for effecting the movements or actions. The control part provides for the processing of the information and generates the automated cycle and the laws of variation of the reference signals, in accordance with the governing procedures implemented and with the enabling and feedback signals arriving from the sensors deployed on the operative part. The order signals coming from the control part are sent to the operative part by means of the interface devices which convert and amplify the signals, where necessary, so that they can be used directly by the actuators. These interfaces can be the speed drives or the contactors of the electric motors, the distributor valves in hydraulic and pneumatic actuators.

Figure 4.42 illustrates a fluid actuation system. The power part consists of the actuator—a double acting cylinder in the case in the figure—the front and rear chambers of which are fed by a 4/2 distributor valve, which constitutes the fluid power adjustment interface. The valve switching command is the order from the control part. This order is sent in accordance with the movement strategy, determined by the desired operating cycle of the cylinder in the control part, on the basis of the feedback signals from the sensors in the cylinder, represented in the figure by the limit switches.

Then there are discontinuous actuation systems and continuous actuation systems, depending on the type of automation realized, while retaining the control part and the actuation part. The first are effective when used in discontinuous automation, typical of assembly lines and lines for the alternating handling of machine parts or components; on the other hand, continuous actuation systems are found in continuous process plants and as continuous or analog control devices for the desired magnitudes, and constitute fluid servosystems.

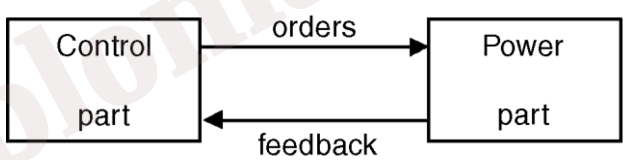


FIGURE 4.41 Actuation system.

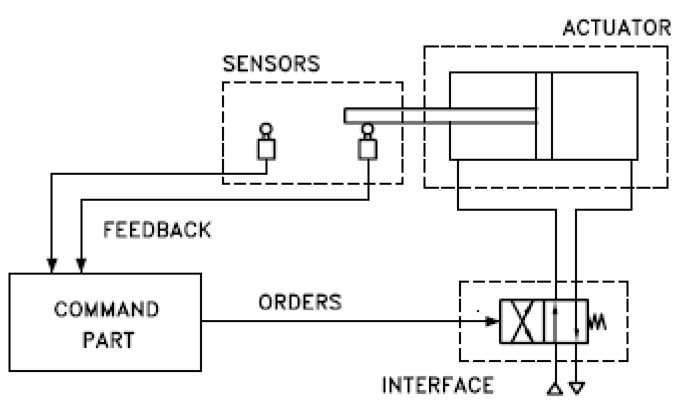


FIGURE 4.42 Fluid power actuation system.

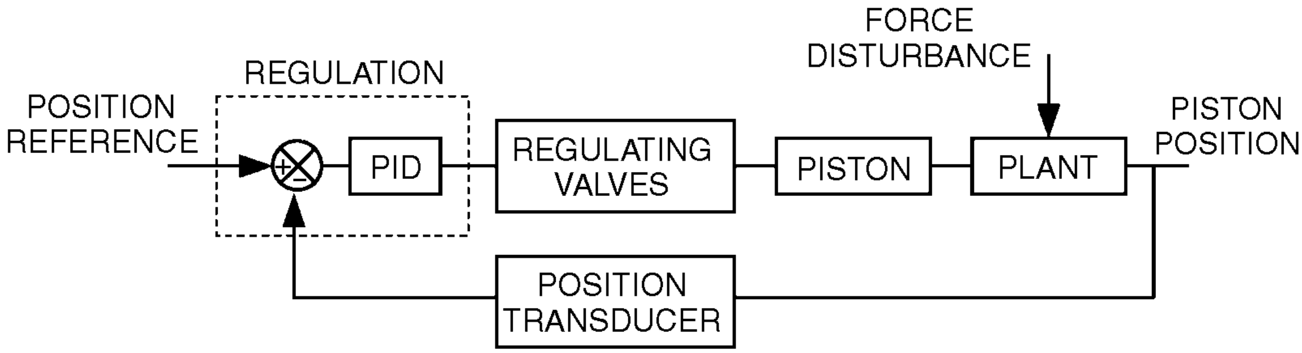


FIGURE 4.43 Scheme of a fluid power servosystem.

Fluid actuators, whether they are linear (cylinders) or rotary (motors) are continuous systems as they can determine the positioning of the mobile component (of the rod with respect to the cylinder liner; of the shaft with respect to the motor casing) at any point in the stroke. Performance of the usual cylinders and motors is currently highly influenced by the action of friction (static and dynamic) developed by contacts between mobile parts. This action, in pneumatic systems in particular, gives rise to the well-known phenomenon of stick-slip, or intermittent motion at very low movement speeds, due to the alternation of conditions of friction and adherence in the motion of the mobile element in the actuator.
Given the nature of the friction itself, the presence of devices suitable for sustaining the mobile components of the actuator and maintaining the correct pressure conditions, such as supports and gaskets, gives rise to nonlinear conditions in the equilibrium of the actuator, increasing the level of difficulty in obtaining high precision in positioning the system. To overcome these problems in specific applications it is necessary to use actuators without seals, for example, with fluid static and/or fluid dynamic bearings.

The interface element, indicated as a distributor in the figure, takes on a crucial role in the definition of the operating mode of the actuator. Indeed, in the case in which it is only necessary to create reciprocating movements, with positioning of the actuator at the end of its stroke, it is only necessary to use a two- or three-position distributor valve, with digital operation. This is the solution shown in Fig. 4.42.

If, on the other hand, it is necessary to have continuous control of the position and force transmitted, it is necessary to use devices which are not digital now, but which are continuous, such as proportional valves and servovalves, or it is necessary to use digital devices operating with control signal modulation, for example those of the PWM (Pulse Width Modulation) type. The actuation system therefore becomes a fluid servosystem, such as the one outlined in Fig. 4.43, for example. A practical construction of a hydraulic linear servoactuator having the same working scheme of Fig. 4.43 is shown in Fig. 4.44. It consists of a cylinder, a valve, and a position transducer integrated in a single device.

A controlled, fluid-actuated system is a classical mechatronic system, as it combines mechanical and fluid components, and control and sensing devices, and normally requires a simulation period for defining the size and characteristics of the various elements so as to comply with the desired specifications. The standardized symbols for the different components of hydraulic and pneumatic fluid systems, and the definitions of the associated circuits, are defined in the standard, ISO 1219 "Fluid power systems and components—Graphic symbols and circuit diagrams; Part 1: Graphics symbols, Part 2: Circuit diagrams."

FIGURE 4.44 Hydraulic servocylinder (Hanchen).

4.2.3.1 Hydraulic Actuation Systems

The components of a hydraulic actuation system are:

- the pump, that is, the hydraulic power generation system;
- the actuator, that is, the element which converts hydraulic power into mechanical power;
- the valve, that is, the hydraulic power regulator;
- the pipes for connecting the various components of the actuation system;
- the filters, accumulators, and reservoirs;
- the fluid, which transfers the power between the various circuit elements;
- the sensors and transducers;
- the system display, measurement, and control devices.

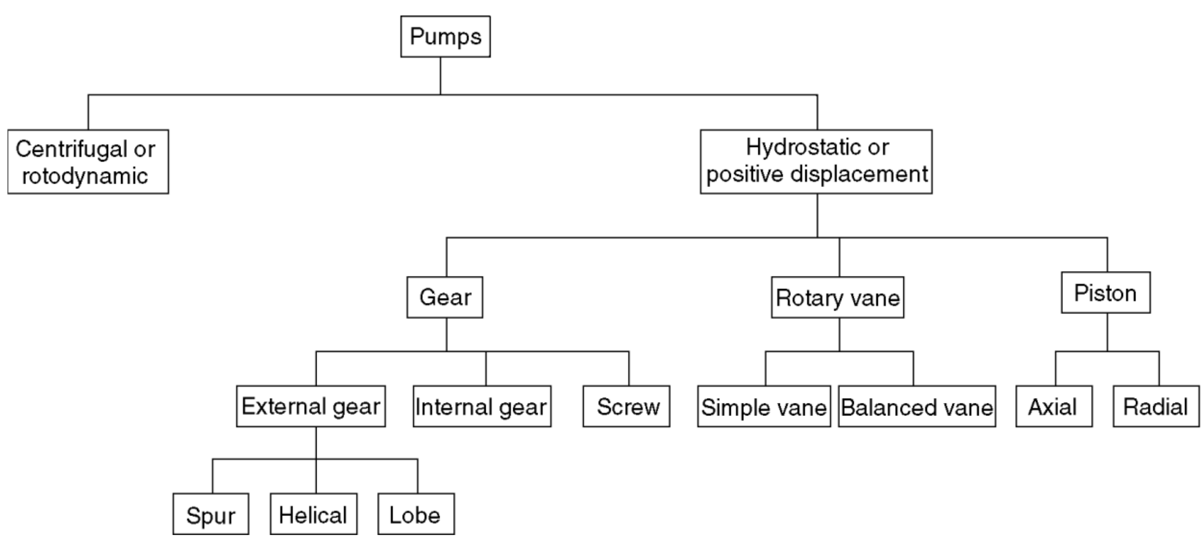


FIGURE 4.45 Pumps classification.

Pumps

Pumps transform electrical or mechanical energy into hydraulic energy. They constitute the fluid flow generator of the hydraulic system, as the pressure is determined by the fluid resistance downstream from the generator. The main types of pumps are shown in Fig. 4.45.

Centrifugal pumps permit high deliveries with low pressures. They do not have internal valves but have a large clearance between the rotary part and stator part and guarantee a sufficiently stationary flow. Vice versa, hydrostatic or positive displacement pumps, which are those most commonly used, guarantee high pressures with limited deliveries. They have elements such as valves and caps, which permit separation of the delivery zone from the intake zone, and they may introduce pulses in the flow in the delivery line and generally require the use of a fluid with sufficient lubricating properties and load capacity, so as to reduce the friction between the sliding parts of the pump. There are constant displacement and variable displacement pumps. The main positive displacement pumps belong to the gear, rotary vane, and piston types.

Gear Pumps

Gear pumps are subdivided into pumps with external gears, pumps with internal gears, and screw pumps. In all cases, the pump is made up of two toothed wheels inserted into a casing with little slack so as to minimize leakage. Figure 4.46 is a photograph of a pump with external gears. The opposed rotation of the wheels causes the transfer of the oil trapped in the space between the teeth and walls of the gear from the intake to the outlet. Depending on the form of the teeth, there are external gear pumps of the spur gear, helical gear, and lobe gear types.

Pumps with internal gears are functionally similar to the above, but in this case the gears rotate in the same direction. Figure 4.47 is a section plane of a two-stage pump. In screw pumps, which may have one or more rotors, the elements have helical toothing similar to a threaded worm screw. Transfer of the fluid takes place in an axial direction following rotation of the screw. These types of pump guarantee very smooth transfer of the flow, with reduced pulsation and low noise levels.

The usual rotation speeds are between 1000 and 3000 rpm, with powers between 1 and 100 kW.

Delivery pressures can reach 250 bar, with higher values in the case of the pumps with external gears.

The flow transferred is a function of the pump displacement and the angular input speed, with values comprised between 0.1 and 1000 cm2/rev. Double pumps can be used to increase these values. Gear pumps have high performance levels, with values around 90%.

Rotary Vane Pumps

Vane pumps (Fig. 4.48) generally consist of a stator and a rotor, which can rotate eccentrically with respect to one another. Vanes can move in special slits placed radially in the stator or in the rotor and delimit appropriate variable volumes. In Fig. 4.48, as in most constructions, the vanes are borne by the rotor which can rotate inside the stator. Rotation leads to the displacement of volumes of fluid enclosed between two consecutive vanes from the intake environment to input into the delivery environment. This type of pump permits a range of working pressures up to 100 bar and, compared with gear pumps, guarantees lower pulsing of the delivery flow and greater silence.

FIGURE 4.46 External spur gear pump (Casappa).

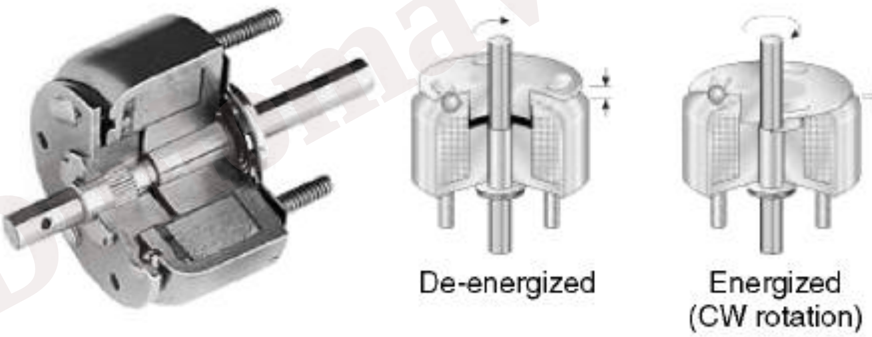


FIGURE 4.47 Internal gear pump (Truninger).

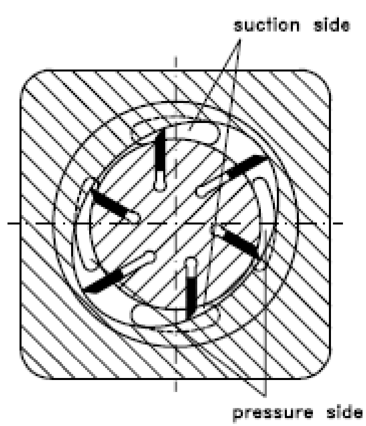


FIGURE 4.48 Rotary vane pump.

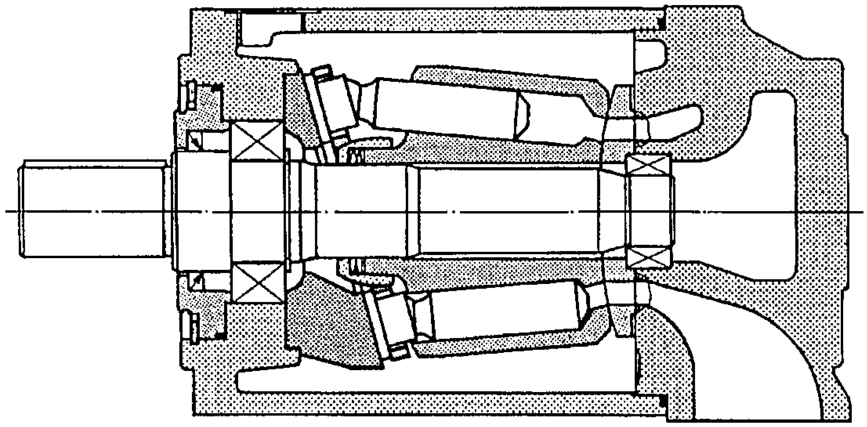


FIGURE 4.49 Axial piston swash plate pump (Bosch Rexroth).

Piston Pumps

Volumetric piston pumps can have one or more cylinders; that is, there may be one or more cylinders with a piston sliding in each of them. Transfer of the volume of fluid from intake to delivery is determined by the displacement of the piston inside the cylinder, which is provided with input and output valves or shutters. Depending on the geometrical arrangement of the cylinders with respect to the rotating motor shaft, piston pumps are subdivided into axial pumps (bent axis type and swash plate type) and radial pumps. Figure 4.49 shows the plan of a fixed-displacement axial piston pump, of the swash plate type. The working pressure range available with the aid of piston pumps is greater than in the previous cases, being able to reach pressures in the order of 400–500 bar but with the disadvantage of more uneven flow.

Motion Actuators

Motion actuators convert the hydraulic energy of the liquid under pressure into mechanical energy. These actuators are therefore volumetric hydraulic motors and are distinguished, on the basis of the type of movement generated, similar to what has been said about pumps, into rotary motors, semi-rotary motors or oscillating ones, which produce limited rotation by the output shaft, and into linear reciprocating motors, that is hydraulic cylinders.

Rotary and Semi-rotary Motors

In construction terms, rotary motors are identical to rotary pumps. Therefore gear, vane, and piston motors, radial or axial, are available. Obviously, the operating principle is the opposite of what has been said for pumps. The symbols of hydraulic rotary motors are shown in Fig. 4.50. Semi-rotary motors generate the oscillating motion either directly, by means of the rotation of a vane connected to the output shaft, or indirectly, by coupling with a rack, driven by a piston, with a toothed wheel connected to the output shaft, as in the example in Fig. 4.51. The semi-rotary vane motors produce high instantaneous torsional torque on the output shaft; for this reason they are also called hydraulic torque-motors.

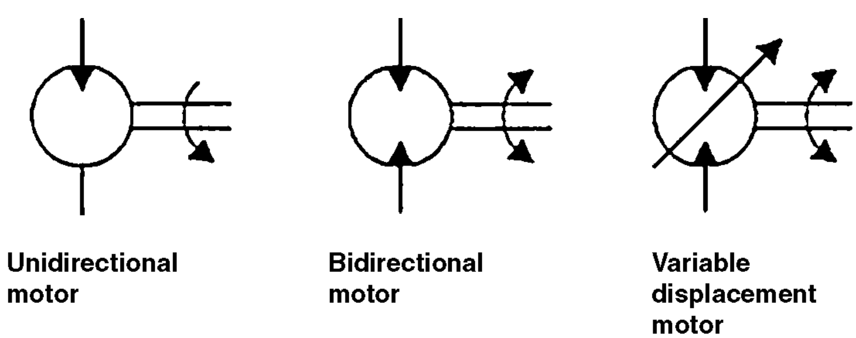


FIGURE 4.50 Symbols of hydraulic rotary motors.

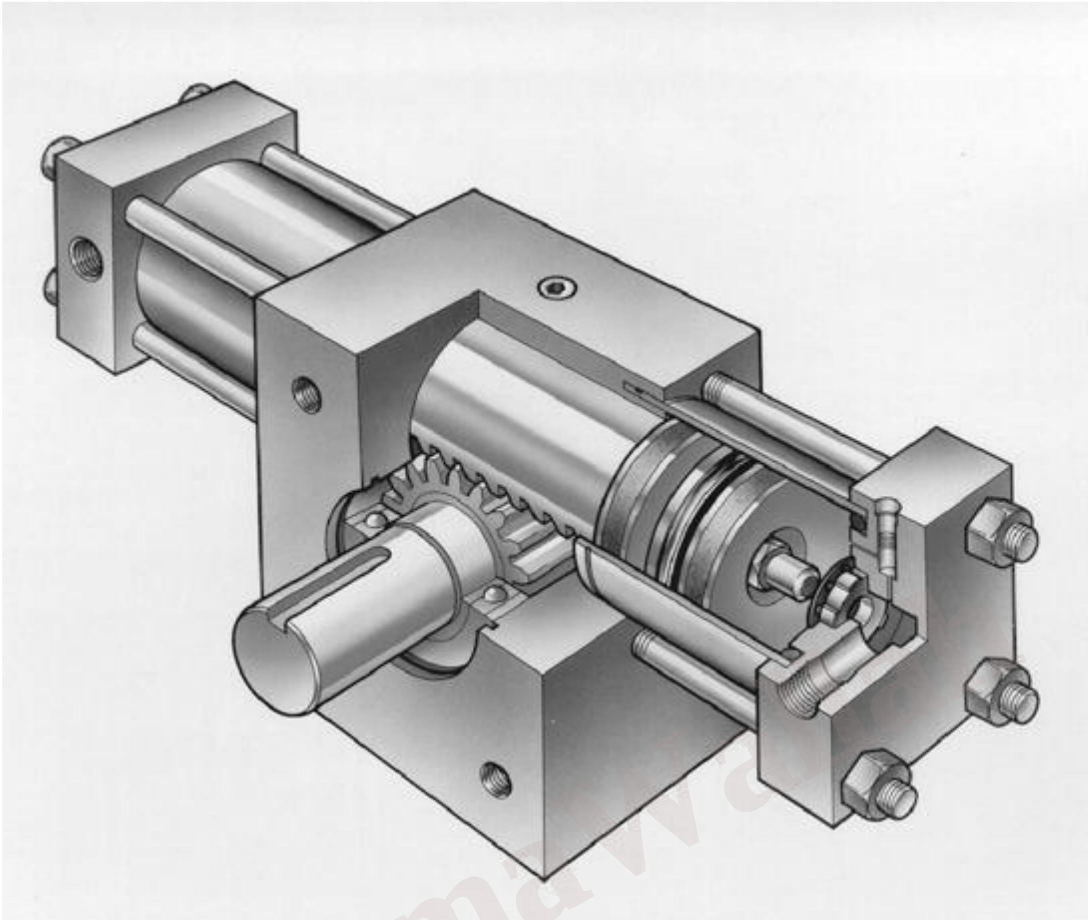


FIGURE 4.51: Hydraulic rotary actuator (Parker Hannifin).

Linear Actuators

Linear hydraulic motors constitute the most commonly used type of actuator. They provide a rectilinear movement realized by the stroke of a rod connected to a piston sliding inside the cylinder. A distinction is made between single acting and double acting cylinders. The former only permit a single work stroke and therefore the pressure of the fluid is exerted on the surface of the piston in one single direction; the retract stroke is made by means of the force applied externally to the cylinder rod, or with the aid of a helical spring incorporated with the actuator inside a chamber. The latter permit both strokes, so that the fluid acts alternately on both faces of the piston, generating both the advance and retract strokes.

Double acting cylinders may have a single rod or a double through rod. These are composed of a tube closed at the ends by two heads, and a mobile piston inside the barrel bearing one or two rods connected externally to the load to move. As it is fitted with sealing gaskets, the piston divides the cylinder into two chambers. By sending the oil under pressure into one of the chambers through special pipes in the heads, a pressure difference is generated between the two surfaces of the piston and a thrust transmitted to the outside by the rod. Figure 4.52 shows the constructional solution of a hydraulic double acting cylinder with a single rod. Single rod actuators are also known as asymmetrical cylinders because the working area on the rod side is smaller than the area of the piston, as it is reduced by the section of the rod itself.

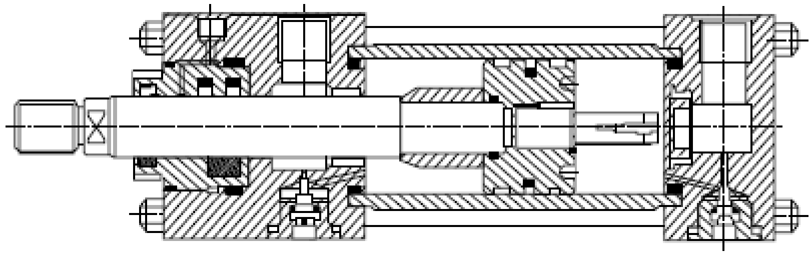


FIGURE 4.52 Single rod double-acting piston actuator (Atos).

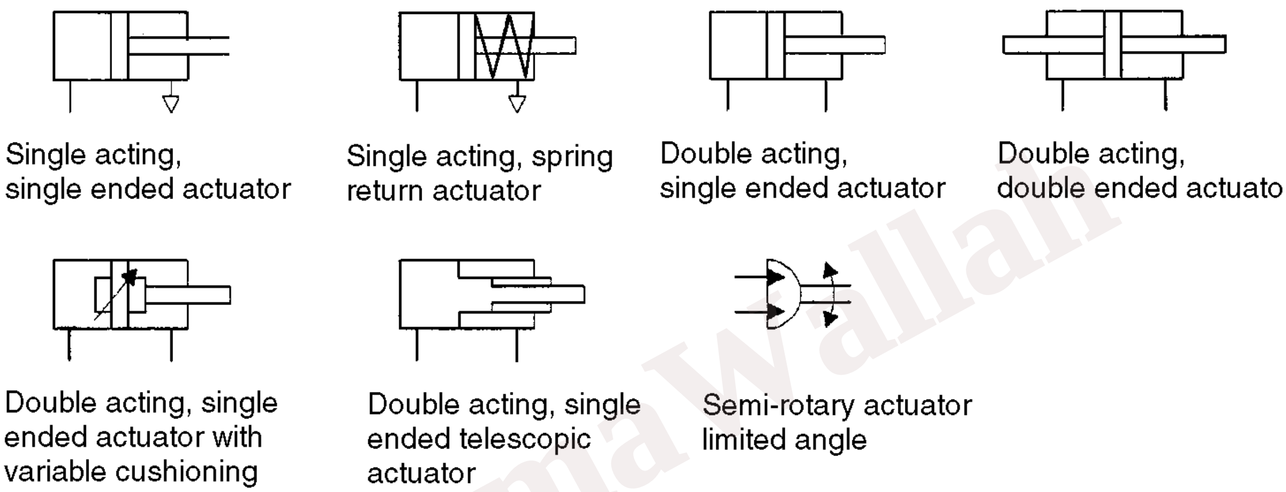


FIGURE 4.53 Actuators symbols.

This involves actuating forces and feed speeds which are different in the two directions, with the same feed pressure in the two thrust chambers. Hydraulic actuators are able to support external overloads, as, if the load exceeds the available thrust force, the rod stops or reverses, but generally does not suffer any damage. Cylinders may get damaged however, or at least suffer a drop in performance, when they have to support loads which are not applied along the axis of the rod, that is, with components in the radial direction, as reactions are generated on the rod supports and piston bearings, which leads to fast wear of the same and reduces the tightness with oil leakage as a result. The main features of a linear actuator are its bore, its stroke, its maximum working pressure, the type of working fluid, and the way its connections are fitted. The symbols of the different types of actuators can be seen in Fig. 4.53.

4.2.4 Apply control method for Pneumatic Actuators

Pneumatic Actuation Systems

Just as described for the hydraulic system, the components of a pneumatic actuation system are:

- the compressed air generation system, consisting of the compressor, the cooler, possibly a dryer,
- the storage tank, and the intake and output filters;
- the compressed air treatment unit, usually consisting of the FRL assembly (filter, pressure regulator,
- and possibly a lubrifier), which permits filtration and local regulation of the supply pressure to
- the actuator valve;
- the valve, that is, the regulator of the pneumatic power;
- the actuator, which converts the pneumatic power into mechanical power;
- the piping;
- the sensors and transducers;
- the system display, physical magnitude measurement, and control devices.

Some of the components of the pneumatic actuation system such as the compressors, treatment units, and some valves used in pneumatic servosystems are described below. The actuators are similar in function and construction to hydraulic ones, though they are built slightly lighter because of the lower working pressure.

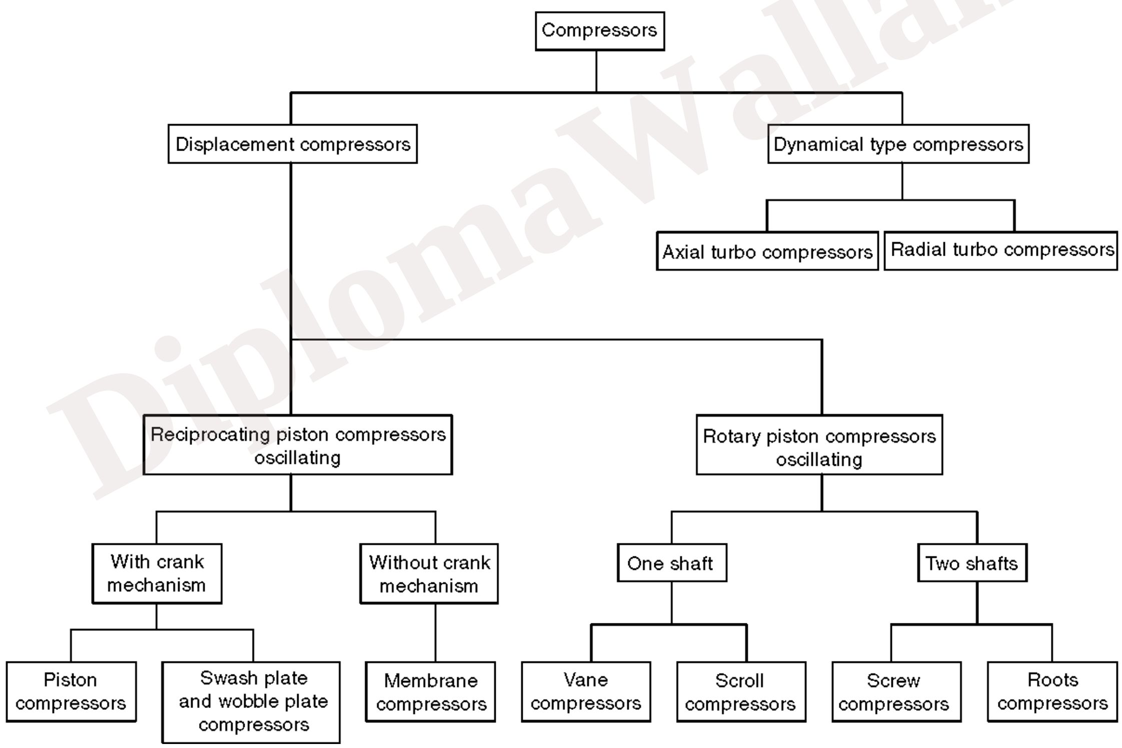


FIGURE 4.54 Classification of pneumatic compressors.

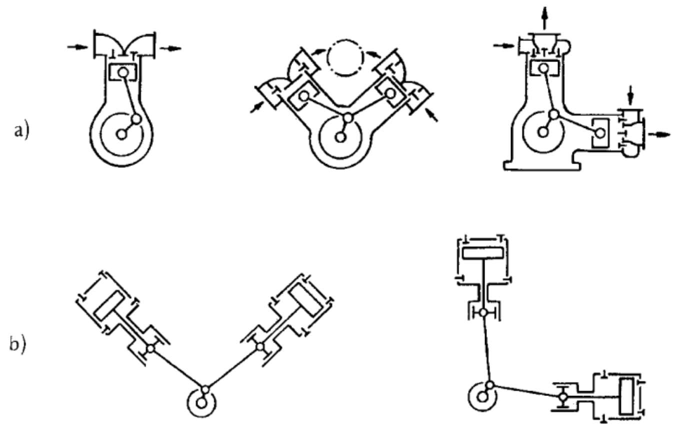


FIGURE 4.55 Piston compressors: (a) single action, (b) double action.

Compressors

The types of compressors used to produce compressed air are summarized in Fig. 4.54. In volumetric compressors, the air or gas is sucked in by means of a valve in the compression chamber where its volume is reduced to cause compression of the gas. Opening of the delivery valve, when a predetermined pressure has been reached, results in the distribution of the air mass to the user.

Vice versa, in dynamic compressors or turbo compressors, the kinetic energy is converted into pressure energy transferred to the gas as a result of the rotary motion of the impeller. Alternating piston compressors determine the compression of the gas as an effect of the motion of the piston, moved by a connecting rod and crank mechanism, inside a gas-tight cylinder. They can be single and double acting, with one or more pistons and one or more stages (Fig. 4.55). They make it possible to obtain pressures of hundreds bar, where there are several stages, and flow rates of thousands of cubic meters per hour, in the case of several cylinders. Vane compressors (Fig. 4.56) have a rotor, fitted eccentrically with respect to the axis of the cylinder in which it rotates, which leads to a certain number of vanes which can move radially with respect to its axis. In the continuous rotation motion of the rotor, the vanes are centrifuged in contact with the seat of the stator, isolating chambers whose volume varies progressively with the angular stroke, guaranteeing input suction on the one hand, and a compressed gas output on the other. The compression pressures are below 15 bar, with maximum flow rates of 500 m3/h. Compared with reciprocating piston compressors, they have less flow pulsation, fewer vibrations, and are more compact.

Screw compressors have two rotors rotating in opposite directions inside a stator, one with convex lobes and the other with concave lobes. The coupling of the profiles of the two rotors leads to a reduction of the volume during the angular stroke and consequent compression of the gas. With pressures typically below 15 bar, they provide a sufficiently continuous flow, up to values of about 3000 m3/h.
In the same way, Roots compressors, also known as superchargers, are made up of two figure-of eight shaped rotors counter-rotating inside a stator in such a way as to transport volumes of gas from suction to delivery. Their efficiency is low because of the leakage between the rotors themselves and between the lobes and the casing, and they are, therefore, used for low compression pressures, below 2 bar. However, they do permit operation without lubrication, like screw compressors, so that oil-free air can be obtained. Both axial and radial dynamic compressors are used to obtain high compressed air flow rates from a few thousand to 100,000 m3/h.

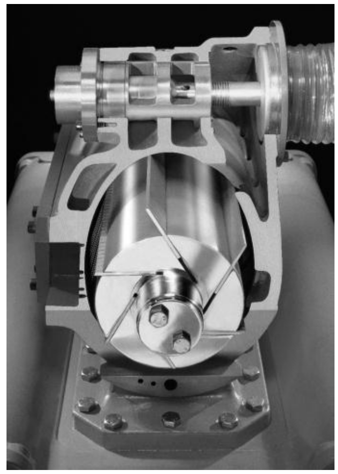


FIGURE 4.56 Rotary vane compressor (Pneumofore).

Compressed Air Treatment Units

Pneumatic supply to a servosystem is generally provided by a local gas treatment unit, consisting of a filter, connected to a compressed gas distribution and generation network, a pressure regulator, and in case a lubricator *L*. Figure 4.57 shows an example of an integrated filter device and pressure regulator. The air first passes through the filter and is filtered by the deflector while the solid and liquid impurities in contact with the walls are deposited on the bottom of the cup, also as an effect of the conical bottom screen, located below the porous cylindrical element in sintered bronze or fabric. The filtered air then flows into the inlet of the pressure regulator, made up of an obturator in equilibrium between pressure.forces. Control of the downstream pressure is determined by the position of the main obturator, which regulates the flow towards the outlet. The passage aperture is closed when the force due to the downstream pressure, acting on a diaphragm and on a translating piston, is in equilibrium with the force of the top spring, the preload of which is set by the rotation of the control knob. Vice versa, if the pressure force is below the desired value, the flow sent to the user tends to compensate for the pressure error with consequent closing of the obturator again when the set point has been reached. The opposite situation occurs if the regulated pressure is above the requested value, so that an aperture passage opens between the user and discharge.

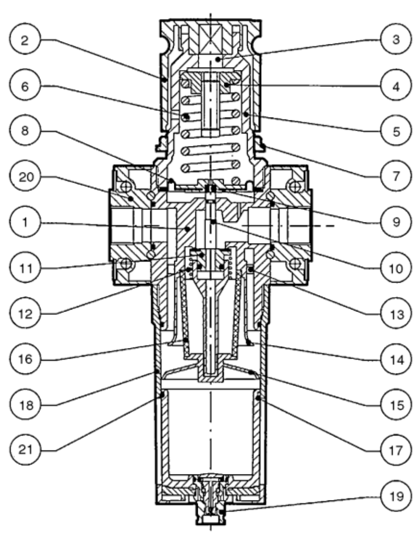


FIGURE 4.57 Pneumatic filter/pressure reducer (Metal Work).

Pneumatic Valves

Pneumatic valves are functionally similar to those used in hydraulic systems, so that reference should be made to the general considerations described above. In particular, this is also valid for the directional valves of the digital and proportional types. Even in pneumatic systems, there are digital spools or poppet two-, three-, or four-way distributors, with two or three working positions, and actuated manually, mechanically, pneumatically, and electrically.

Flow proportional valves are substantially similar to hydraulic ones and are available both with the torque motor electromechanical converter (servovalve), and with servosolenoid acting directly on the spool. As well as these components for controlling the gas flow, digital electrically controlled two- or threeway valves are also used, and their control signals are modulated using PWM (pulse width modulation), PFM (pulse frequency modulation), PCM (pulse code modulation), PNM (pulse number modulation), or a combination of these.

As far as pressure regulation valves are concerned, three-way pressure proportional valves are available for pneumatic actuation which convert an electrical reference signal with standardized input into a controlled output pressure with good dynamics and high precision.

PWM (Pulse Width Modulation) Valves

The structure of PWM valves is similar to the corresponding electrically controlled unistable digital valves, but uses a technique for modulating the width of the pulses sent to the solenoid for supplying proportional control of the flow rate. This technique envisages that the input voltage reference analog signal *V*REF (for example 0–10 V) is converted by a special driver into a digital *V*PWM (ON/OFF) signal with pulse duration proportional to the input signal. Alternatively, the modulated signal can be generated directly by a digital controller, such as a PLC.

Figure 4.58 shows the PWM operating principle. The digital voltage signal sent to the valve solenoid is made up of a pulse train, with constant amplitude, with a constant period T, but with the duration t of every pulse being a linear function of the analog value of the reference voltage. The average valve opening value, and therefore an initial approximation of the generated flow, is a function of the duration t of the pulse, in particular of the duty cycle t/T, and increases as the latter increases.

PWM valves generally do not have any feedback, so that the value of the downstream pressure, and therefore of the flow rate, depends on the type of pneumatic circuit present. Figure 4.59 shows two plans which depict operation of the two-way, two-position valves, with PWM, used as a flow regulator (Fig. 4.59(a)) and as a pressure regulator (Fig. 4.59(b)). In plan a, the valve proportionally controls the flow which transits between the two points at pressure *PS* (feed pressure) and at pressure *PV* (downstream pressure) maintained constant. In this case, this flow is only a function of the aperture of the valve and therefore the proportionality is of the linear type. In plan b, the cross-fitted valves control the pressure *PR*, for example, within a fluid capacity of volume V, respectively regulating the mass flow rate $G1$ and discharge flow $G2$. The time gradient for the controlled pressure *PR* corresponds to the resulting flow G entering the reservoir.

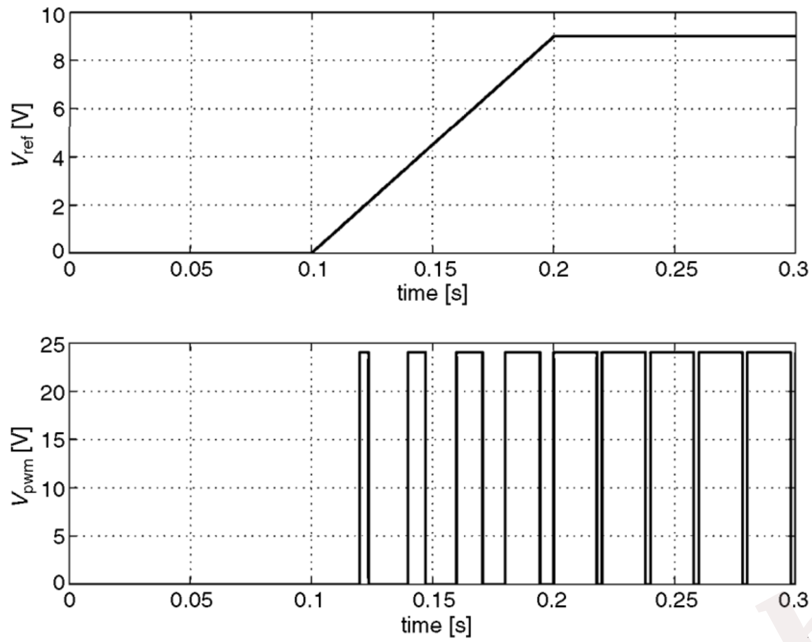


FIGURE 4.58 PWM (pulse width modulation) input and output signals.

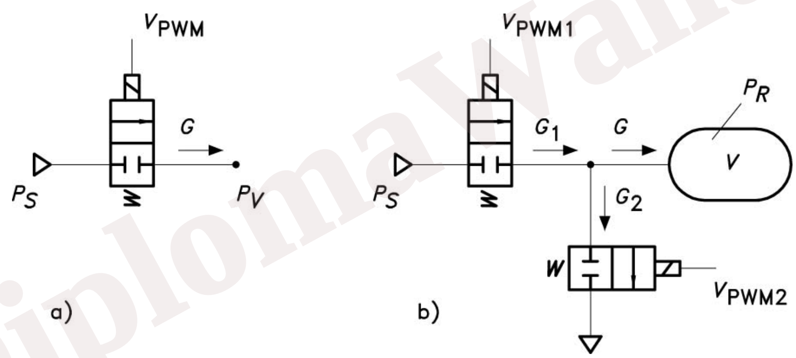


FIGURE 4.59 PWM (pulse width modulation) digital valves: (a) flow regulator, (b) pressure regulator.

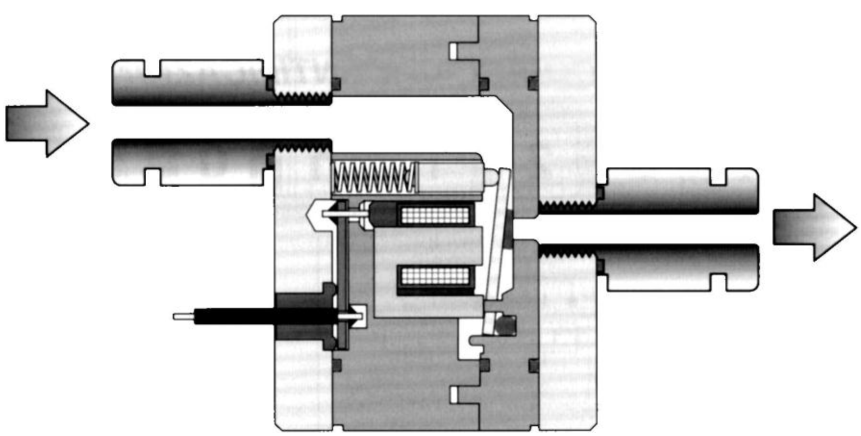


FIGURE 4.60 Two-way digital poppet valve (Matrix).

The parameters affecting the performance of a regulation made by a PWM on/off valve are as follows:

- valve opening and closing times
- dependence on the opening/closing times of the upstream and downstream pressures
- valve size
- period *T* or modulation carrier frequency $f = 1/T$
- working life of the valve

While small opening/closing times and high flow capacity are always antithetical characteristics in an on/off valve, when designing the system it is always necessary to find a compromise between the need for good control resolution and linearity and a high response dynamic.

In pneumatic servosystem applications, typical carrier frequency values $f = 1/T$ range between 20 and 100 Hz, so that valves with opening/closing times of 1 ÷ 5 ms are used. An example of a normally closed 2/2 valve that guarantees minimum opening times (below 1 ms with speed-up command) can be seen in Fig. 4.60. These characteristics are obtained by reducing the mass of the moving parts, with the use of a poppet connected to a small oscillating bar, while practically eliminating the friction between the parts in relative motion.

Proportional Pressure Regulator Valves

These valves are normally three-way, with double poppets or with spool. Poppet valves operate in a similar way to pressure regulator valves. In the same way as with pressure regulators, the poppet which separates the high pressure environment from the regulated pressure one is in equilibrium between the force due to the regulated pressure and that exerted by the action of the control block. The latter can directly be the force of the servo-solenoid armature, or that due to a pressure controlled by the control block which acts on a piston or on a diaphragm linked with the poppet.

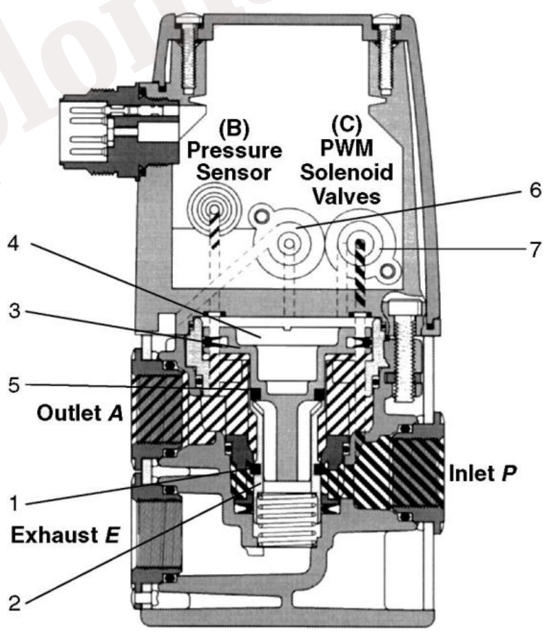


FIGURE 4.61 Pressure proportional valves (Parker).

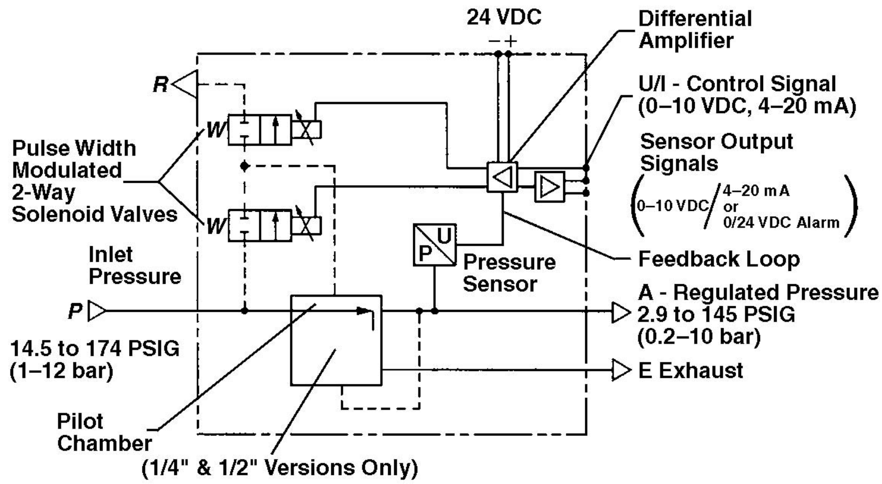


FIGURE 4.62 Pneumatic control scheme of the pressure proportional valve.

Figure 4.61 shows an example of a pressure proportional valve with double poppets. The ports at supply pressure, controlled pressure, and discharge are respectively indicated by *P*, *A*, and *E*. In the position indicated in the figure, the supply poppet 2 is at the top end of its stroke, as the seal 1 is against the fixed seat. In the same way, the regulating poppet 3 is in contact with the poppet 2 by means of the seal 5.

The opening of the feed aperture, between the port P and the port A, is determined by the equilibrium of the forces acting on the piston of poppet 3, in particular the force *FR* of the regulation pressure *PR* in the servochamber 4 directed downwards, and the force *FC* due to the action of the regulated pressure *Pc* on the outlet, directed upwards. If *FR* = *FC*, the moving bodies in the valve are in the positions shown in the figure, so that the chamber at controlled pressure *PC* is isolated both from supply and discharge. If *FR* > *FC*, then the two poppets move downwards and the feed aperture is opened so as to convey the air mass towards the output and rebalance the pressure *PC* at the desired value. In the opposite case, if *FR* < *FC*, the regulating poppet moves upwards, but while remaining at the top end of its stroke, the seal 5 opens and permits the passage of the masses from port A to the exhaust E.

In Fig. 4.61, 6 and 7 indicate the PWM on/off valves, which regulate the pressure *PR* of the servo chamber 4. The pneumatic control plan of the valve is shown in Fig. 4.62. The two 2-way PWM valves receive the modulated control signal from the regulation block. These are fitted in such a way that one controls a flow entering the servo chamber 4 (see Fig. 4.61) while the other controls the flow exiting towards discharge. By means of appropriate action, the control signal is converted into a pressure proportional signal.

Table 4.8: Type of Actuators and Their Features

Actuator			Features
Electrical			
Diodes, thyristor, bipolar transistor, triacs, diacs, power MOSFET, solid state relay, etc.			Electronic type Very high frequency response Low power consumption
Electromechanical			
DC motor	Wound field	Separately excited	Speed can be controlled either by the voltage across the armature winding or by varying the field current
		Shunt	Constant-speed application
		Series	High starting torque, high acceleration torque, high speed with light load
		Compound	Low starting torque, good speed regulation Instability at heavy loads
	Permanent magnet	Conventional PM motor	High efficiency, high peak power, and fast response
		Moving-coil PM motor	Higher efficiency and lower inductance than conventional DC motor
		Torque motor	Designed to run for a long periods in a stalled or a low rpm condition
	Electronic commutation (brushless motor)		Fast response High efficiency, often exceeding 75% Long life, high reliability, no maintenance needed Low radio frequency interference and noise production
AC motor	AC induction motor		The most commonly used motor in industry Simple, rugged, and inexpensive
	AC synchronous motor		Rotor rotates at synchronous speed Very high efficiency over a wide range of speeds and loads Need an additional system to start
	Universal motor		Can operate in DC or AC Very high horsepower per pound ratio Relatively short operating life
Stepper motor	Hybrid		Change electrical pulses into mechanical movement Provide accurate positioning without feedback
	Variable reluctance		Low maintenance
Electromagnetic			
Solenoid type devices			Large force, short duration
Electromagnets, relay			On/off control
Hydraulic and Pneumatic			
Cylinder			Suitable for liner movement
Hydraulic motor	Gear type		Wide speed range
	Vane type		High horsepower output
	Piston type		High degree of reliability
Air motor	Rotary type		No electric shock hazard
	Reciprocating		Low maintenance
Valves	Directional control valves Pressure control valves Process control valves		
Smart Material actuators			
Piezoelectric & Electrostrictive			High frequency with small motion High voltage with low current excitation High resolution

END OF CHAPTER 4

1. Indicate the correct answer:
   a. The major advantage of a permanent magnet step motor is that it can provide holding torque (True/ False).
   b. The torque generated in a step motor under start-stop mode is more than under slewing mode (True/ False).
   c. A variable reluctance type step motor requires less number of switches than of permanent magnet type (True/ False).
   d. Damping of a step motor refers to slow acceleration during starting (True/ False).
   e. 3-phase a.c. excitation is needed to drive a 3-phase step motor (True/ False).

2. What stepper motor has to offer in the field of engineering?

3. Explain how can we identify a stepper motor through physical outlook?

4. Explain the operation principles of a stepper motor.

5. Why are the step positions likely to be less well defined when a motor is operated in 'two-phase-on' mode as compared with one-phase-on mode?

6. Explain how rotation is achieved in a stepper motor.

7. A stepper motor cannot be bench-checked directly from a power source. Why?

8. What is meant by detent torque, and in what type of motors does detent torque occur?

9. What is meant by the 'holding torque' of a stepping motor?

10. Find the step angle of the following stepping motors: (a) 3-phase, VR, 12 stator teeth and 8 rotor teeth; (b) 3-phase, VR, three-stack, 16 rotor teeth; (c) 4-phase unipolar, hybrid, 50 rotor teeth.

11. Named two methods of Damping Rotor Oscillations and explain one method of damping.

12. Explain briefly about linear stepper motor.

13. A step motor has 130 steps per revolution. Find the input digital pulse rate that produces continuous rotation at a speed of 10.5 revolutions/ sec.

14. Explain the operation principles of a Brushless DC motor.

15. Named three types of direct drive actuator.

16. Explain how solenoid actuator works.

17. Draw a schematic diagram for fluid power actuation system.

18. Draw a schematic of fluid power servo system.

19. Draw symbol for hydraulic rotary actuator.

20. Draw symbol for double acting, single ended actuator.

21. List three advantages and three disadvantages of using fluid power for power transfer mechanism.

22. Name several pumps that used hydraulic fluid.

23. Draw a schematic diagram pneumatic actuation system.

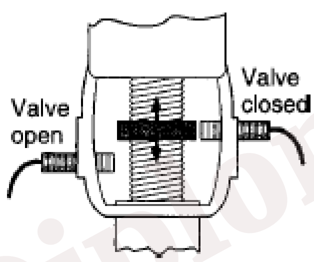
24. Given three advantages and three disadvantages for pneumatic system.

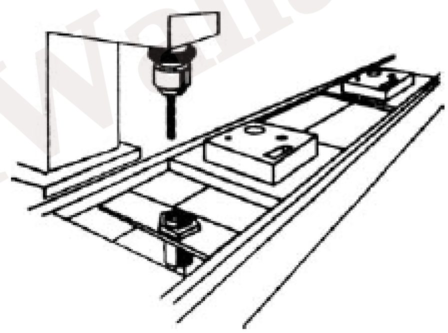
25. How proportional regulator valve works for a pneumatic system?

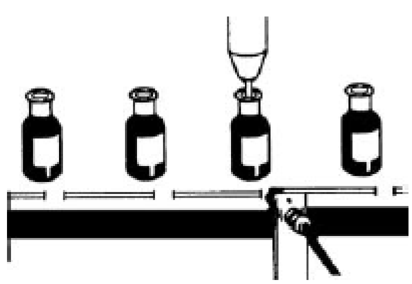
CHAPTER 5
AUTOMATION SENSORY DEVICES

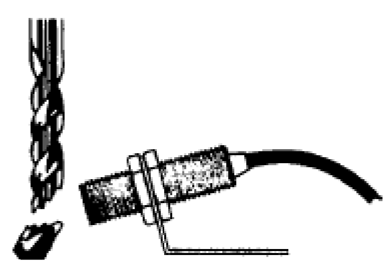
Upon completion of this course, students should be able to:-

- Describe the General Characteristics of sensor
- Explain the Angular and Linear Position Sensors
- Explain the Velocity and Acceleration Sensors
- Explain the Contact Sensors
- Explain the distance and velocity sensor









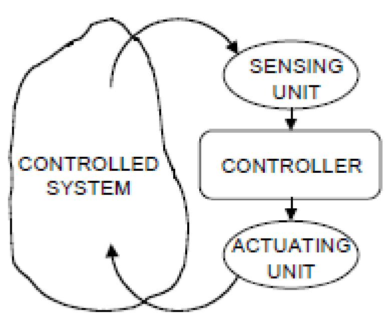
5.1 General Characteristics of sensor

The sensor is an element that tells the guard about the status of the processor either continuously or at the end of each movement. This information is used by regulators to ensure that each movement is accurately directed at the target. Information presented in the form of analog, digital or a combination of both.

	Use of sensor	Example
1	Detect a situation where operators were injured by robots or tools of production.	Sensors that stop the robot operation where employees into the work area while the robot are operating.
2	Detect situations where the robot can be damaged by the presence of other production tools.	Sensors for the presses to move before the complete material handling robots.
3	Control a process so that the quality of preserved products.	Control the temperature for heat treatment process.
4	Control operations and detect damage to the whole system.	Sensors located on the detector material forms so that operations can be carried out.

Sensors and actuators are two critical components of every closed loop control system. Such a system is also called a control system. A typical control system as shown in Fig. 5.1 consists of a sensing unit, a controller, and an actuating unit. A sensing unit can be as simple as a single sensor or can consist of additional components such as filters, amplifiers, modulators, and other signal conditioners. The controller accepts the information from the sensing unit, makes decisions based on the control algorithm, and outputs commands to the actuating unit. The actuating unit consists of an actuator and optionally a power supply and a coupling mechanism.

Sensor is a device that when exposed to a physical phenomenon (temperature, displacement, force, etc.) produces a proportional output signal (electrical, mechanical, magnetic, etc.). The term transducer is often used synonymously with sensors. However, ideally, a sensor is a device that responds to a change in the physical phenomenon. On the other hand, a transducer is a device that converts one form of energy into another form of energy. Sensors are transducers when they sense one form of energy input and output in a different form of energy. For example, a thermocouple responds to a temperature change (thermal energy) and outputs a proportional change in electromotive force (electrical energy). Therefore, a thermocouple can be called a sensor and or transducer.



Figurer 5.1: A typical control system

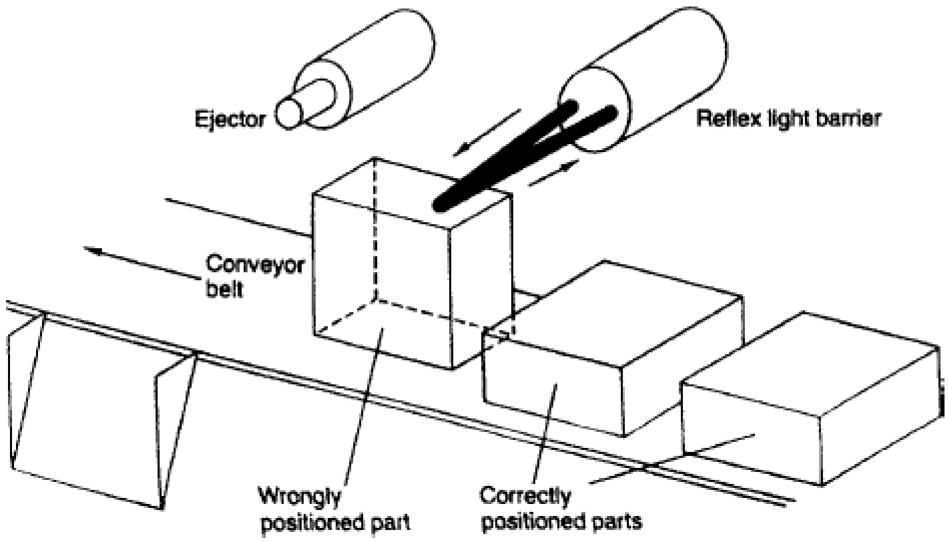


Figure 5.2: Sensors providing machine status

5.1.1 Sensor Classification

a. Classification of sensors (internal, external etc.)

Industrial robot requires sensory feedback to:

1. Locate randomly placed object;
2. Allow for variations in shape of objects;
3. Protect against dangerous and unexpected situations. Especially if the robot must work close to humans:
4. Allow “intelligent” recovery form error conditions;
5. Perform quality control.

Sensors will make robots more intelligent. But the associated robotic software must have the ability to receive data from the sensors and to process the necessary real time information and commands needed for the decision making.

In general, robotic sensors can be divided into two classes:

i. **Internal state sensors** - device being used to measure the position, velocity and acceleration of the robot joint and/or end-effector. These devices are potentiometer, tachometers, resolvers, encoders, differential transformers, optical interrupters, optical encoders and accelerometer.
ii. **External state sensors** – device being used to monitor the relationship between the robot kinematics and/or dynamics with its task, surrounding, or the object being manipulated.

Besides that, sensor also can be classified according to:

1- **Mechanical** sensors measure such quantities as position, shape, velocity, force, torque, pressure, vibration, strain and mass.
2- **Electrical** sensors measure voltage, current, charge and conductivity
3- **Magnetic** sensors measure magnetic field, flux and permeability
4- **Thermal** sensors measure temperature, flux, conductivity and specific heat.
5- Other types are acoustic, ultrasonic, chemical, optical, radiation, laser and fiber optic.

Depending on its application, a sensor may consist of metallic, nonmetallic, organic or inorganic, materials and fluids, plasmas, or semiconductors.

INTERNAL SENSOR	Function	Example
Internal sensor is a sensor located at all joints of robots and it is used to control the position and speed of tool-center point of the robot.	It forms a feedback control loop between the processor robot control units. It is usually *found on the manipulator that uses closed loop control* in which, the sensor will identify the position and speed of processing and sending data to the controller. The next controller to process the data and displays it on display "teach pendant" or a computer.	- **Position sensor** is '*potentiometer*' and *'optical encoders'* - **Speed sensor** is the choice of *tachometer*s.

EXTERNAL SENSOR

External sensor is the sensor that is placed outside the manipulator. They *adjust operations between the manipulator with other devices in a cell work* like Conveyor, actuator, CNC machines. It consists of two: **touch sensors** and **contactless sensor**.

	Type of external sensor	Sensor classification	Function	Example
1	**Proximity sensors**	Contact and contactless sensor	These sensors detect the presence of a substance either through touch or not by touching.	'Optical proximity sensors'
2	**Tactile sensor @ Physical sensor**	Sensors contact type	This sensor detects an object touching it.	Limit switches, pressure sensors
3	**Range sensor**	Without a touch sensor	Sensors that detect the presence of an object and measuring the position of the object and the sensor correctly	'Tellurometers' and 'interferometric'.
4	**Miscellaneous sensors**	Contact and contactless	Includes a variety of other sensors used in robotics	Equipment to measure temperature, fluid pressure, electrical voltage, current, and so on.

Sensors allow a PLC to detect the state of a process. Logical sensors can only detect a state that is either true or false. Examples of physical phenomena that are typically detected are listed below.

- inductive proximity - is a metal object nearby?
- capacitive proximity - is a dielectric object nearby?
- optical presence - is an object breaking a light beam or reflecting light?
- mechanical contact - is an object touching a switch?

CLASSIFICATION

Table 5.1 lists various types of sensors that are classified by their measurement objectives. Although this list is by no means exhaustive, it covers all the basic types including the new generation sensors such as smart material sensors, microsensors, and nanosensors.

Sensors can also be classified as *passive* or *active*. In passive sensors, the power required to produce the output is provided by the sensed physical phenomenon itself (such as a thermometer) whereas the active sensors require external power source (such as a strain gage).

Furthermore, sensors are classified as *analog* or *digital* based on the type of output signal. Analog sensors produce continuous signals that are proportional to the sensed parameter and typically require analog-to-digital conversion before feeding to the digital controller. Digital sensors on the other hand produce digital outputs that can be directly interfaced with the digital controller. Often, the digital outputs are produced by adding an analog-to-digital converter to the sensing unit. If many sensors are required, it is more economical to choose simple analog sensors and interface them to the digital controller equipped with a multi-channel analog-to-digital converter.

Table 5.1: Type of Sensors for Various Measurement Objectives

Sensor	Features
	Linear/Rotational sensors
Linear/Rotational variable differential transducer (LVDT/RVDT)	High resolution with wide range capability Very stable in static and quasi-static applications
Optical encoder	Simple, reliable, and low-cost solution Good for both absolute and incremental measurements
Electrical tachometer	Resolution depends on type such as generator or magnetic pickups
Hall effect sensor	High accuracy over a small to medium range
Capacitive transducer	Very high resolution with high sensitivity Low power requirements Good for high frequency dynamic measurements
Strain gauge elements	Very high accuracy in small ranges Provides high resolution at low noise levels
Interferometer	Laser systems provide extremely high resolution in large ranges Very reliable and expensive
Magnetic pickup	Output is sinusoidal
Gyroscope	
Inductosyn	Very high resolution over small ranges
	Acceleration sensors
Seismic accelerometer	Good for measuring frequencies up to 40% of its natural frequency
Piezoelectric accelerometer	High sensitivity, compact, and rugged Very high natural frequency (100 kHz typical)
	Force, torque, and pressure sensor
Strain gauge	Good for both static and dynamic measurements
Dynamometers/load cells	They are also available as micro- and nanosensors
Piezoelectric load cells	Good for high precision dynamic force measurements
Tactile sensor	Compact, has wide dynamic range, and high
Ultrasonic stress sensor	Good for small force measurements
	Flow sensors
Pitot tube	Widely used as a flow rate sensor to determine speed in aircrafts
Orifice plate	Least expensive with limited range
Flow nozzle, venturi tubes	Accurate on wide range of flow More complex and expensive
Rotameter	Good for upstream flow measurements Used in conjunction with variable inductance sensor
Ultrasonic type	Good for very high flow rates Can be used for both upstream and downstream flow measurements
Turbine flow meter	Not suited for fluids containing abrasive particles Relationship between flow rate and angular velocity is linear
Electromagnetic flow meter	Least intrusive as it is noncontact type Can be used with fluids that are corrosive, contaminated, etc. The fluid has to be electrically conductive
	Temperature sensors
Thermocouples	This is the cheapest and the most versatile sensor Applicable over wide temperature ranges (-200∞C to 1200∞C typical)
Thermistors	Very high sensitivity in medium ranges (up to 100∞C typical) Compact but nonlinear in nature
Thermodiodes, thermo transistors	Ideally suited for chip temperature measurements Minimized self heating
RTD—resistance temperature detector	More stable over a long period of time compared to thermocouple Linear over a wide range

(*continued*)

Table 5.1: Type of Sensors for Various Measurement Objectives (continued)

Sensor	Features
Infrared type	Noncontact point sensor with resolution limited by wavelength
Infrared thermography	Measures whole-field temperature distribution
	Proximity sensors
Inductance, eddy current, hall effect, photoelectric, capacitance, etc.	Robust noncontact switching action The digital outputs are often directly fed to the digital controller
	Light sensors
Photoresistors, photodiodes, photo transistors, photo conductors, etc.	Measure light intensity with high sensitivity Inexpensive, reliable, and noncontact sensor
Charge-coupled diode	Captures digital image of a field of vision
	Smart material sensors
Optical fiber	
As strain sensor	Alternate to strain gages with very high accuracy and bandwidth Sensitive to the reflecting surface's orientation and status
As level sensor	Reliable and accurate
As force sensor	High resolution in wide ranges
As temperature sensor	High resolution and range (up to 2000∞C)
Piezoelectric	
As strain sensor	Distributed sensing with high resolution and bandwidth
As force sensor	Most suitable for dynamic applications
As accelerometer	Least hysteresis and good setpoint accuracy
Magnetostrictive	
As force sensors	Compact force sensor with high resolution and bandwidth Good for distributed and noncontact sensing applications
As torque sensor	Accurate, high bandwidth, and noncontact sensor
	Micro- and nano-sensors
Micro CCD image sensor	Small size, full field image sensor
Fiberscope	Small (0.2 mm diameter) field vision scope using SMA coil actuators
Micro-ultrasonic sensor	Detects flaws in small pipes
Micro-tactile sensor	Detects proximity between the end of catheter and blood vessels

b. Sensor generalities (absolute, incremental, etc)

Several industrial sensing devices enable the robot to place objects at desired locations or perform various manufacturing processes:

- Transducers. Sensors that convert nonelectrical signals into electrical energy.
- Contact sensors (limit switches). Switches designed to be turned ON or OFF by an object exerting pressure on a lever or roller that operates the switch.
- Noncontact sensors. Devices that sense through changes in pressure, temperature, or electromagnetic field.
- Proximity sensors. Devices that sense the presence of a nearby object by inductance, capacitance, light reflection, or eddy currents.
- Range sensors. Devices such as laser-interferometric gauges that provide a precise distance measurement.
- Tactile sensors. Devices that rely on touch to detect the presence of an object; strain gauges can be used as tactile sensors.
- Displacement sensors. Provide the exact location of a gripper or manipulator. Resistive sensors are often used—usually wire-wound resistors with a slider contact. As force is applied to the slider arm, it changes the circuit resistance.
- Speed sensors. Devices such as tachometers that detect the motor shaft speed.
- Torque sensors. Measure the turning effort required to rotate a mass through an angle.

- Vision sensors. Enable a robot to see an object and generate adjustments suitable for object manipulation; include dissectors, flying-spot scanners, vidicons, orthicons, plumbicons, and charge-coupled devices.

Encoders use rotating disks with optical windows. The encoder contains an optical disk with fine windows etched into it. Light from emitter's passes through the openings in the disk to detectors. As the encoder shaft is rotated, the light beams are broken. The encoder shown here is a quadrature encode, and it will be discussed later. There are two fundamental types of encoders; absolute and incremental. An absolute encoder will measure the position of the shaft for a single rotation. The same shaft angle will always produce the same reading. The output is normally a binary or grey code number. An incremental (or relative) encoder will output two pulses that can be used to determine displacement. Logic circuits or software is used to determine the direction of rotation, and count pulses to determine the displacement. The velocity can be determined by measuring the time between pulses.

i. Absolute Encoders

A position can be read from an absolute encoder if the application requires knowledge of the position of the motors immediately upon system start-up. An absolute encoder is similar to an incremental encoder, except that the disk used has multiple concentric code tracks and a separate photodetector is used with each code track. The number of code tracks is equivalent to the binary resolution of the encoder, as shown in Figure 5.6.

An 8-bit absolute encoder has eight code tracks. The 8-bit output is read to form an 8-bitword indicating absolute position. While absolute encoders are available in a wide variety of resolutions, 8-, 10-, and 12-bit binary are the most common. Due to their complexity, absolute encoders are typically more expensive than quadrature encoders. Absolute encoders may output position in either parallel or serial format. Because of the variety of output formats available for absolute encoders, it is important to ensure that the robot controller or intelligent drive is compatible with the particular model of the absolute encoder.

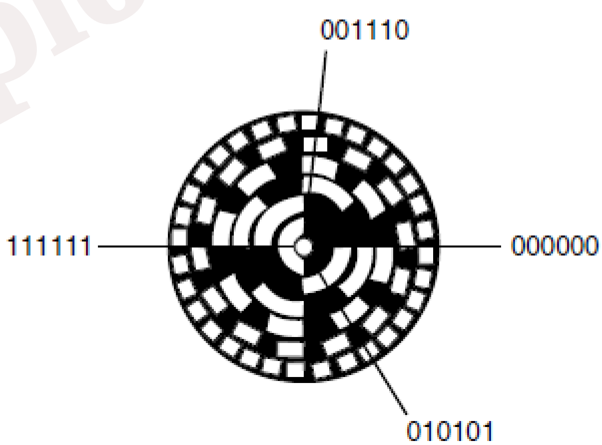


Figure 5.6: Absolute encoder example

ii. Analog Sensors

Analog sensors commonly used in robotic applications include displacement, force, torque, acceleration, and strain sensors. As in the case of encoders, these sensors may be used in either an open-loop or closed loop fashion within the robot system. For example, a force sensor may be used to measure the weight of objects being assembled for quality control. Or, a force sensor may be added to the gripper in a robot end effector as feedback to the gripper actuator. The gripper control system would allow objects to be held with a constant force.

Analog Displacement Sensors
These sensors can include both angular and translation position measurement relative to a reference position. They provide a continuously varying output signal which is proportional to the position of the sensed object. The most common sensors and technologies used for displacement measurement include:

- Potentiometer
- LVDT—contact sensor
- Resolvers
- Inductive
- Capacitive
- Optical
- Ultrasonic
- Hall effect

iii. Digital Sensors
A digital sensor will output either an "on" or an "off" electrical state. Apart from encoders, the majority of digital sensors used in robotic applications are static digital sensors in that their value is solely based on the digital state of the output as opposed to the frequency of an output pulse train. Static digital sensors do not require counter electronics for acquisition. Digital sensors can be used in a wide variety of applications within robotics. These include

- Switches as Digital Sensors - A mechanical switch is the simplest and lowest cost type of digital sensor used in robotics.
- Noncontact Digital Sensors - In order to reduce problems of contact wear and switch bounce, noncontact digital sensors are frequently used in robotics such as Inductive, Capacitive, Optical, Hall -effect.
- Solid State Output - Digital sensors frequently use transistors as the output driver technology.
- Proximity sensors
- Limit sensors
- Safety sensors such as light curtains

c. Sensor characteristics (linearity, resolution, dynamic characteristics etc.)

i. Range —Difference between the maximum and minimum value of the sensed parameter.
The range (or span) of a sensor is the difference between the minimum (or most negative) and maximum inputs that will give a valid output. Range is typically specified by the manufacturer of the sensor. For example, a common type K thermocouple has a range of 800°C (from 50°C to 750°C). A ten-turn potentiometer would have a range of 3600 degrees.

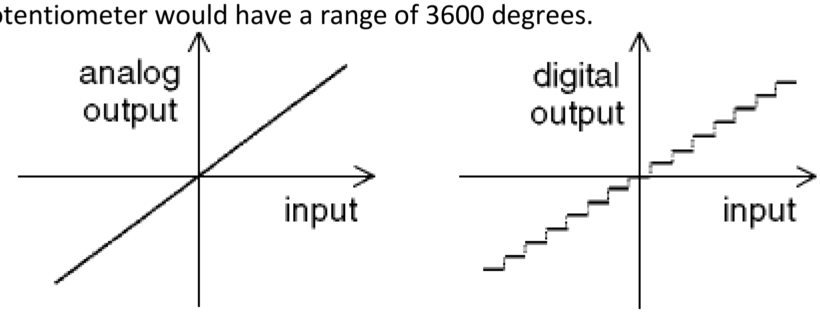


Figure 5.7: Analog and digital sensor outputs

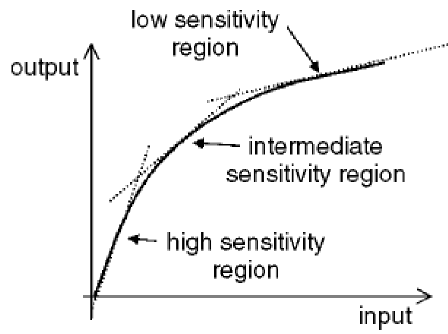


Figure 5.8: Sensor sensitivity

ii. Precision —Ability to reproduce repeatedly with a given accuracy. This considers accuracy, resolution and repeatability or one device relative to another The resolution of a sensor is the smallest increment of input that can be reliably detected. Resolution is also frequently known as the least count of the sensor. Resolution of digital sensors is easily determined. A 1024 ppr (pulse per revolution) incremental encoder would have a resolution of

$$\frac{1\,revolution}{1024\,pulses}\,x\,\frac{360\,degrees}{1\,revolution} = 0.3516\frac{degrees}{pulse}$$

The resolution of analog sensors is usually limited only by low-level electrical noise and is often than equivalent digital sensors.

iii. Sensitivity —Ratio of change in output to a unit change of the input
Sensor sensitivity is defined as the change in output per change in input. The sensitivity of digital sensors is closely related to the resolution. The sensitivity of an analog sensor is the slope of the output versus input line. A sensor exhibiting truly linear behavior has a constant sensitivity over the entire input range. Other sensors exhibit nonlinear behavior where the sensitivity either increases or decreases as the input is changed, as shown in Fig. 5.8.

iv. Error - is the difference between a measured value and the true input value. Two classifications of errors are bias (or systematic) errors and precision (or random) errors. Bias errors are present in all measurements made with a given sensor, and cannot be detected or removed by statistical means. These bias errors can be further subdivided into

- calibration errors (a zero or null point error is a common type of bias error created by a nonzero output value when the input is zero),
- loading errors (adding the sensor to the measured system changes the system), and
- errors due to sensor sensitivity to variables other than the desired one (e.g., temperature effects on strain gages).

v. Accuracy - This is the maximum difference between the indicated and actual reading.
For example, if a sensor reads a force of 100N with a ±1% accuracy, then the force could be anywhere from 99N to 101N.

vi. Resolution is used for systems that step through readings. This is the smallest increment that the sensor can detect; this may also be incorporated into the accuracy value. For example if a sensor measures up to 10 inches of linear displacements, and it outputs a number between 0 and 100, then the resolution of the device is 0.1 inches.

vii. Repeatability - When a single sensor condition is made and repeated, there will be a small variation for that particular reading. If we take a statistical range for repeated readings (e.g., ±3 standard deviations) this will be the repeatability. For example, if a flow rate sensor has a repeatability of 0.5cfm, readings for an actual flow of 100cfm should rarely be outside 99.5cfm to 100.5cfm.

viii. Linearity - In a linear sensor the input phenomenon has a linear relationship with the output signal. In most sensors this is a desirable feature. When the relationship is not linear, the conversion from the sensor output (e.g., voltage) to a calculated quantity (e.g., force) becomes more complex.

ix. Dynamic Response - The frequency range for regular operation of the sensor. Typically sensors will have an upper operation frequency; occasionally there will be lower frequency limits. For example, our ears hear best between 10Hz and 16KHz.

x. Environmental - Sensors all have some limitations over factors such as temperature, humidity, dirt/oil, corrosives and pressures. For example many sensors will work in relative humidities (RH) from 10% to 80%.

xi. Calibration - When manufactured or installed, many sensors will need some calibration to determine or set the relationship between the input phenomena, and output. For example, a temperature reading sensor may need to be zeroed or adjusted so that the measured temperature matches the actual temperature. This may require special equipment, and need to be performed frequently.

xii. Cost - Generally more precision costs more. Some sensors are very inexpensive, but the signal conditioning equipment costs are significant.

5.2 Angular and Linear Position Sensors

Linear and angular (rotational) position sensors are two of the most fundamental of all measurements used in a typical mechatronics system. The most common type position sensors are listed in Table 5.1. In general, the position sensors produce an electrical output that is proportional to the displacement they experience. There are contact type sensors such as strain gage, LVDT, RVDT, tachometer, etc. The noncontact type includes encoders, hall effect, capacitance, inductance, and interferometer type. They can also be classified based on the range of measurement. Usually the high-resolution type of sensors such as *hall effect, fiber optic inductance, capacitance*, and strain *gage* are suitable for only very small range (typically from 0.1 mm to 5 mm). The *differential transformers* on the other hand, have a much larger range with good resolution.

Interferometer type sensors provide both very high resolution (in terms of microns) and large range of measurements (typically up to a meter). However, interferometer type sensors are bulky, expensive, and require large set up time.

Among many linear displacement sensors, strain gage provides high resolution at low noise level and is least expensive. A typical resistance strain gage consists of resistive foil arranged as shown in the Fig. 5.9. A typical setup to measure the normal strain of a member loaded in tension is shown in Fig. 5.10. Strain gage 1 is bonded to the loading member whereas strain gage 2 is bonded to a second member made of same material, but not loaded. This arrangement compensates for any temperature effect. When the member is loaded, the gage 1 elongates thereby changing the resistance of the gage. The change in resistance is transformed into a change in voltage by the voltage sensitive wheat stone bridge circuit. Assuming that the resistance of all four arms are equal initially, the change in output voltage (Dv_0)due to change in resistance (DR_1) of gage 1 is

$$\frac{D_{V_0}}{V_i} = \frac{DR_1/R}{4 + 2(\frac{DR_1}{R})}$$

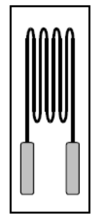


Figure 5.9: Bonded strain gauge

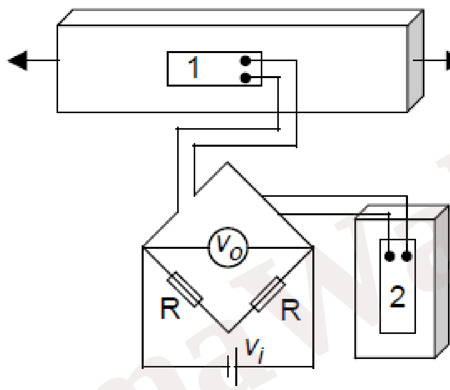


Figure 5.10: Experimental setup to measure normal strain using strain gauges

5.2.1 Methods of angular position measurement (resistive, capacitive, inductive, and optical)

By far the most common motions in mechanical systems are linear translation along a fixed axis and angular rotation about a fixed axis. More complex motions are usually accomplished by composing these simpler motions. In this chapter we provide a summary of some of the many technologies available for sensing linear and rotational motion along a single axis. We have arranged the sensing modalities according to the physical effect exploited to provide the measurement.

i. Resistive

One of the simplest and least expensive ways to measure rotational or linear motion is using a variable resistor called a *potentiometer* or *rheostat*. We focus on rotary potentiometers, or "pots" for short, but the principle of operation is the same in the linear case.

A pot consists of three terminals (Fig. 5.11(a,b)). Two end terminals, call them terminals 1 and 3, connect to either end of a length of resistive material, such as partially conductive plastic, ceramic, or a long thin wire. (For compactness, the long wire is wound around in loops to make a coil, leading to the name *wirewound* potentiometer.)

The other terminal, terminal 2, is connected to a *wiper*, which slides over the material as the pot shaft rotates. The total resistance of the pot R_{13} is equal to the sum of the resistance R_{12} between terminal 1 and the wiper, and the resistance R_{23} between the wiper and terminal 3. Typically the wiper can rotate from one end of the resistive material (R_{13}=R_{12}) to the other (R_{13}=R_{23}). If the full motion of the wiper is caused by one revolution of the shaft or less, the pot is called a *single-turn* pot. If the full motion is caused by multiple revolutions, it is called a *multi-turn* pot. Typically a pot is used by connecting terminal 1 to a voltage V, terminal 3 to ground, and using the voltage at the wiper as a measure of the rotation. The voltage observed at the wiper is *V* (R_{23}/R_{13}) and is a linear function of the rotation of the shaft. A remarkably simple absolute sensor for a wide range of distances is the string pot or draw-wire sensor (Fig. 5.12). It consists of a string wrapped on a spool, with a potentiometer to monitor rotations of the spool. A return spring keeps the string taut. Lengths up to many meters may be measured, using sensors incorporating multi-turn pots. The same technique is similarly useful for short distances (a few centimeters) using compact single-turn pots and a small spool. Both tolerate misalignment or arc-like motion well. String pots are susceptible to damage to the string in exposed applications, but the sensor element is small and unobtrusive. Manufacturers include RDP Electronics, SpaceAge Control, and UniMeasure.

Another type of resistive sensor is the flexible bend sensor. Conductive ink between two electrical contacts on a flexible material changes resistance as the material bends and stretches. Used in a voltage divider with a fixed resistor, the analog voltage may be used as a measure of the bend. Such a sensor could be used to detect contact (like a whisker) or as a rough measure of the deformation of a surface to which it is attached.

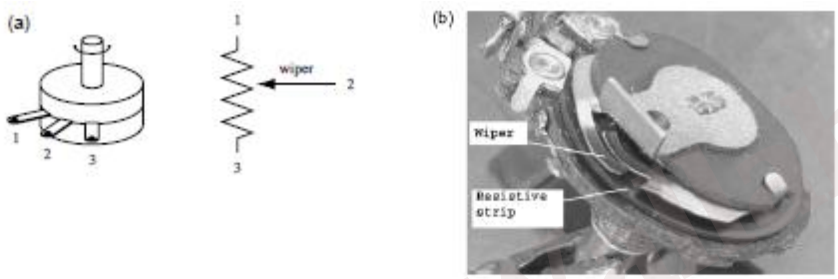


Figure 5.11 (a) As the shaft of the potentiometer rotates, the wipers moves from one end of the resistive material to the other. (b) The inside of a typical potentiometer, showing the wiper contacting a resistive strip.

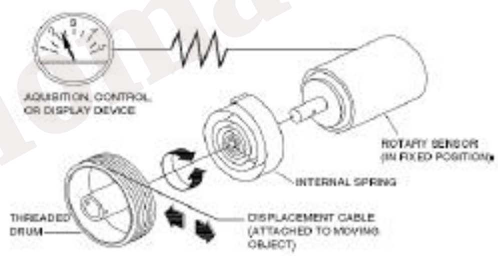


Figure 5.11: A string pot.

ii. Capacitive

Capacitance can be used to measure proximity or linear motions on the order of millimeters. The capacitance *C* of a parallel plate capacitor is given by $C = \frac{\varepsilon_r \varepsilon_0 A}{d}$, where ε_r the relative permittivity of the dielectric between the plates is, ε_0 is the permittivity of free space, *A* is the area of overlap of the two plates, and *d* is the plate separation. As the plates translate in the direction normal to their planes, *C* is a nonlinear function of the distance *d*. As the plates translate relative to each other in their planes, *C* is a linear function of the area of overlap *A*. Used as proximity sensors, capacitive sensors can detect metallic or nonmetallic objects, liquids, or any object with a dielectric constant greater than air.

One common sensing configuration has one plate of the capacitor inside a probe, sealed in an insulator. The external target object forms the other plate of the capacitor, and it must be grounded to the proximity sensor ground. As the sensor approaches the target, the capacitance increases, modifying the oscillation of a detector circuit including the capacitor. This altered oscillation may be used to signal proximity or to obtain a distance measurement. Manufacturers of capacitive sensors include Cutler-Hammer and RDP Electronics.

iii. AC Inductive (LVDT, Resolvers)

LVDT

The best known AC inductive sensor is the *linear variable differential transformer*, or LVDT. The LVDT is a tube with a plunger, the displacement of the plunger being the variable to be measured (Fig. 5.13). The tube is wrapped with at least two coils, an excitation coil and a pickup coil. An AC current (typically 1 kHz) is passed through the excitation coil, and an AC signal is detected from the pickup coil and compared in magnitude and in phase (0 or 180°) to the excitation current. Support electronics are needed for the demodulation, which is called synchronous detection. The plunger carries a ferromagnetic slug, which enhances the magnetic coupling from the excitation coil to the pickup coil. Depending on the position of the slug within the pickup coil, the detected signal may be zero (when the ferrite slug is centered in the pickup coil), or increasing in amplitude in one or the other phase, depending on displacement of the slug. LVDTs are a highly evolved technology and can be very accurate, in some cases to the micron level. They have displacement ranges of millimeters up to a meter. They do not tolerate misalignment or nonlinear motion, as a string pot does.

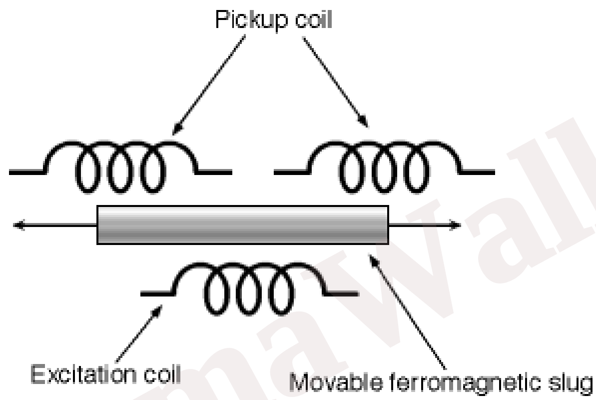


Figure 5.13: Operating principle of an LVDT

A linear variable differential transformer (LVDT) is a sensing transformer consisting of a primary winding, two adjacent secondary windings, and a ferromagnetic core that can be moved axially within the windings, as shown in the cutaway view Figure 5.14. LVDTs are capable of measuring position, acceleration, force, or pressure, depending on how they are installed. In motion control systems, LVDTs provide position feedback by measuring the variation in mutual inductance between their primary and secondary windings caused by the linear movement of the ferromagnetic core.

The core is attached to a spring-loaded sensing shaft. When depressed, the shaft moves the core axially within the windings, coupling the excitation voltage in the primary (middle) winding P1 to the two adjacent secondary windings S1 and S2. Figure 5.15 is a schematic diagram of an LVDT. When the core is centered between S1 and S2, the voltages induced in S1 and S2 have equal amplitudes and are 180º out of phase. With a series-opposed connection, as shown, the net voltage across the secondaries is zero because both voltages cancel. This is called the null position of the core.

However, if the core is moved to the left, secondary winding S1 is more strongly coupled to primary winding P1 than secondary winding S2, and an output sine wave in phase with the primary voltage is induced. Similarly, if the core is moved to the right and winding S2 is more strongly coupled to primary winding P1, an output sine wave that is 180° out-of-phase with the primary voltage is induced. The amplitudes of the output sine waves of the LVDT vary symmetrically with core displacement, either to the left or right of the null position.

Linear variable differential transformers require signal conditioning circuitry that includes a stable sine wave oscillator to excite the primary winding P1, a demodulator to convert secondary AC voltage signals to DC, a low-pass filter, and an amplifier to buffer the DC output signal. The amplitude of the resulting DC voltage output is proportional to the magnitude of core displacement, either to the left or right of the null position. The phase of the DC voltage indicates the position of the core relative to the null (left or right). An LVDT containing an integral oscillator/demodulator is a DC-to-DC LVDT, also known as a DCDT. Linear variable differential transformers can make linear displacement (position) measurements as precise as 0.005 in. (0.127 mm). Output voltage linearity is an important LVDT characteristic, and it can be plotted as a straight line within a specified range. Linearity is the characteristic that largely determines the LVDT's absolute accuracy.

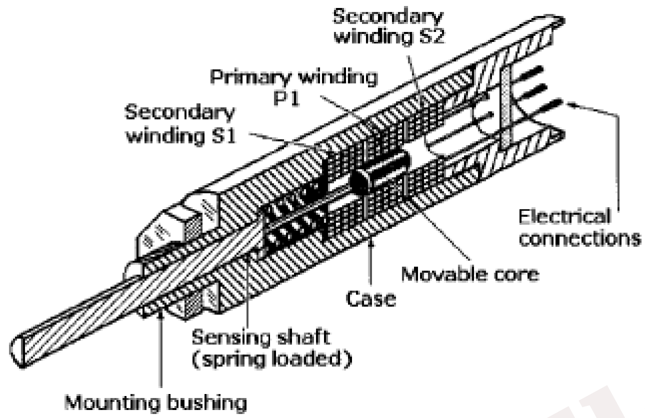


Figure 5.14: Cuteway view of linear variable displacement transformer (LVDT)

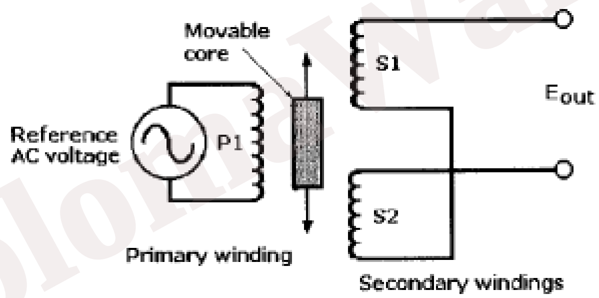


Figure 5.15: Schematic for a linear variable differential transformer

Resolvers

A resolver is essentially a rotary transformer that can provide position feedback in a servosystem as an alternative to an encoder. Resolvers resemble small AC motors, as shown in Figure 5.16, and generate an electrical signal for each revolution of their shaft. Resolvers that sense position in closed-loop motion control applications have one winding on the rotor and a pair of windings on the stator, oriented at 90º. The stator is made by winding copper wire in a stack of iron laminations fastened to the housing, and the rotor is made by winding copper wire in a stack of laminations mounted on the resolver's shaft.

Figure 5.17 is an electrical schematic for a brushless resolver showing the single rotor winding and the two stator windings 90º apart. In a servosystem, the resolver's rotor is mechanically coupled to the drive motor and load. When a rotor winding is excited by an AC reference signal, it produces an AC voltage output that varies in amplitude according to the sine and cosine of shaft position. If the phase shift between the applied signal to the rotor and the induced signal appearing on the stator coil is measured, that angle is an analog of rotor position. The absolute position of the load being driven can be determined by the ratio of the sine output amplitude to the cosine output amplitude as the resolver shaft turns through one revolution. (A single-speed resolver produces one sine and one cosine wave as the output for each revolution.)

Connections to the rotor of some resolvers can be made by brushes and slip rings, but resolvers for motion control applications are typically brushless. A rotating transformer on the rotor couples the signal to the rotor inductively. Because brushless resolvers have no slip rings or brushes, they are more rugged than encoders and have operating lives that are up to ten times those of brush-type resolvers. Bearing failure is the most likely cause of resolver failure. The absence of brushes in these resolvers makes them insensitive to vibration and contaminants. Typical brushless resolvers have diameters from 0.8 to 3.7 in. Rotor shafts are typically threaded and splined.

Most brushless resolvers can operate over a 2- to 40-volt range, and their winding is excited by an AC reference voltage at frequencies from 400 to 10,000 Hz. The magnitude of the voltage induced in any stator winding is proportional to the cosine of the angle, q, between the rotor coil axis and the stator coil axis. The voltage induced across any pair of stator terminals will be the vector sum of the voltages across the two connected coils. Accuracies of ±1 arc-minute can be achieved. In feedback loop applications, the stator's sinusoidal output signals are transmitted to a resolver-to-digital converter (RDC), a specialized analog-to-digital converter (ADC) that converts the signals to a digital representation of the actual angle required as an input to the motion controller.

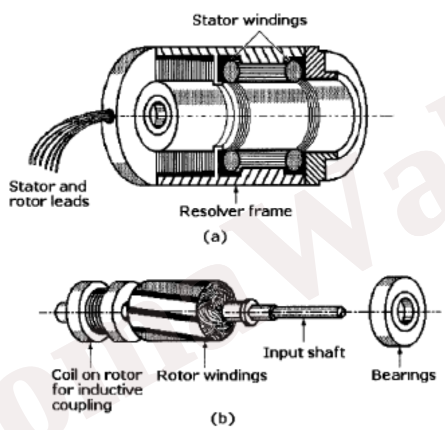


Figure 5.16: Exploded view of a brushless resolver frame (a) and rotor and bearings. (b) The coil on the rotor couples speed data inductively to the frame for processing

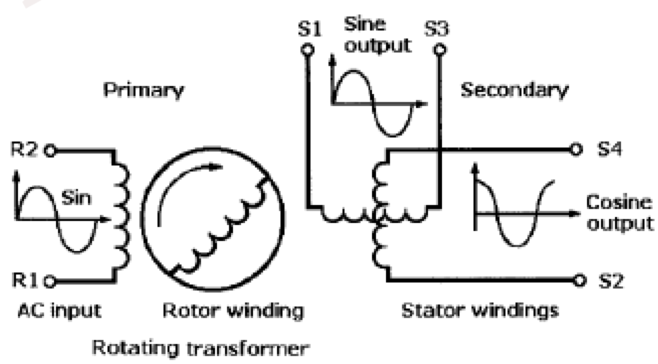


Figure 5.17: Schematic for a resolver shows how rotor position is transformed into sine and cosine outputs that measure rotor position

iv. Optical

Optical sensing techniques primarily rely on modulating the properties of an optical frequency electromagnetic wave. In the case of optical sensors, the measurand directly modulates the properties of the electromagnetic wave. In the case of micro-sensors, which use optical interfacing, the miniaturized sensor interacts with the measurand. The microsensor then modulates a property of the optical signal in order to provide an indication of the measurand. The following properties of the electromagnetic wave can be altered:

1. Intensity;
2. Phase;
3. Wavelength;
4. Spatial position;
5. Frequency;
6. Polarization.

For decades, optical sensors have been finding their way into an increasing number of applications. The development of semiconductors in the 1940s and '50s led to lower-cost, compact and efficient light-sensing devices. Photodetectors were used in camera light meters, street lights and traffic counters. Fiber optics allowed sensitive equipment to work in electrically noisy environments. Sensors packaged with tiny integrated circuits yielded detectors that were simpler to use. Optical sensors have improved efficiency and reliability of control systems at a reasonable cost.

Optical sensor is a general term for a group of sensors that use various wavelengths of light to perform certain tasks. Some rely on lasers while others use available light to perform their jobs, but all forms of optical sensors can be roughly categorized according to the job they perform.

Optical (Photoelectric) Sensors

Light sensors have been used for almost a century - originally photocells were used for applications such as reading audio tracks on motion pictures. But modern optical sensors are much more sophisticated. Optical sensors require both a light source (emitter) and detector. Emitters will produce light beams in the visible and invisible spectrums using LEDs and laser diodes.

Detectors are typically built with photodiodes or phototransistors. The emitter and detector are positioned so that an object will block or reflect a beam when present. A basic optical sensor is shown in Figure 5.18.

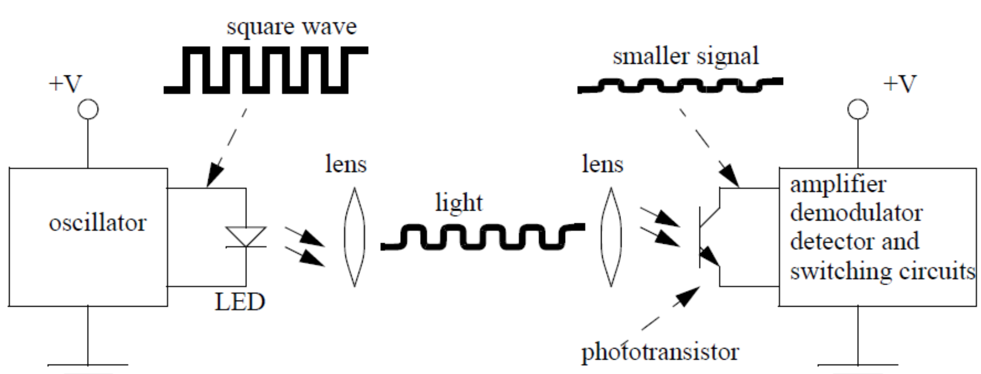


Figure 5.18: A Basic Optical Sensor

In the figure the light beam is generated on the left, focused through a lens. At the detector side the beam is focused on the detector with a second lens. If the beam is broken the detector will indicate an object is present. The oscillating light wave is used so that the sensor can filter out normal light in the room. The light from the emitter is turned on and off at a set frequency. When the detector receives the light it checks to make sure that it is at the same frequency. If light is being received at the right frequency then the beam is not broken. The frequency of oscillation is in the KHz range, and too fast to be noticed. A side effect of the frequency method is that the sensors can be used with lower power at longer distances.

An emitter can be set up to point directly at a detector, this is known as opposed mode. When the beam is broken the part will be detected. This sensor needs two separate components, as shown in Figure 5.19. This arrangement works well with opaque and reflective objects with the emitter and detector separated by distances of up to hundreds of feet.

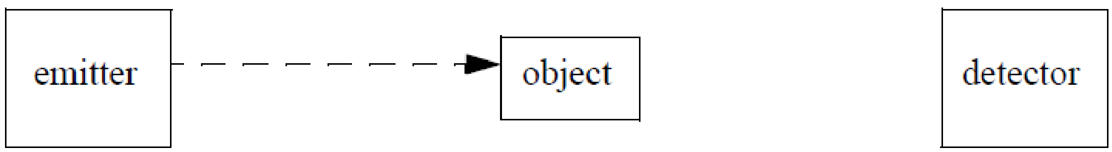
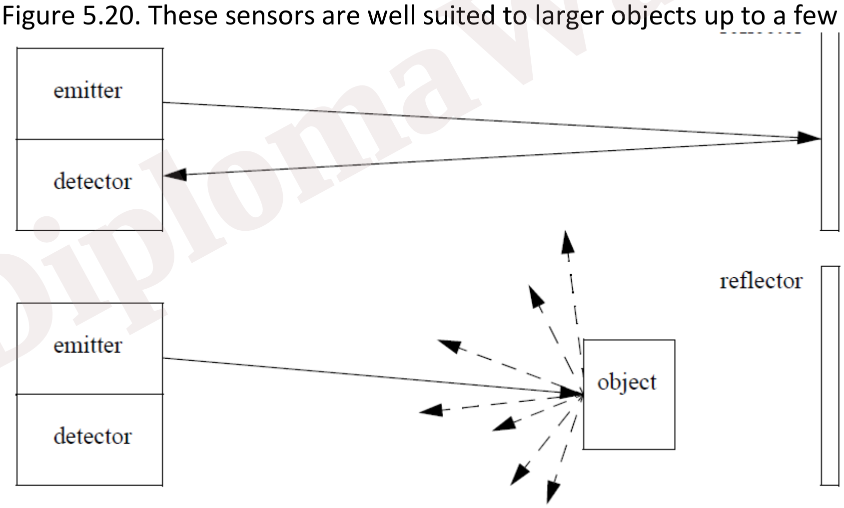


Figure 5.19: Opposed mode optical sensor

Having the emitter and detector separate increases maintenance problems and alignment is required. A preferred solution is to house the emitter and detector in one unit. But, this requires that light be reflected back as shown in Figure 5.20. These sensors are well suited to larger objects up to a few feet away.



Note: the reflector is constructed with polarizing screens oriented at 90 deg. angles. If the light is reflected back directly the light does not pass through the screen in front of the detector. The reflector is designed to rotate the phase of the light by 90 deg., so it will now pass through the screen in front of the detector.

Figure 5.20: Retroreflective Optical Sensor

In the figure, the emitter sends out a beam of light. If the light is returned from the reflector most of the light beam is returned to the detector. When an object interrupts the beam between the emitter and the reflector the beam is no longer reflected back to the detector, and the sensor becomes active. A potential problem with this sensor is that reflective objects could return a good beam. This problem is overcome by polarizing the light at the emitter (with a filter), and then using a polarized filter at the detector. The reflector uses small cubic reflectors and when the light is reflected the polarity is rotated by 90 degrees. If the light is reflected off the object the light will not be rotated by 90 degrees. So the polarizing filters on the emitter and detector are rotated by 90 degrees, as shown in Figure 5.21. The reflector is very similar to reflectors used on bicycles.

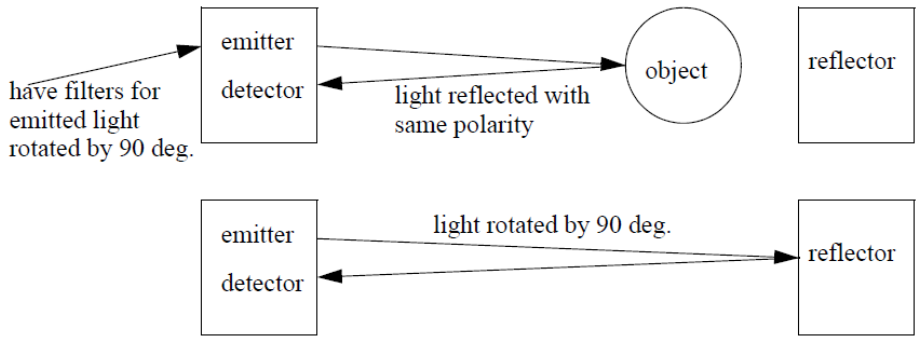


Figure 5.21: Polarized light in Retroreflective Sensors

Photoresistors

- Photoresistors, also known as light intensity sensors, do exactly what their name implies: They detect and measure the intensity of light. These sensors are used in science and modern technology in a number of different applications. For instance, intensity sensors are used in the liquid crystal displays on many mobile phones to adjust the screen brightness according the brightness of the surroundings. They also are used in modern cameras to adjust the exposure of the image.

Proximity Sensors

- Proximity sensors read changes in the ambient light environment to detect movement. These sensors have many modern applications. They are used in the hunting industry in trap cameras, which use the sensor to trigger the camera to take a photo of passing animals. They also are used in traffic cameras, using the same principle. Perhaps the most obvious use is in holiday decorations that are triggered when a person approaches.
-

Photodiodes

- Photodiodes are a type of optical sensor that converts light into electric current. This type of sensor can be found in many types of consumer electronics. They are present in remote control sensors on compact discs and other components. Photodiodes also are found in various forms of disc players. Photodiodes are more sensitive than photoresistors, but they are used in many of the same applications.

Advantages and Disadvantages of Optical Sensors

R&D in the optical sensor field is motivated by the expectation that optical sensors have significant advantages compared to conventional sensor types, in terms of their properties. Table 6.2 lists some of the advantages of optical over nonoptical sensors. Taking advantage of the capacity of optical fibers to send and receive optical signals over long distances, a current trend is to create networks of sensors, or sensor arrays. This avoids having to convert between electronics and photonics separately at each sensing site, thereby reducing costs and increasing flexibility. A difficulty of all sensors, both optical and non-optical, is interference from multiple effects. A sensor intended to measure strain or pressure may be very temperature-sensitive. Intense R&D over the last five years to provide means for distinguishing between various effects has been conducted for optical sensors. Advantages of Optical sensor:

- Greater sensitivity
- Electrical passiveness
- Freedom from electromagnetic
- Wide dynamic range
- Both point and distributed configuration
- Multiplexing capabilities

5.2.2 Encoding schemes (incremental, absolute)

Encoders Schemes
Encoders use rotating disks with optical windows, as shown in Figure 5.22. The encoder contains an optical disk with fine windows etched into it. Light from emitters passes through the openings in the disk to detectors. As the encoder shaft is rotated, the light beams are broken. The encoder shown here is a quadrature encode, and it will be discussed later.

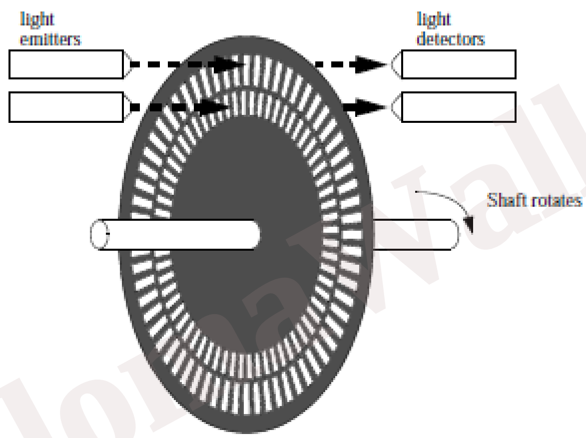


Figure 5.22: An encoder disk

There are two fundamental types of encoders; absolute and incremental. An absolute encoder will measure the position of the shaft for a single rotation. The same shaft angle will always produce the same reading. The output is normally a binary or grey code number. An incremental (or relative) encoder will output two pulses that can be used to determine displacement. Logic circuits or software is used to determine the direction of rotation, and count pulses to determine the displacement. The velocity can be determined by measuring the time between pulses. Encoder disks are shown in Figure 5.23. The absolute encoder has two rings, the outer ring is the most significant digit of the encoder, the inner ring is the least significant digit. The relative encoder has two rings, with one ring rotated a few degrees ahead of the other, but otherwise the same. Both rings detect position to a quarter of the disk. To add accuracy to the absolute encoder more rings must be added to the disk, and more emitters and detectors. To add accuracy to the relative encoder we only need to add more windows to the existing two rings. Typical encoders will have from 2 to thousands of windows per ring.

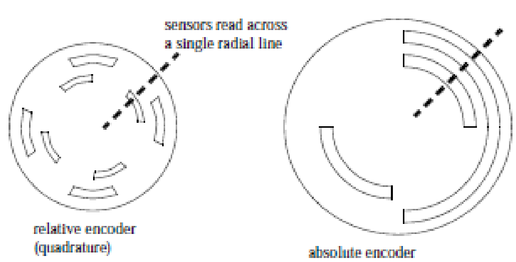
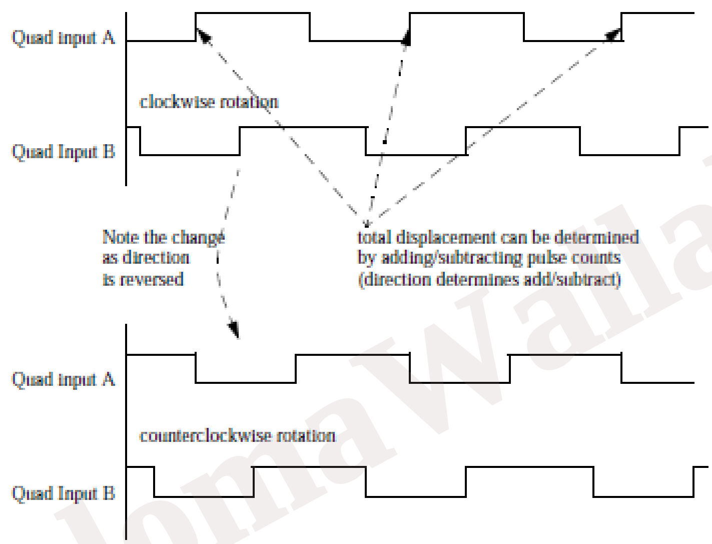


Figure 5.23: Encoder disk

When using absolute encoders, the position during a single rotation is measured directly. If the encoder rotates multiple times then the total number of rotations must be counted separately.

When using a relative encoder, the distance of rotation is determined by counting the pulses from one of the rings. If the encoder only rotates in one direction then a simple count of pulses from one ring will determine the total distance. If the encoder can rotate both directions a second ring must be used to determine when to subtract pulses. The quadrature scheme, using two rings, is shown in Figure 5.24. The signals are set up so that one is out of phase with the other. Notice that for different directions of rotation, input *B* either leads or lags *A*.



Note: To determine direction we can do a simple check. If both are off or on, the first to change state determines direction. Consider a point in the graphs above where both A and B are off. If A is the first input to turn on the encoder is rotating clockwise. If B is the first to turn on the rotation is counterclockwise.

Aside: A circuit (or program) can be built for this circuit using an up/down counter. If the positive edge of input *A* is used to trigger the clock, and input *B* is used to drive the up/down count, the counter will keep track of the encoder position.

Figure 5.24: Quadrature Encoders

Normally absolute and relative encoders require a calibration phase when a controller is turned on. This normally involves moving an axis until it reaches a logical sensor that marks the end of the range. The end of range is then used as the zero position. Machines using encoders, and other relative sensors, are noticeable in that they normally move to some extreme position before use.

Rotary Encoders

Rotary encoders, also called rotary shaft encoders or rotary shaft-angle encoders, are electromechanical transducers that convert shaft rotation into output pulses, which can be counted to measure shaft revolutions or shaft angle. They provide rate and positioning information in servo feedback loops. A rotary encoder can sense a number of discrete positions per revolution. The number is called points per revolution and is analogous to the steps per revolution of a stepper motor. The speed of an encoder is in units of counts per second. Rotary encoders can measure the motor-shaft or leadscrew angle to report position indirectly, but they can also measure the response of rotating machines directly.

The most popular rotary encoders are incremental optical shaft-angle encoders and the absolute optical shaft-angle encoders. There are also direct contact or brush-type and magnetic rotary encoders, but they are not as widely used in motion control systems. Commercial rotary encoders are available as standard or catalog units, or they can be custom made for unusual applications or survival in extreme environments. Standard rotary encoders are packaged in cylindrical cases with diameters from 1.5 to 3.5 in. Resolutions range from 50 cycles per shaft revolution to 2,304,000 counts per revolution. A variation of the conventional configuration, the *hollow-shaft encoder,* eliminates problems associated with the installation and shaft run out of conventional models. Models with hollow shafts are available for mounting on shafts with diameters of 0.04 to 1.6 in. (1 to 40 mm).

The basic parts of an incremental optical shaft-angle encoder are shown in Figure 5.25. A glass or plastic code disk mounted on the encoder shaft rotates between internal light sources, typically a light-emitting diode (LED), on one side and a mask and matching photo-detector assembly on the other side. The incremental code disk contains a pattern of equally spaced opaque and transparent segments or spokes that radiate out from its center as shown. The electronic signals that are generated by the encoder's electronics board are fed into a motion controller that calculates position and velocity information for feedback purposes. An exploded view of an industrial-grade incremental encoder is shown in Figure 5.25. Glass code disks containing finer graduations capable of 11- to more than 16-bit resolution are used in high-resolution encoders, and plastic (Mylar) disks capable of 8- to 10-bit resolution are used in the more rugged encoders that are subject to shock and vibration.

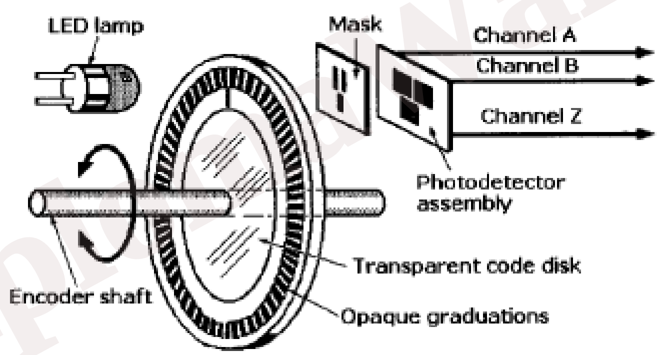


Figure 5.25: Basic elements of an incremental optical rotary encoder

The quadrature encoder is the most common type of incremental encoder. Light from the LED passing through the rotating code disk and mask is "chopped" before it strikes the photodetector assembly. The output signals from the assembly are converted into two channels of square pulses (A and B) as shown in Figure 5.26. The number of square pulses in each channel is equal to the number of code disk segments that pass the photodetectors as the disk rotates, but the waveforms are 90º out of phase. If, for example, the pulses in channel A lead those in channel B, the disk is rotating in a clockwise direction, but if the pulses in channel A lag those in channel B lead, the disk is rotating counterclockwise. By monitoring both the number of pulses and the relative phases of signals A and B, both position and direction of rotation can be determined. Many incremental quadrature encoders also include a third output Z channel to obtain a zero reference or index signal that occurs once per revolution. This channel can be gated to the A and B quadrature channels and used to trigger certain events accurately within the system. The signal can also be used to align the encoder shaft to a mechanical reference.

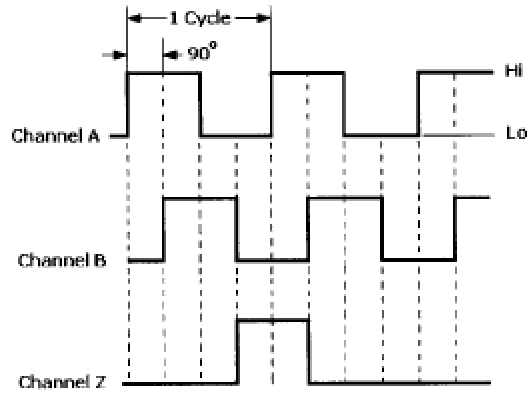


Figure 5.26: Channels A and B provides bidirectional position sensing.

Absolute Encoders

An absolute shaft-angle optical encoder contains multiple light sources and photodetectors, and a code disk with up to 20 tracks of segmented patterns arranged as annular rings, as shown in Figure 1-37. The code disk provides a binary output that uniquely defines each shaft angle, thus providing an absolute measurement. This type of encoder is organized in essentially the same way as the incremental encoder shown in Figure 1-35, but the code disk rotates between linear arrays of LEDs and photodetectors arranged radially, and a LED opposes a photodetector for each track or annular ring.

The arc lengths of the opaque and transparent sectors decrease with respect to the radial distance from the shaft. These disks, also made of glass or plastic, produce either the natural binary or Gray code. Shaft position accuracy is proportional to the number of annular rings or tracks on the disk. When the code disk rotates, light passing through each track or annular ring generates a continuous stream of signals from the detector array. The electronics board converts that output into a binary word.

The value of the output code word is read radially from the most significant bit (MSB) on the inner ring of the disk to the least significant bit (LSB) on the outer ring of the disk. The principal reason for selecting an absolute encoder over an incremental encoder is that its code disk retains the last angular position of the encoder shaft whenever it stops moving, whether the system is shut down deliberately or as a result of power failure. This means that the last readout is preserved, an important feature for many applications. Figure 5.27 Binary-code disk for an absolute optical rotary encoder. Opaque sectors represent a binary value of 1, and the transparent sectors

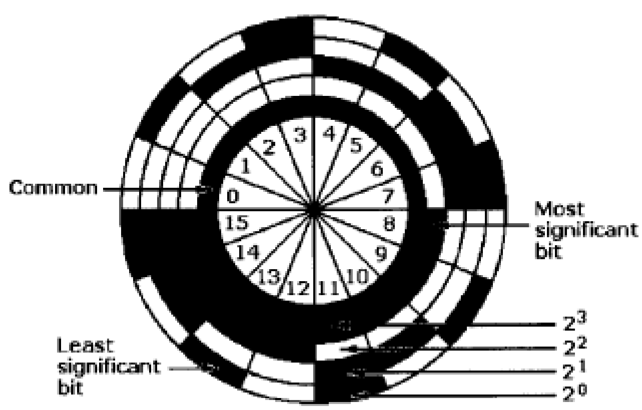


Figure 5.27: Binary-code disk for absolute optical rotary encoder

5.3 Velocity and Acceleration Sensors

Acceleration relating to motion is an important section of kinematic quantities: position, velocity, acceleration, and jerk. Each one of these quantities has a linear relationship with the neighboring ones. That is, all the kinematic quantities can be derived from a single quantity. For example, acceleration can be obtained by differentiating the corresponding velocity or by integrating the jerk. Likewise, velocity can be obtained by differentiating the position or by integrating the acceleration. In practice, only integration is widely used since it provides better noise characteristics and attenuation.

There are two classes of acceleration measurements techniques: *direct* measurements by specific accelerometers and *indirect* measurements where velocity is differentiated. The applicability of these techniques depends on the type of motion (rectilinear, angular, or curvilinear motion) or equilibrium centered vibration. For rectilinear and curvilinear motions, the direct measurement accelerometers are preferred. However, the angular acceleration is usually measured by indirect methods.

Acceleration is an important parameter for general-purpose absolute motion measurements, vibration, and shock sensing. For these measurements, accelerometers are commercially available in a wide range and many different types to meet diverse application requirements, mainly in three areas: (1) *Commercial applications*—automobiles, ships, appliances, sports and other hobbies; (2) *Industrial applications*—robotics, machine control, vibration testing and instrumentation; and (3) *High reliability applications*—military, space and aerospace, seismic monitoring, tilt, vibration and shock measurements.

Accelerometers have been in use for many years. Early accelerometers were mechanical types relying on analog electronics. Although early accelerometers still find many applications, modern accelerometers are essentially semiconductor devices within electronic chips integrated with the signal processing circuitry. Mechanical accelerometers detect the force imposed on a mass when acceleration occurs. A new type of accelerometer, the thermal type, senses the position through heat transfer.

Measurement of acceleration is important for systems subject to shock and vibration. Although acceleration can be derived from the time history data obtainable from linear or rotary sensors, the accelerometer whose output is directly proportional to the acceleration is preferred. Two common types include the seismic mass type and the piezoelectric accelerometer. The seismic mass type accelerometer is based on the relative motion between a mass and the supporting structure. The natural frequency of the seismic mass limits its use to low to medium frequency applications. The piezoelectric accelerometer, however, is compact and more suitable for high frequency applications.

5.3.1 Tachogenerator, optical incremental encoder, Sagnac interferometer, micromechanical angular velocity and acceleration sensor

i. Tachogenerator (tachometer)

A tachogenerator (tachometer) is a DC generator that can provide velocity feedback for a servo system. The tachometer's output voltage is directly proportional to the rotational speed of the armature shaft that drives it. In a typical servo system application, it is mechanically coupled to the DC motor and feeds its output voltage back to the controller and amplifier to control drive motor and load speed. A cross-sectional drawing of a tachometer built into the same housing as the DC motor and a resolver is shown in Figure 5.28. Encoders or resolvers are part of separate loops that provide position feedback.
As the tachometer's armature coils rotate through the stator's magnetic field, lines of force are cut so that an electromotive force is induced in each of its coils. This emf is directly proportional to the rate at which

the magnetic lines of force are cut as well as being directly proportional to the velocity of the motor's drive shaft. The direction of the emf is determined by Fleming's generator rule.

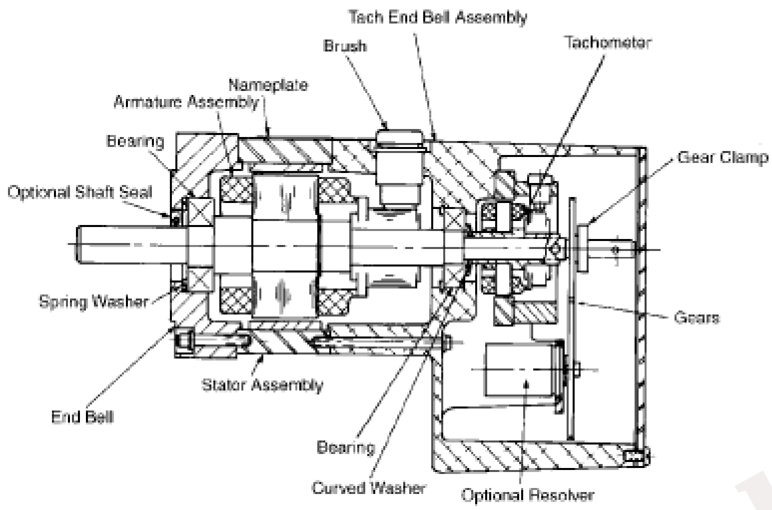
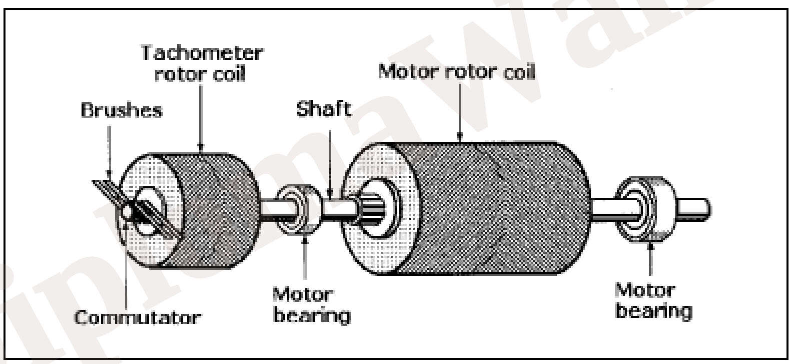


Figure 5.28: Section view of a resolver and tachometer in the same frame as the servo meter



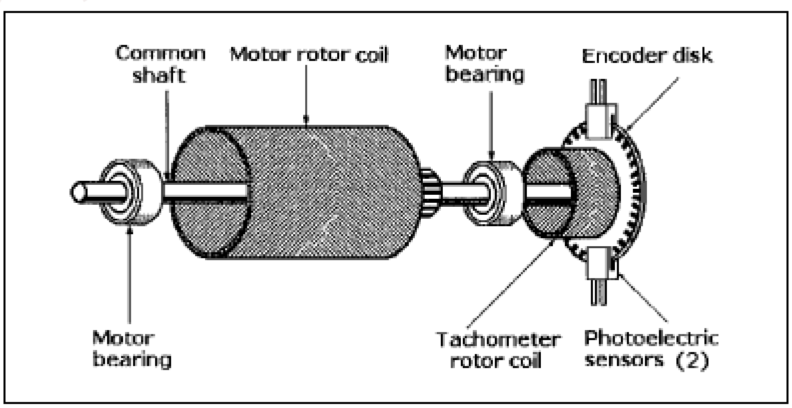


Figure 5.29: The rotors of the Dc motor and tachogenerator share a common shaft

The AC generated by the armature coil is converted to DC by the tachometer's commutator, and its value is directly proportional to shaft rotation speed while its polarity depends on the direction of shaft rotation. There are two basic types of DC tachometer: shunt wound and permanent magnet (PM), but PM tachometers are more widely used in servosystems today. There are also moving-coil tachometers which, like motors, have no iron in their armatures. The armature windings are wound from fine copper wire and bonded with glass fibers and polyester resins into a rigid cup, which is bonded to its coaxial shaft. Because this armature contains no iron, it has lower inertia than conventional copper and iron armatures, and it exhibits low inductance. As a result, the moving- coil tachometer is more responsive to speed changes and provides a DC output with very low ripple amplitudes.

Tachometers are available as standalone machines. They can be rigidly mounted to the servomotor housings, and their shafts can be mechanically coupled to the servomotor's shafts. If the DC servomotor is either a brushless or moving-coil motor, the standalone tachometer will typically be brushless and, although they are housed separately, a common armature shaft will be shared. A brush-type DC motor with feedback furnished by a brush-type tachometer is shown in Figure 5.29. Both tachometer and motor rotor coils are mounted on a common shaft. This arrangement provides a high resonance frequency. Moreover, the need for separate tachometer bearings is eliminated. In applications where precise positioning is required in addition to speed regulation, an incremental encoder can be added on the same shaft, as shown in Figure 5.29.

ii. Optical Incremental Encoders

Incremental encoders are the most common feedback devices for robotic systems. They typically output digital pulses at TTL levels. Rotary encoders are used to measure the angular position and direction of a motor or mechanical drive shaft. Linear encoders measure linear position and direction. They are often used in linear stages or in linear motors. In addition to position and direction of motion, velocity can also be derived from either rotary or linear encoder signals.

In a rotary incremental encoder, a glass or metal disk is attached to a motor or mechanical drive shaft. The disk has a pattern of opaque and transparent sectors known as a code track. A light source is placed on one side of the disk and a photodetector is placed on the other side. As the disk rotates with the motor shaft, the code track interrupts the light emitted onto the photodetector, generating a digital signal output. The number of opaque/transparent sector pairs, also known as line pairs, on the code track corresponds to the number of cycles the encoder will output per revolution. The number of cycles per revolution (CPR) defines the base resolution of the encoder.

Quadrature Encoders

Quadrature encoders are another type of incremental encoder. A two-channel quadrature encoder uses two photo detectors to sense both position and direction. The photo detectors are offset from each other by90◦ relative to one line pair on the code track. Since the two output signals, A and B, are 90◦ out of phase, one signal will lead the other as the disk rotates. If A leads B, the disk is rotating in a clockwise direction, as shown in Figure 5.30. If B leads A, the disk is rotating in a counterclockwise direction, as shown in Figure 5.30. Figure 5.30 and Figure 5.31 illustrate four separate pulse edges occurring during each cycle. A separate encoder count is generated with each rising and falling edge, which effectively quadruples the encoder resolution such that a 500 CPR (cycles per revolution) encoder provides 2000 counts per revolution with quadrature decoding.

The electrical complements of channels A and B may be included as differential signals to improve noise immunity. This is especially important in applications featuring long cable lengths between the encoder and the motion controller or electrical drive.

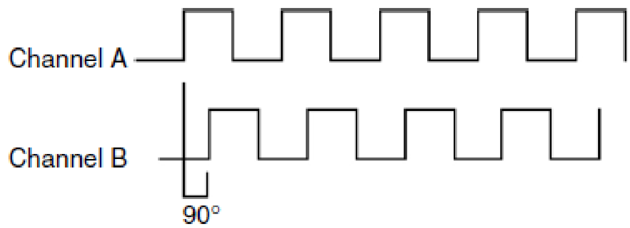


Figure 5.30: Clockwise motion

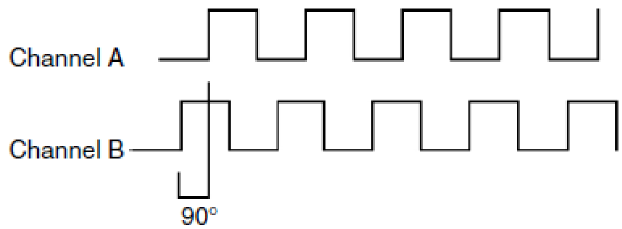


Figure 5.31: Counterclockwise motion

Some quadrature encoders include a third output channel, known as an index or zero pulse. This signal supplies a single pulse per revolution and is used for referencing the position of the system. During power up of the system, the motor can be rotated until the index pulse occurs. This specifies the current position of the motor in relation to the revolution.

Moving to the index position is not enough to determine the position of the motor or mechanical system during system startup. The absolute position is usually determined through a homing routine where the system is moved to limit switches or sensors. Once the robot is commanded to move to the limits, then the encoder readings can be reset to zero or some other position value to define the absolute position. Subsequent motion is measured by the encoders to be relative to this absolute position.

Encoders can be attached to components other than motors throughout a mechanical system for measurement and control. For example, a rotary motor may be attached to a belt drive that is in turn attached to a payload under test. A rotary encoder is attached to the motor to provide feedback for control, but a second rotary encoder can also be attached to the payload to return additional feedback for improved positioning. This technique is known as dual-loop feedback control and can reduce the effects of backlash in the mechanical components of the motion system.

Some encoders output analog sine and cosine signals instead of digital pulses. These types of encoders are typically used in very high precision applications that require positioning accuracy at the submicron level. In this case, an interpolation module is necessary between the encoder output and the robot controller or intelligent drive. This functionality may be included in the robot controller. Interpolation modules increase the resolution of the encoder by an integer value and provide a digital quadrature output for use by the robot controller. Some encoder manufacturers offer interpolation modules in a wide range of multiplier values. Manufacturers of encoders include BEI, Renco, US Digital, Renishaw, Heidenhain, and
Micro-E

An optical encoder uses photo-interrupters to convert motion into an electrical pulse train. These electrical pulses “encode” the motion, and the pulses are counted or “decoded” by circuitry to produce the displacement measurement. The motion may be either linear or rotational, but we focus on more common rotary optical encoders.

There are two basic configurations for rotary optical encoders, the *incremental* encoder and the *absolute* encoder. In an incremental encoder, a disk (or code-wheel) attached to a rotating shaft spins between two photo-interrupters. The disk has a radial pattern of lines, deposited on a clear plastic or glass disk or cut out of an opaque disk, so that as the disk spins, the radial lines alternately pass and block the infrared light to the photo-detectors. (Typically there is also a stationary mask, with the same pattern as the rotating code-wheel, in the light path from the emitters to the detectors.) This results in pulse trains from each of the photo-detectors at a frequency proportional to the angular velocity of the disk. These signals are labeled A and B, and they are 1/4 cycle out of phase with each other. The signals may come from photo-interrupters aligned with two separate tracks of lines at different radii on the disk, or they may be generated by the same track, with the photo-interrupters placed relative to each other to give out of phase pulse trains.

By counting the number of pulses and knowing the number of radial lines in the disk, the rotation of the shaft can be measured. The direction of rotation is determined by the phase relationship of the A and B pulse trains, i.e., which signal leads the other. For example, a rising edge of A while B = 1 may indicate counterclockwise rotation, while a rising edge of A while B = 0 indicates clockwise rotation. The two out-of-phase signals are known as quadrature signals. Incremental encoders commonly have a third output signal called the index signal, labeled I or Z. The index signal is derived from a separate track yielding a single pulse per revolution of the disk, providing a home signal for absolute orientation. In practice, multiple photo-interrupters can be replaced by a single source and a single array detecting device.

IC decoder chips are available to decode the pulse trains. The inputs to the chip are the A and B signals, and the outputs are one or more pulse trains to be fed into a counter chip. For example, the US Digital LS7083 outputs two pulse trains, one each for clockwise and counterclockwise rotation, which can be sent to the inputs of a 74193 counter, chip (Fig. 5.32). Standard decoding methods for the quadrature input are 1X, 2X, and 4X resolution. In 1X resolution, a single count is generated for the rising or falling edge of just one of the pulse trains, so that the total number of encoder counts for a single revolution of the disk is equal to the number of lines in the disk. In 4X resolution, a count is generated for each rising and falling edge of both pulse trains, resulting in four times the angular resolution. An encoder with 1000 lines on the code wheel being decoded at 4X resolution yields an angular resolution of 360 ∞ / (4 x 1000) = 0.09° . While a *single-ended output* encoder provides the signals A, B, and possibly Z, a *differential output* encoder also provides the complementary outputs A', B', and Z'. Differential outputs, when used with a differential receiver, can increase the electrical noise immunity of the encoder. A drawback of the incremental encoder is that there is no way to know the absolute position of the shaft at power-up without rotating it until the index pulse is received. Also, if pulses are momentarily garbled due to electrical noise, the estimate of the shaft rotation is lost until the index pulse is received. A solution to these problems is the absolute encoder. An absolute encoder uses *k* photo interrupters and *k* code tracks to produce a *k* -bit binary word uniquely representing 2*k* different orientations of the disk, giving an angular resolution of 360 °/2*k* (Fig. 5.33). Unlike an incremental encoder, an absolute encoder always reports the absolute angle of the encoder.

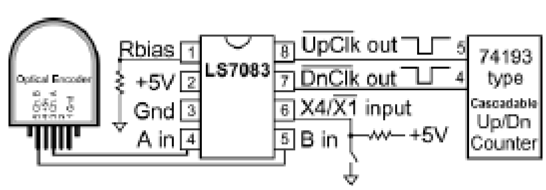


Figure: 5.32: An optical encoder, US Digital LS7083

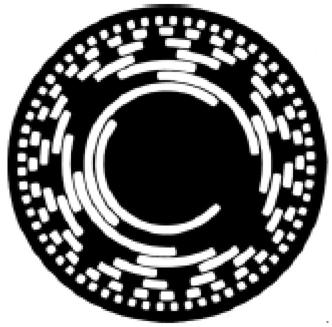


Figure 5.33: An 8-bit Gray code absolute encoder disk

The radial patterns on the tracks are arranged so that as the encoder rotates in one direction, the binary word increments or decrements according to a binary code. Although natural binary code is a possibility, the Gray code is a more common solution. With natural binary code, incrementing by one may change many or all of the bits, e.g., 7 to 8 in decimal is 0111 to 1000 in natural binary. With the Gray code, only one bit changes as the number increments or decrements, e.g., 7 to 8 in decimal is 0100 to 1100 in Gray code. The rotational uncertainty during a Gray code transition is only a single count, or 360°/2*k*. With the natural binary code, an infinitesimal misalignment between the lines and the photo interrupters may cause the reading to briefly go from 0111 (7) to 1111 (15) during the transition to 1000 (8). In general, incremental encoders provide higher resolution at a lower cost and are the most common choice for many industrial and robotic applications.

iii. Sagnac interferometer

The Sagnac effect (also called Sagnac interference), named after French physicist Georges Sagnac, is a phenomenon encountered in interferometry that is elicited by rotation. The Sagnac effect manifests itself in a setup called ring interferometry. A beam of light is split and the two beams are made to follow a trajectory in opposite directions. To act as a ring the trajectory must enclose an area. On return to the point of entry the light is allowed to exit the apparatus in such a way that an interference pattern is obtained. The position of the interference fringes is dependent on the angular velocity of the setup. This arrangement is also called a Sagnac interferometer. The Sagnac effect is the electromagnetic counterpart of the mechanics of rotation. A gimbal mounted gyroscope remains pointing in the same direction after spinning up, and thus can be used as the reference for an inertial guidance system. A Sagnac interferometer measures its own angular velocity with respect to the local inertial frame; hence just like a gyroscope it can provide the reference for an inertial guidance system. The principles behind the two devices are different, however. A gyroscope uses the principle of conservation of angular momentum whereas the interferometer is affected by relativistic phenomena.

Usually several mirrors are used, so that the light beams follow a triangular or square trajectory. Fiber optics can also be employed to guide the light. The ring interferometer is located on a platform that can rotate. When the platform is rotating the lines of the interference pattern are displaced as compared to the position of the interference pattern when the platform is not rotating. The amount of displacement is proportional to the angular velocity of the rotating platform. The axis of rotation does not have to be inside the enclosed area.

When the platform is rotating, the point of entry/exit moves during the transit time of the light. So one beam has covered less distance than the other beam. This creates the shift in the interference pattern. Therefore, the interference pattern obtained at each angular velocity of the platform features a different phase-shift particular to that angular velocity. In the above discussion, the rotation mentioned is rotation with respect to an inertial reference frame.

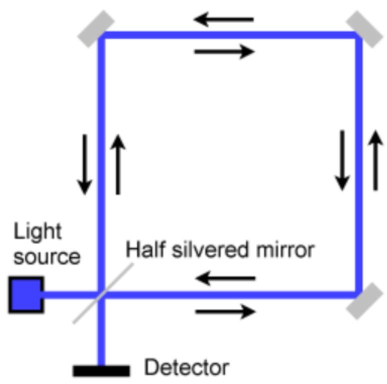
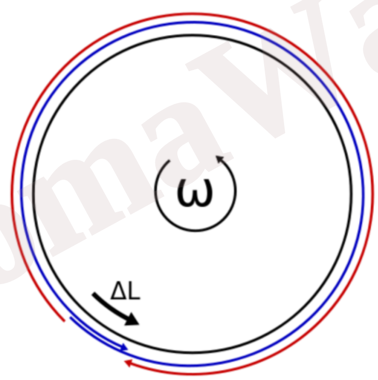


Figure 5.34: Sagnac interferometer

The shift in interference fringes can be viewed simply as a consequence of different *distances* light travels due to the rotation of the observer. The simplest derivation is for a circular ring rotating at an angular velocity of ω, but the result is general for loop geometries with other shapes. If a light source emits in both directions from one point on the rotating ring, light traveling with the rotation direction will travel more than one circumference around the ring and hit the light source from behind after a time t_1



Light traveling opposite directions go different distances before reaching the moving source again.

$$t_1 = \frac{2\pi R + \Delta L}{c}$$

$\Delta L$is the distance (black bold arrow in the figure) the mirror has moved in that same time:

$$\Delta L = R\omega t_1.$$

Eliminating $\Delta L$from the two equations above we get:

$$t_1 = \frac{2\pi R}{c - R\omega}.$$

Likewise, the light traveling against the rotation will travel less than one circumference before hitting the light source on the front side. So the time for this direction of light to reach the moving source again is:

$$t_2 = \frac{2\pi R}{c + R\omega}.$$

The time difference is

$$\Delta t = t_1 - t_2 = \frac{4\pi R^2 \omega}{c^2 - R^2 \omega^2}.$$

Since $R\omega = v \ll c$ this reduces to

$$\Delta t \approx \frac{4\pi R^2 \omega}{c^2} = \frac{4A\omega}{c^2},$$

where *A* is the area of the ring. This result happens to be general for any shape of loop with area *A*.

We imagine a screen for viewing fringes placed at the light source (or we use a beam splitter to send light from the source point to the screen). If the light were pulses shorter than Δt, there would be no interference. But applications use steady light, and shifting interference fringes are seen due to the presence of the two beams of light on the screen that left the source at different times and hence have different phases at the screen. The phase shift is $\Delta\phi = \frac{2\pi c \Delta t}{\lambda}$, which causes fringes to shift in proportion to A and ω.

In the case of light propagating in vacuum pre-relativistic theories and relativistic physics predict the same. In other words, in the case of propagation in vacuum a Sagnac experiment does not distinguish between pre-relativistic physics and relativistic physics.

When light propagates in fibre optic cable the setup is effectively a combination of a Sagnac experiment and the Fizeau experiment. In glass the speed of light is slower than in vacuum, and the optical cable is the moving medium. In that case the relativistic velocity addition rule applies. Pre-relativistic theories of light propagation cannot account for the Fizeau effect. (By 1900 Lorentz could account for the Fizeau effect, but by that time his theory had evolved to a form where in effect it was mathematically equivalent to special relativity.)

A clock attached to the ring would run slower due to its velocity than an inertial observer's, the light frequency of the moving source would increase to cancel that.
Also, Doppler effects cancel out, so the Sagnac effect does not involve the Doppler effect. In the case of ring laser interferometry it is important to be aware of this. When the ring laser setup is rotating the counter propagating beams undergo frequency shifts in opposite directions. This frequency shift is not a Doppler shift.

iv. Micromechanical Angular Velocity (Gyroscope)

Virtually all micromachined gyroscopes rely on a mechanical structure that is driven into resonance and excites a secondary oscillation in either the same structure or in a second one, due to the Coriolis force. The amplitude of this secondary oscillation is directly proportional to the angular rate signal to be measured. The Coriolis force is a virtual force that depends on the inertial frame of the observer. Imagine a person on a spinning disk, rolling a ball radially away from himself, with a velocity, v_r. The person in the rotating frame will observe a curved trajectory of the ball. This is due to the Coriolis acceleration that gives rise to a Coriolis force acting perpendicularly to the radial component of the velocity vector of the ball. A way of explaining the origin of this acceleration is to think of the current angular velocity of the ball on its way from the center of the disk to its edge, as shown in Figure 5.35.

The angular velocity v_{ang} increases with the distance of the ball from the center (v_{ang}= rΩ), but any change in velocity inevitably gives rise to acceleration in the same direction. This acceleration is given by the cross product of the angular velocity **Ω** of the disk and the radial velocity v_r of the ball:

$$\text{Coriolis acceleration: } \vec{a}_c = 2\vec{\Omega} \times \vec{v}_r; \quad \text{Coriolis force: } \vec{F}_c = 2m\vec{\Omega} \times \vec{v}_r$$

Macroscopic mechanical gyroscopes typically use a flywheel that has a high mass and spin speed and hence a large angular momentum which counteracts all external torque and creates an inertial reference frame that keeps the orientation of the spin axis constant. This approach is not very suitable for a micromachined sensor since the scaling laws are unfavorable where friction is concerned, and hence, there are no high-quality micromachined bearings. Consequently, nearly all MEMS gyroscopes use a vibrating structure that couples energy from a primary, forced oscillation mode into a secondary, sense oscillation mode. In Figure 5.36, a lumped model of a simple gyroscope suitable for a micromachined implementation is shown.

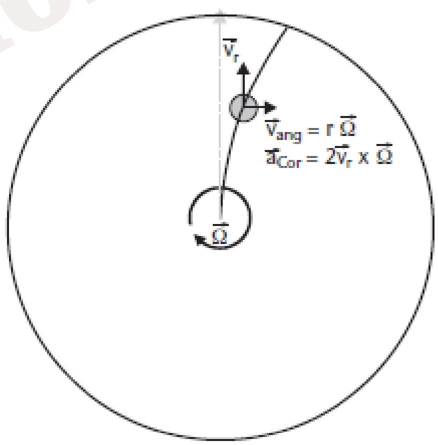


Figure 5.35: A ball rolling from the center of a spinning disk is subjected to Coriolis acceleration and hence shows a curved trajectory.

The proof mass is excited to oscillate along the x-axis with a constant amplitude and frequency. Rotation about the z-axis couples energy into an oscillation along the y-axis whose amplitude is proportional to the rotational velocity. Similar to closed loop micromachined accelerometers, it is possible to incorporate the sense mode in a force-feedback loop. Any motion along the sense axis is measured and a force is applied to counterbalance this sense motion. The magnitude of the required force is then a measure of the angular rate signal.

One problem is the relatively small amplitude of the Coriolis force compared to the driving force. Assuming a sinusoidal drive vibration given by $x(t) = x_0 \sin(\omega_d t)$, where x_0 is the amplitude of the oscillation and ω_d is the drive frequency, the Coriolis acceleration is given by $a_c = 2v(t)x\,\Omega = 2\Omega x_0 \omega_d \cos(\omega_d t)$. Using typical values of $x_0 = 1\mu m$, and $\omega_d = 2\pi 20$ kHz, the Coriolis acceleration is only 4.4 mm/s^2. If the sensing element along the sense axis is considered as a second order mass-spring-damper system with a Q = 1, the resulting displacement amplitude is only 0.0003 nm. One way to increase the displacement is to fabricate sensing elements with a high Q structure and then tune the drive frequency to the resonant frequency of the sense mode. Very high Q structures, however, require vacuum packaging, making the fabrication process much more demanding. Furthermore, the bandwidth of the gyroscopes is proportional to ω_d/Q; hence, if a quality factor of 10,000 or more is achieved in vacuum, the bandwidth of the sensor is reduced to only a few hertz. Lastly, it is difficult to design structures for an exact resonance frequency, due to manufacturing tolerances. A solution is to design the sense mode for a higher resonant frequency than the drive mode and then decrease the resonant frequency of the sense mode by tuning the mechanical spring constant using electrostatic forces. An acceptable compromise between bandwidth and sensitivity is to tune the resonant frequency of the sense mode close to the drive frequency (within 5% to 10%). A second fundamental problem with vibratory rate micromachined gyroscopes is due to so-called quadrature error. This type of error originates from manufacturing tolerances manifesting themselves as a misalignment of the axis of the driven oscillation from the nominal drive axis. As a result, a small proportion of the driven motion will be along the sense axis. Even though the misalignment angle is very small, due to the minute Coriolis acceleration, the resulting motion along the sense axis may be much larger than the motion caused by the Coriolis acceleration.

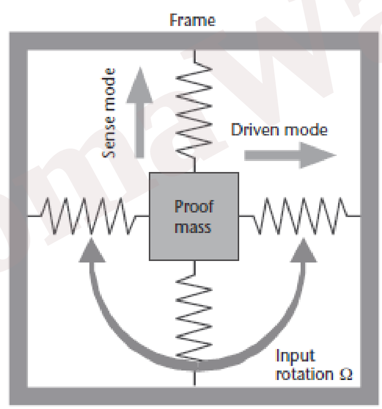


Figure 5.36: Lumped model of a vibratory gyroscope

A conventional-scale gyroscope utilizes the spatial coupling of the angular momentum-based gyroscopic effect to measure angular rate. In these devices, a disk is spun at a constant high rate about its primary axis, so that when the disk is rotated about an axis not colinear with the primary (or spin) axis, a torque results in an orthogonal direction that is proportional to the angular velocity. These devices are typically mounted in gimbals with low-friction bearings, incorporate motors that maintain the spin velocity, and utilize strain gages to measure the gyroscopic torque (and thus angular velocity). Such a design would not be appropriate for a microsensor due to several factors, some of which include the diminishing effect of inertia (and thus momentum) at small scales, the lack of adequate bearings, the lack of appropriate micromotors, and the lack of an adequate three-dimensional microfabrication processes. Instead, microscale angular rate sensors are of the vibratory type, which incorporate Coriolis-type effects rather thanthe angular momentum-based gyroscopic mechanics of conventional-scale devices. A Coriolis acceleration results from linear translation within a coordinate frame that is rotating with respect to an inertial reference frame. In particular, if the particle in Fig. 5.37 is moving with a velocity, *v* within the frame *xyz*, and if the frame *xyz* is rotating with an angular velocity of with respect to the inertial reference frame *XYZ*, then a Coriolis acceleration will result equal to $\mathbf{a_c}$= 2ω x *v*. If the object has a mass *m*, a Coriolis inertial force will result equal to **F***c* = -2*m*ω x *v* (minus sign because direction is opposite $\mathbf{a}_c$). A vibratory gyroscope utilizes this effect as illustrated in Fig. 5.38. A flexure-suspended inertial mass is vibrated in the *x*-direction, typically with an electrostatic comb drive. An angular velocity about the *z*-axis will generate a Coriolis acceleration, and thus force, in the *y*-direction. If the "external" angular velocity is constant and the velocity in the *x*-direction is sinusoidal, then the resulting Coriolis force will be sinusiodal, and the suspended inertial mass will vibrate in the *y*-direction with amplitude proportional to the angular velocity. The motion in the *y*-direction, which is typically measured capacitively, is thus a measure of the angular rate. Note that though vibration is an essential component of these devices, they are not technically resonant sensors, since they measure amplitude of vibration rather than frequency.

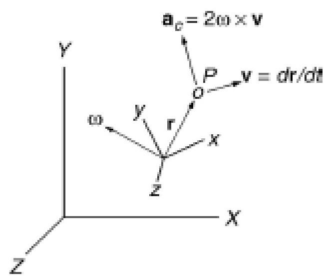


Figure 5.37: Illustration of Coriolis acceleration which result from translation within a reference frame that is rotating with respect to an inertial reference frame

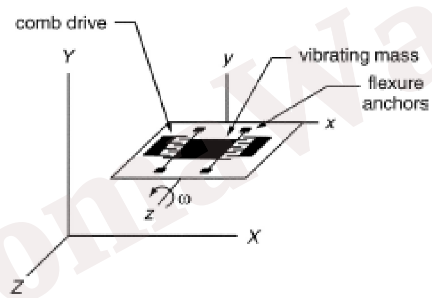


Figure 5.38: Schematic of a vibratory gyroscope

Micromachining technology has made possible very small cost-effective devices to sense angular rotation. The implementation of one such gyroscopic device for measuring roll/ pitch/yaw is shown in Fig. 5.39. This device is fabricated on a silicon substrate using the same surface micromachining techniques as used for the accelerometer. In this case, three layers of polysilicon are also used, with the first and third layers being fixed and the second layer being free to vibrate about its center. The center is held in position by four spring arms attached to four mounting posts, as shown. This device can sense rotation about two axes—that is, the x- and y-axes—and sense acceleration in the direction of the z-axis. The center layer of polysilicon is driven into oscillation about the z-axis by the electrostatic forces that are produced by voltages applied between the fixed comb fingers and the comb fingers of the second polysilicon. Capacitor plates as shown are formed between the first and third polysilicon on the x- and y-axes and the second layer of polysilicon. Differential capacitive sensing techniques are used to sense any displacement of the vibrating disc caused by angular rotation. For example, if angular rotation takes place about the x-axis, Coriolis forces produce a deflection of the disc about the y-axis. This deflection can then be detected by the capacitor plates on the x-axis. The sensing of the three functions is achieved by using a common sensing circuit that alternatively senses the x-rotation, y-rotation, and acceleration. The gyroscope is designed to have a resolution of <1°/s for angular rate measurements and an acceleration resolution of <20 mg.

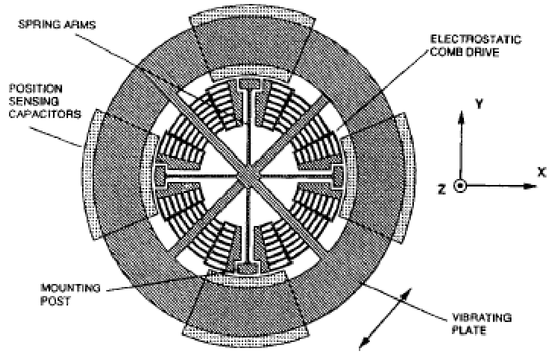


Figure 5.39 A three-layer micromachined gyroscope structure

V. Acceleration Sensors

A basic accelerometer consists of a mass that is free to move along a sensitive axis within a case. The technology is largely based on this basic accelerometer and can be classified in a number of ways, such as mechanical or electrical, active or passive, deflection or null-balance accelerometers, etc. The majority of industrial accelerometers are classified as either deflection or null-balance types. Accelerometers used in vibration and shock measurements are usually the deflection types, whereas those used for the measurement of motions of vehicles, aircraft, and so on for navigation purposes may be either deflection or null-balance type.

Figure 5.40: Commercial bulk-micromachined accelerometer from Colibrys.

Accelerometers measure acceleration using a mass suspended on a force sensor, as shown in Figure 5.41. When the sensor accelerates, the inertial resistance of the mass will cause the force sensor to deflect. By measuring the deflection the acceleration can be determined. In this case the mass is cantilevered on the force sensor. A base and housing enclose the sensor. A small mounting stud (a threaded shaft) is used to mount the accelerometer.

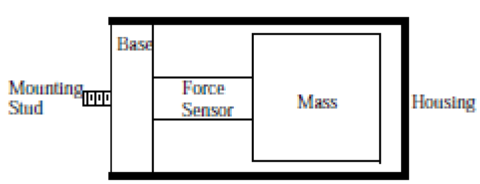


Figure 5.41: A cross section of an accelerometer

Accelerometers are dynamic sensors; typically used for measuring vibrations between 10Hz to 10KHz. Temperature variations will affect the accuracy of the sensors. Standard accelerometers can be linear up to 100,000 m/s**2: high shock designs can be used up to 1,000,000 m/s**2. There is often a trade-off between wide frequency range and device sensitivity (note: higher sensitivity requires a larger mass). Figure 5.42 shows the sensitivity of two accelerometers with different resonant frequencies. A smaller resonant frequency limits the maximum frequency for the reading. The smaller frequency results in a smaller sensitivity. The units for sensitivity is charge per m/s**2.

Resonant freq (Hz)	Sensitivity
22 Khz	4.5pc/(m/s**2)
180 Khz	.004

Figure 5.42: Piezoelectric Accelerometer Sensitivities

The force sensor is often a small piece of piezoelectric material (discussed later in this chapter). The piezoelectic material can be used to measure the force in shear or compression. Piezoelectric based accelerometers typically have parameters such as, -100 to 250°C operating range 1mV/g to 30V/g sensitivity operate well below one fourth of the natural frequency. The accelerometer is mounted on the vibration source as shown in Figure 5.43. The accelerometer is electrically isolated from the vibration source so that the sensor may be grounded at the amplifier (to reduce electrical noise). Cables are fixed to the surface of the vibration source, close to the accelerometer, and are fixed to the surface as often as possible to prevent noise from the cable striking the surface. Background vibrations can be detected by attaching control electrodes to *non-vibrating* surfaces. Each accelerometer is different, but some general application guidelines are;

- The control vibrations should be less than 1/3 of the signal for the error to be less than 12%).
- Mass of the accelerometers should be less than a tenth of the measurement mass.
- These devices can be calibrated with shakers, for example a 1g shaker will hit a peak velocity of 9.81 m/s*2.

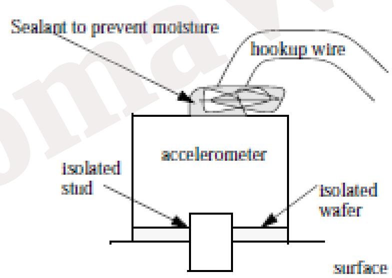


Figure 5.43: Mounting an Accelerometer

Equipment normally used when doing vibration testing is shown in Figure 5.44. The sensor needs to be mounted on the equipment to be tested. A pre-amplifier normally converts the charge generated by the accelerometer to a voltage. The voltage can then be analyzed to determine the vibration frequencies.

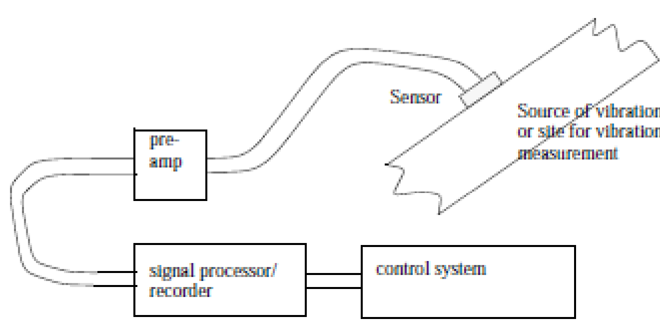


Figure 5.44: Typical connection for accelerometers

Accelerometers are commonly used for control systems that adjust speeds to reduce vibration and noise. Computer Controlled Milling machines now use these sensors to actively eliminate chatter, and detect tool failure. The signal from accelerometers can be integrated to find velocity and acceleration. Currently accelerometers cost hundreds or thousands per channel. But, advances in micromachining are already beginning to provide integrated circuit accelerometers at a low cost. Their current use is for airbag deployment systems in automobiles.

5.4 Contact Sensors

Force, which is a vector quantity, can be defined as an action that will cause acceleration or a certain reaction of a body. This chapter will outline the methods that can be employed to determine the magnitude of these forces. The determination or measurement of forces must yield to the following considerations: if the forces acting on a body do not produce any acceleration, they must form a system of forces in equilibrium. The system is then considered to be in static equilibrium. The forces experienced by a body can be classified into two categories: internal, where the individual particles of a body act on each other and external otherwise. If a body is supported by other bodies while subject to the action of forces, deformations and/or displacements will be produced at the points of support or contact. The internal forces will be distributed throughout the body until equilibrium is established, and then the body is said to be in a state of tension, compression, or shear. In considering a body at a definite section, it is evident that all the internal forces act in pairs, the two forces being equal and opposite, whereas the external forces act singly.

Basic Methods of Force Measurement

An unknown force may be measured by the following means:

1. balancing the unknown force against a standard mass through a system of levers,
2. measuring the acceleration of a known mass,
3. equalizing it to a magnetic force generated by the interaction of a current-carrying coil and a magnet,
4. distributing the force on a specific area to generate pressure and then measuring the pressure,
5. converting the applied force into the deformation of an elastic element.

The aforementioned methods used for measuring forces yield a variety of designs of measuring equipment. The challenge involved with the task of measuring force resides primarily in sensor design.

The basics of sensor design can be resolved into two problems:

1. primary geometric, or physical constraints, governed by the application of the force sensor device;
2. the means by which the force can be converted into a workable signal form (such as electronic signals or graduated displacements).

The remaining sections will discuss the types of devices used for force-to-signal conversion and finally illustrate some examples of applications of these devices for measuring forces.

Force sensors are required for a basic understanding of the response of a system. For example, cutting forces generated by a machining process can be monitored to detect a tool failure or to diagnose the causes of this failure in controlling the process parameters, and in evaluating the quality of the surface produced. Force sensors are used to monitor impact forces in the automotive industry. Robotic handling and assembly tasks are controlled by detecting the forces generated at the end effector. Direct measurement of forces is useful in controlling many mechanical systems.

Some types of force sensors are based on measuring a deflection caused by the force. Relatively high deflections (typically, several micrometers) would be necessary for this technique to be feasible. The excellent elastic properties of helical springs make it possible to apply them successfully as force sensors that transform the load to be measured into a deflection. The relation between force and deflection in the elastic region is demonstrated by Hooke's law. Force sensors that employ strain gage elements or piezoelectric (quartz) crystals with built-in microelectronics are common. Both impulsive forces and slowly varying forces can be monitored using these sensors.

5.4.1 Piezoresistive and capacitive tactile sensors, optical tactile sensors, force measurement by deformation of contact sensors: principle and applications of strain gage sensors

i. Piezoresistive Sensor and Capacitive Tactile Sensors

In a piezoresistive sensor, the magnitude of a mechanical displacement is measured by the amount of stress it induces in a mechanical member. A stress-sensitive resistor (called a piezoresistor) located strategically on the mechanical member experiences a change of resistance as a result of the applied stress.
Many materials, including metals, alloys, and doped silicon, exhibit piezoresistive characteristics. The applied stress causes the lattice of a material to deform, thereby inducing changes in the resistivity as well as the dimensions of a resistor. The change of resistance (ΔR) as a function of applied strain ε is

$$\frac{\Delta R}{R_0} = G\varepsilon$$

where R_0 is the value of the resistor in the unstressed state, and G is the piezoresistive gauge factor. Using doped silicon as a piezoresistive sensor, the overall footprint of the sensor can be made quite small and yet have a respectable value, i.e., 1 kΩ. Unlike the capacitive sensor method, which requires significant plate area to achieve significant capacitance value, the piezoresistive sensor is more area efficient. As a result, the piezoresistive sensing is more likely to be used for sensors with characteristic length below 10 μm. However, the capacitive measurement method is more generally applicable, whereas the optimal piezoresistive sensors involve silicon with proper doping concentration.

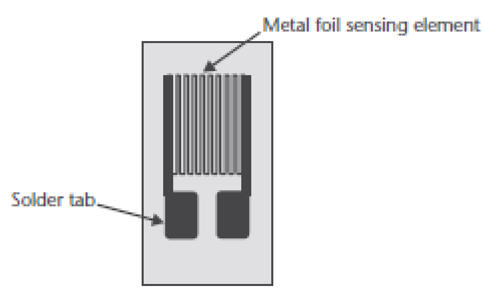


Figure 5.45: Illustration of a metal foil strain gauge

Capacitive Tactile Sensor
A transducer that uses capacitance variation can be used to measure force. The force is directed onto a membrane whose elastic deflection is detected by a capacitance variation. A highly sensitive force transducer can be constructed because the capacitive transducer senses very small deflections accurately. An electronic circuit converts the capacitance variations into DC-voltage variations.

A capacitance sensor consists of two metal plates separated by an air gap. The capacitance *C* between terminals is given by the expression:

$$\mathbf{C = \varepsilon_0 \varepsilon_r \frac{A}{h}}$$

where
C = capacitance in farads (F),
ε_0= dielectric constant of free space,
ε_r= relative dielectric constant of the insulator,
A = overlapping area for the two plates,
h = thickness of the gap between the two plates.

The sensitivity of capacitance-type sensors is inherently low. Theoretically, decreasing the gap *h* should increase the sensitivity; however, there are practical electrical and mechanical conditions that preclude high sensitivities. One of the main advantages of the capacitive transducer is that moving of one of its plates relative to the other requires an extremely small force to be applied. A second advantage is stability and the sensitivity of the sensor is not influenced by pressure or temperature of the environment.

To maximize the change in capacitance as force is applied, it is preferable to use a high permittivity, dielectric in a coaxial capacitor design. . In this type of sensor, as the size is reduced to increase the spatial resolution, the sensor's absolute capacitance will decrease. With the limitations imposed by the sensitivity of the measurement techniques, and the increasing domination of stray capacitance, there is an effective limit on the resolution of a capacitive array. The figure shows the cross section of the capacitive touch transducer in which the movement of a one set of the capacitors' plates is used to resolve the displacement and hence applied force. The use of a highly dielectric polymer such as polyvinylidene fluoride maximizes the change capacitance. From an application viewpoint, the coaxial design is better as its capacitance will give a greater increase for an applied force than the parallel plate design.

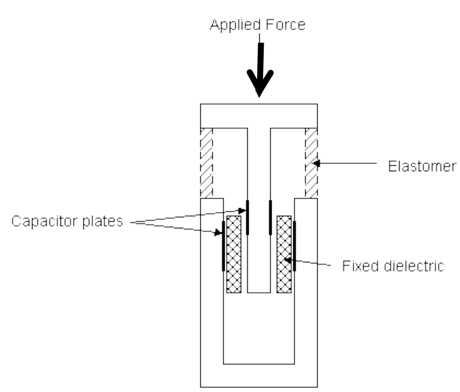


Figure 5.46: Force applied at the parallel plate design

Force Sensing Resistors (Conductive Polymers)

Force sensing resistors (FSRs) utilize the fact that certain polymer thick-film devices exhibit decreasing resistance with the increase of an applied force. A force sensing resistor is made up of two parts. The first is a resistive material applied to a film. The second is a set of digitating contacts applied to another film. Figure 5.47 shows this configuration. The resistive material completes the electrical circuit between the two sets of conductors on the other film. When a force is applied to this sensor, a better connection is made between the contacts; hence, the conductivity is increased. Over a wide range of forces, it turns out that the conductivity is approximately a linear function of force. Figure 5.48 shows the resistance of the sensor as a function of force. It is important to note that there are three possible regions for the sensor to operate. The first abrupt transition occurs somewhere in the vicinity of 10 g of force. In this region, the resistance changes very rapidly. This behavior is useful when one is designing switches using force sensing resistors.

FSRs should not be used for accurate measurements of force because sensor parts may exhibit 15–25% variation in resistance between each other. However, FSRs exhibit little hysteresis and are considered far less costly than other sensing devices. Compared to piezofilm, the FSR is far less sensitive to vibration and heat.

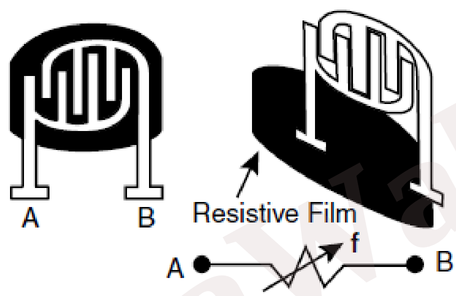


Figure 5.47: diagram of a typical force sensing resistor (FSR)

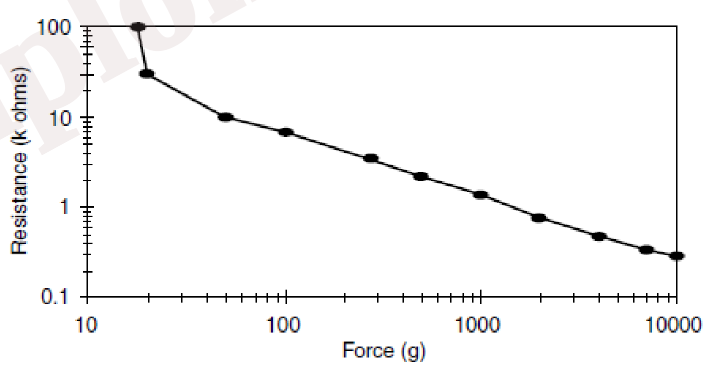


Figure 5.48: Resistance as a function of force for typical force sensing resistor

Tactile Sensors

Tactile sensors are usually interpreted as a touch sensing technique. Tactile sensors cannot be considered as simple touch sensors, where very few discrete force measurements are made. In tactile sensing, a force “distribution” is measured using a closely spaced array of force sensors.

Tactile sensing is important in both grasping and objects identification operations. Grasping an object must be done in a stable manner so that the object is not allowed to slip or get damaged. Object identification includes recognizing the shape, location, and orientation of a product, as well as identifying surface properties and defects. Ideally, these tasks would require two types of sensing:

1. continuous sensing of contact forces,
2. sensing of the surface deformation profile.

These two types of data are generally related through stress–strain relations of the tactile sensor. As a result, almost continuous variable sensing of tactile forces (the sensing of the tactile deflection profile) is achieved.

Tactile Sensor Requirements

Significant advances in tactile sensing are taking place in the robotics area. Applications include automated inspection of surface profiles, material handling or parts transfer, parts assembly, and parts identification and gaging in manufacturing applications and fine-manipulation tasks. Some of these applications may need only simple touch (force–torque) sensing if the parts being grasped are properly oriented and if adequate information about the process is already available. Naturally, the main design objective for tactile sensing devices has been to mimic the capabilities of human fingers. Typical specifications for an industrial tactile sensor include:

1. Spatial resolution of about 2 mm
2. Force resolution (sensitivity) of about 2 g
3. Maximum touch force of about 1 kg
4. Low response time of 5 ms
5. Low hysteresis
6. Durability under extremely difficult working conditions
7. Insensitivity to change in environmental conditions (temperature, dust, humidity, vibration, etc.)
8. Ability to monitor slip

Tactile Array Sensor

Tactile array sensors (Fig. 5.49) consist of a regular pattern of sensing elements to measure the distribution of pressure across the fingertip of a robot. The 8 × 8 array of elements at 2 mm spacing in each direction provides 64 force sensitive elements. Table 5.2 outlines some of the characteristics of early tactile array sensors. The sensor is composed of two crossed layers of copper strips separated by strips of thin silicone rubber. The sensor forms a thin, compliant layer that can be easily attached to a variety of fingertip shapes and sizes. The entire array is sampled by computer.

Table 5.2: Summary of some of the Characteristics of Early Tactile Array Sensors

	Size of Array		
Device Parameter	(4×4)	(8×8)	(16×16)
Cell spacing (mm)	4.00	2.00	1.00
Zero-pressure capacitance (fF)	6.48	1.62	0.40
Rupture force (N)	18.90	1.88	0.19
Max. linear capacitance (fF)	4.80	1.20	0.30
Max. output voltage (V)	1.20	0.60	0.30
Max. resolution (bit)	9.00	8.00	8.00
Readout (access) time (μs)	—	<20	—

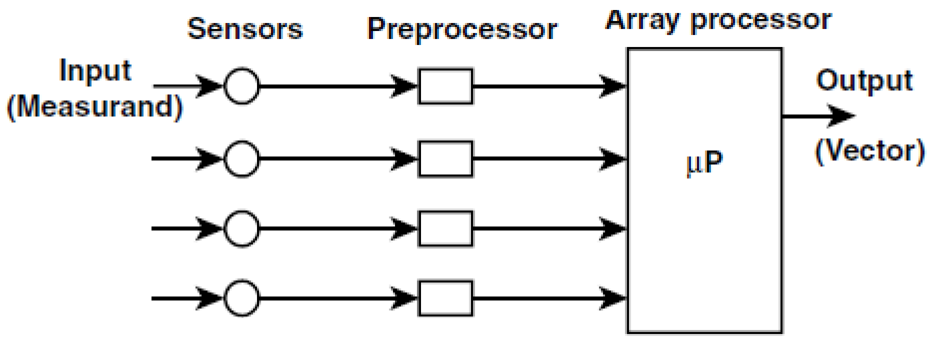


Figure 5.49: Tactile array sensor

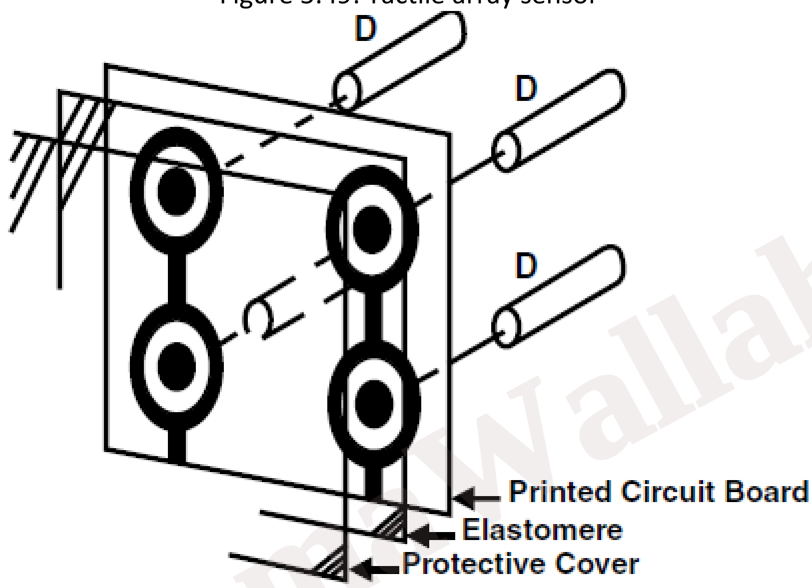


Figure 5.50: Typical taxel sensor array

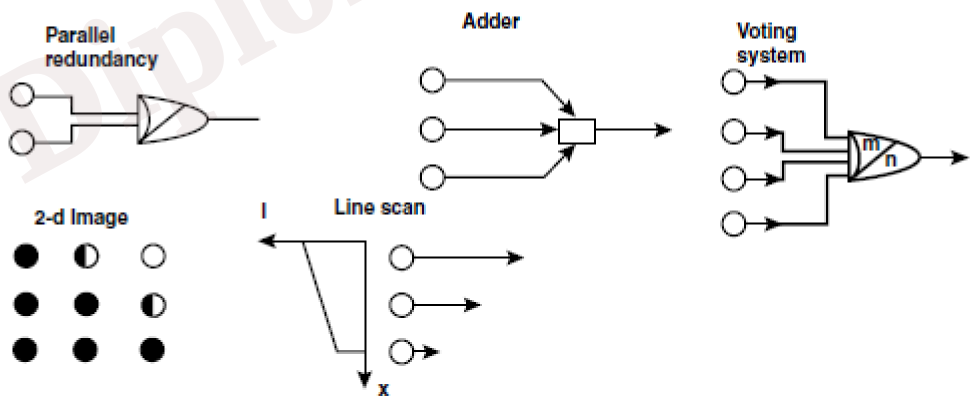


Figure 5.51: General arrangement of an intelligent array system

A typical tactile sensor array can consist of several sensing elements. Each element or taxel (Fig. 5.51) is used to sense the forces present. Since tactile sensors are implemented in applications where sensitivity providing semblance to human touch is desired, an elastomer is utilized to mimic the human skin. The elastomer is generally a conductive material whose electrical conductivity changes locally when pressure is applied. The sensor itself consists of three layers: a protective covering, a sheet of conductive elastomer, and a printed circuit board. The printed circuit board consists of two rows of two “bullseyes,” each with conductive inner and outer rings that compromise the taxels of the sensor. The outer rings are connected together and to a column-select transistor. The inner rings are connected to diodes (D) in Fig. 5.51.

Once the column in the array is selected, the current flows through the diodes, through the elastomer, and thence through a transistor to ground. As such, it is generally not possible to excite just one taxel because the pressure applied causes a local deformation in neighboring taxels. This situation is called crosstalk and is eliminated by the diodes. Tactile array sensor signals are used to provide information about the contact kinematics. Several feature parameters, such as contact location, object shape, and the pressure distribution, can be obtained. The general layout of a sensor array system can be seen in Fig. 5.50. An example of this is a contact and force sensing finger. This tactile finger has four contact sensors made of piezoelectric polymer strips on the surface of the fingertip that provide dynamic contact information. A strain gage force sensor provides static grasp force information.

ii. Optical tactile sensors

The rapid expansion of optical technology in recent years has led to the development of a wide range of tactile sensors. The operating principles of optical-based sensors are well known and fall into two classes:

- Intrinsic, where the optical phase, intensity, or polarization of transmitted light are modulated without interrupting the optical path
- Extrinsic, where the physical stimulus interacts with the light external to the primary light path.

Intrinsic and extrinsic optical sensors can be used for touch, torque, and force sensing. For industrial applications, the most suitable will be that which requires the least optical processing. For example the detection of phase shift, using interferometry, is not considered a practical option for robotic touch and force sensors. For robotic touch and force-sensing applications, the extrinsic sensor based on intensity measurement is the most widely used due to its simplicity of construction and the subsequent information processing. The potential benefits of using optical sensors can be summarized as follow:

Immunity to external electromagnetic interference, which is widespread in robotic applications:

- Intrinsically safe.
- The use of optical fiber allows the sensor to be located some distance from the optical source and receiver.
- Low weight and volume.

Touch and tactile optical sensors have been developed using a range of optical technologies:

(a) Modulating the intensity of light by moving an obstruction into the light path.

The force sensitivity is determined by a spring or elastomer. To prevent cross-talk from external sources, the sensor can construct around a deformable tube, resulting in a highly compact sensor, as shown in Figure 1.

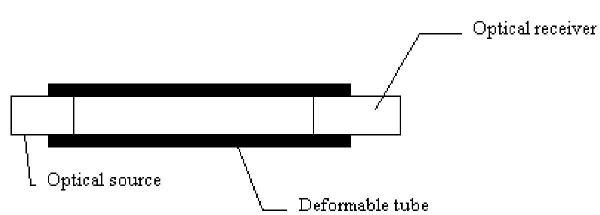


Figure 5.52: Intensity of light by moving an obstruction into the light path

In the reflective touch sensor below, the distance between the reflector and the plane of source and the detector is the variable. The intensity of the received light is a function of distance, and hence the applied force. The U shaped spring was manufactured from spring steel, leading to a compact overall design. This sensor has been successfully used in an anthropomorphic end effector (Figure 2).

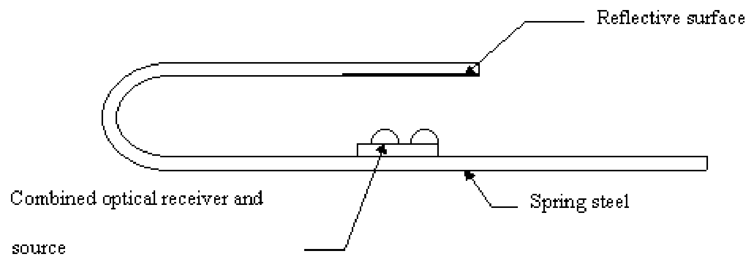


Figure 5.53: Reflective touch sensor

Reflective sensors can be constructed with source-receiver fiber pairs embedded in a solid elastomer structure. As shown below, above the fiber is a layer of clear elastomer topped with a reflective silicon rubber layer. The amount of light reflected to the receiver is determined by applied force that changes the thickness of the clear elastomer. For satisfactory operation the clear elastomer must have a lower compliance that the reflective layer. By the use of a number of matrixes of transmitter-receiver pairs, the tactile image of the contact can be determined:

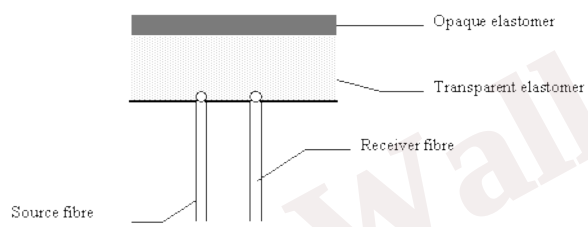


Figure 5.54: Clear elastomer topped with a reflective silicon rubber layer.

(b) Photoelasticity

Photoelasticity is the phenomena where stress or strain causes birefringence in optically transparent materials. Light is passed through the photoelastic medium. As the medium is stressed, the photoelastic medium effectively rotates the plane of polarization and hence the intensity of the light at the detector changes as a function of the applied force. This type of sensor is of considerable importance in the measurement of slip.

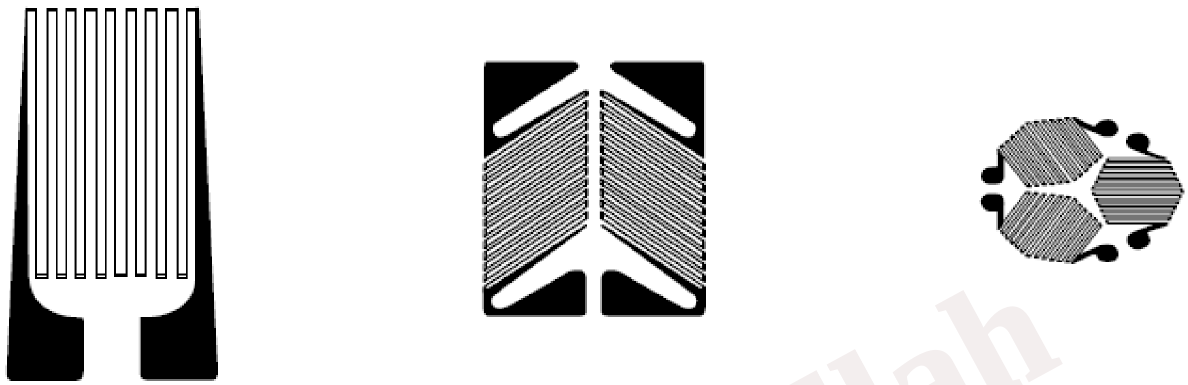
iii. Force measurement by deformation of contact sensors: principle and applications of strain gage sensors

Strain Gage Load Cell

The strain gage load cell consists of a structure that elastically deforms when subjected to a force and a strain gage network that produces an electrical signal proportional to this deformation. Examples of this are beam and ring types of load cells.

Strain Gages

Strain gages use a length of gage wire to produce the desired resistance (which is usually about 120Ω) in the form of a flat coil. This coil is then cemented (bonded) between two thin insulating sheets of paper or plastic. Such a gage cannot be used directly to measure deflection. It has to be first fixed properly to a member to be strained. After bonding the gage to the member, they are baked at about 195°F (90°C) to remove moisture. Coating the unit with wax or resin will provide some mechanical protection. The resistance between the member under test and the gage itself must be at least 50 MΩ. The total area of all conductors must remain small so that the cement can easily transmit the force necessary to deform the wire. As the member is stressed, the resulting strain deforms the strain gage and the cross-sectional area diminishes. This causes an increase in resistivity of the gage that is easily determined. In order to measure very small strains, it is necessary to measure small changes of the resistance per unit resistance ($\Delta R/R$). The change in the resistance of a bonded strain gage is usually less than 0.5%. A wide variety of gage sizes and grid shapes are available, and typical examples are shown in Fig. 5.55.



(a) (b) (c)

Figure 5.55: Configuration of metal-foil resistance gages: (a) single element, (b) two element and (c) three element

The use of strain gages to measure force requires careful consideration with respect to rigidity and environment. By virtue of their design, strain gages of shorter length generally possess higher response frequencies (examples: 660 kHz for a gage of 0.2 mm and 20 kHz for a gage of 60 mm in length). The environmental considerations focus mainly on the temperature of the gage. It is well known that resistance is a function of temperature and, thus, strain gages are susceptible to variations in temperature.

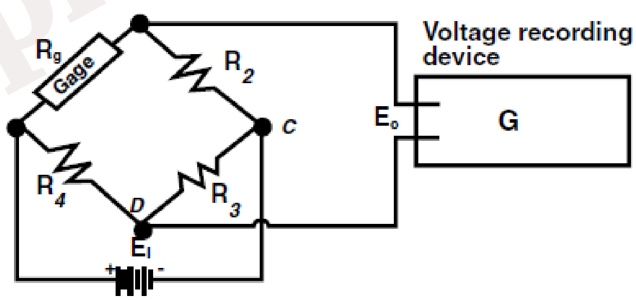
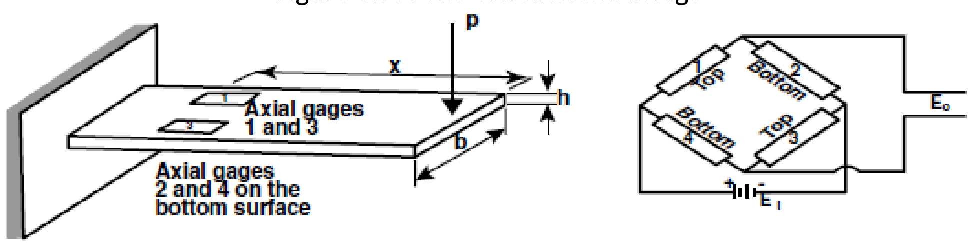


Figure 5.56: The Wheatstone bridge



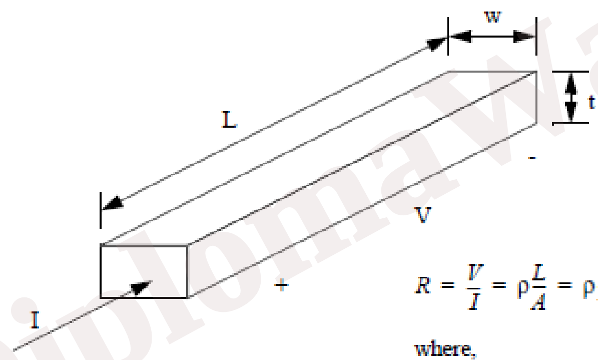
(a) (b)

Figure 5.57: Beam-type load cells; (a) a selection of beam-type load cells (elastic element with strain gages) and (b) gage position in the Wheatstone bridge.

Thus, if it is known that the temperature of the gage will vary due to any influence, temperature compensation is required in order to ensure that the force measurement is accurate. A Wheatstone bridge (Fig. 5.56) is usually used to measure this small order of magnitude. In Fig. 5.56, no current will flow through the galvanometer (G) if the four resistances satisfy a certain condition. In order to demonstrate how a Wheatstone bridge operates, a voltage scale has been drawn at points C and D of Fig. 5.56. Assume that R1 is a bonded gage. If R1 is now stretched so that its resistance increases by one unit (+ΔR), the voltage at point D will be increased from zero to plus one unit of voltage (+ΔV), and there will be a voltage difference of one unit between C and D that will give rise to a current through C. If R_4 is also a bonded gage, and at the same time that R1 changes by +ΔR, R_4 changes by ΔR, the voltage at D will move to +2ΔV. Also, if at the same time, R_2 changes by ΔR_4 and R_3 changes by +ΔR, then the voltage of point C will move to -2ΔV, and the voltage difference between C and D will now be 4ΔV. It is then apparent that although a single gage can be used, the sensitivity can be increased fourfold if two gages are used in tension while two others are used in compression.

Example:

Strain gages measure strain in materials using the change in resistance of a wire. The wire is glued to the surface of a part, so that it undergoes the same strain as the part (at the mount point). Figure 5.58 shows the basic properties of the undeformed wire. Basically, the resistance of the wire is a function of the resistivity, length, and cross sectional area.



$$R = \frac{V}{I} = \rho\frac{L}{A} = \rho\frac{L}{wt}$$

where,

R = resistance of wire

V, I = voltage and current

L = length of wire

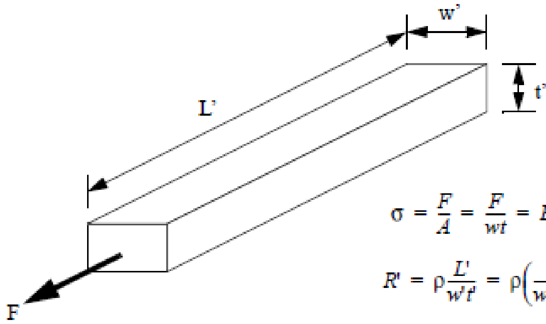
w, t = width and thickness

A = cross sectional area of conductor

ρ = resistivity of material

Figure 5.58 the Electrical Properties of a wire

After the wire in Figure 5.58 has been deformed it will take on the new dimensions and resistance shown in Figure 5.59. If a force is applied as shown, the wire will become longer, as predicted by Young's modulus. But, the cross sectional area will decrease, as predicted by Poison's ratio. The new length and cross sectional area can then be used to find a new resistance.



$$\sigma = \frac{F}{A} = \frac{F}{wt} = E\varepsilon \qquad \therefore \varepsilon = \frac{F}{Ewt}$$

$$R' = \rho\frac{L'}{w't'} = \rho\left(\frac{L(1+\varepsilon)}{w(1-\nu\varepsilon)t(1-\nu\varepsilon)}\right)$$

$$\therefore \Delta R = R' - R = R\left[\frac{(1+\varepsilon)}{(1-\nu\varepsilon)(1-\nu\varepsilon)} - 1\right]$$

where,

ν = poissons ratio for the material
F = applied force
E = Youngs modulus of material
σ, ε = stress and strain of material

Figure 5.59: The Electrical and mechanical Properties of the Deformed Wire

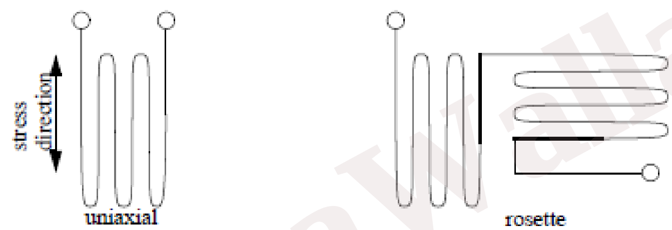


Figure 5.60 Wire Arrangement in strain gages

Design techniques using strain gages are to design a part with a narrowed neck to mount the strain gage on, as shown in Figure 5.61. In the narrow neck the strain is proportional to the load on the member, so it may be used to measure force. These parts are often called load cells.

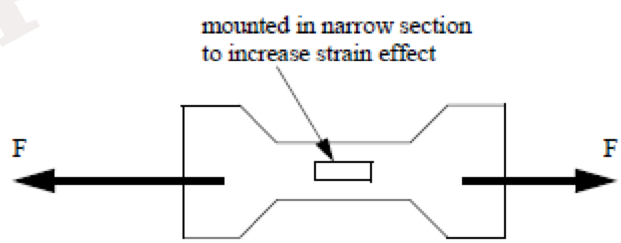


Figure 5.61 Using a Narrow to Increase Strain

Strain gauges are inexpensive, and can be used to measure a wide range of stresses with accuracies under 1%. Gages require calibration before each use. This often involves making a reading with no load, or a known load applied. An example application includes using strain gages to measure die forces during stamping to estimate when maintenance is needed.

5.5 Distance and velocity sensor

Distance Measuring Sensors

Range sensors are used to measure the distance from a reference point to an object. A number of technologies have been applied to develop these sensors, the most prominent being light/optics, computer vision, microwave, and ultrasonic. Range sensors may be of contact or noncontact types.

Noncontact Ranging Sensors

Sensors that measure the actual distance to a target of interest with no direct physical contact are referred to as noncontact ranging sensors. There are at least seven different types of ranging techniques employed in various implementations of such distance measuring devices (Everett et al., 1992):

- Triangulation
- Time of flight (pulsed)
- Phase-shift measurement (CW)
- Frequency modulation (CW)
- Interferometry
- Swept focus
- Return signal intensity

Noncontact ranging sensors can be broadly classified as either *active* (radiating some form of energy into the field of regard) or *passive* (relying on energy emitted by the various objects in the scene under surveillance). The commonly used terms *radar* (radio direction and ranging), *sonar* (sound navigation and ranging), and *lidar* (light direction and ranging) refer to *active* methodologies that can be based on any of several of the above ranging techniques. For example, *radar* is usually implemented using time of- flight, phase-shift measurement, or frequency modulation.

Sonar typically is based on time-of-flight ranging, since the speed of sound is slow enough to be easily measured with fairly inexpensive electronics. *Lidar* generally refers to laser-based schemes using time-of-flight or phase-shift measurement. For any such active (reflective) sensors, effective detection range is dependent not only on emitted power levels, but also the following target characteristics:

- *Cross-sectional area* —determines how much of the emitted energy strikes the target.
- *Reflectivity* —determines how much of the incident energy is reflected versus absorbed or passed through.
- *Directivity* —determines how the reflected energy is redistributed (i.e., scattered versus focused). Many noncontact sensors operate based on the physics of wave propagation. A wave is emitted at a reference point, and the range is determined by measuring either the propagation time from reference to target, or the decrease of intensity as the wave travels to the target and returns to the reference. Propagation time is measured using time-of-flight or frequency modulation methods.

Proximity Sensors

They are used to sense the proximity of an object relative to another object. They usually provide a on or off signal indicating the presence or absence of an object. *Inductance, capacitance, photoelectric*, and *hall effect* types are widely used as proximity sensors. Inductance proximity sensors consist of a coil wound around a soft iron core. The inductance of the sensor changes when a ferrous object is in its proximity.
This change is converted to a voltage-triggered switch. Capacitance types are similar to inductance except the proximity of an object changes the gap and affects the capacitance. Photoelectric sensors are normally aligned with an infrared light source. The proximity of a moving object interrupts the light beam causing the voltage level to change. Hall effect voltage is produced when a current-carrying conductor is exposed to a transverse magnetic field. The voltage is proportional to transverse distance between the hall effect sensor and an object in its proximity.

Proximity sensors, used to determine the presence (as opposed to actual range) of nearby objects, were developed to extend the sensing range beyond that afforded by direct-contact tactile or haptic sensors. Recent advances in electronic technology have significantly improved performance and reliability, thereby increasing the number of possible applications. As a result, many industrial installations that historically have used mechanical limit switches can now choose from a variety of alternative noncontact devices for their close (between a fraction of an inch and a few inches) sensing needs. Such *proximity sensors* are classified into several types in accordance with the specific properties used to initiate a switching action:

- Magnetic
- Inductive
- Ultrasonic
- Microwave
- Optical
- Capacitive

The reliability characteristics displayed by these sensors make them well suited for operation in harsh or otherwise adverse environments, while providing high-speed response and long service lives. Instruments can be designed to withstand significant shock and vibration; with some capable of handling forces over 30,000 Gs and pressures of nearly 20,000 psi (Hall, 1984). Burreson (1989) and Peale (1992) discuss advantages and tradeoffs associated with proximity sensor selection for applications in challenging and severe environments. In addition, proximity devices are valuable when detecting objects moving at high speed, when physical contact may cause damage, or when differentiation between metallic and nonmetallic items is required. Ball (1986), Johnson (1987), and Wojcik (1994) provide general overviews of various alternative proximity sensor types with suggested guidelines for selection.

5.5.1 Triangle sensor, Time-Of-Flight Sensors, Laser-Range Radar, Laser interferometric distance meter, Laser-Doppler Velocimeter

i. Optical Triangulation

Optical triangulation sensors use a light emitter, either a laser or an LED, in combination with a light receiver to sense the position of objects. Both the emitter and receiver are contained in the same housing as shown in Figure 5.62. The emitter directs light waves toward a target. These are reflected off the target, through a lens, to the receiver. The location of the incident light on the receiver is used to determine the position of the target in relation to the sensor face. The type of receiver used may be a position sensitive detector (PSD) or a pixelized array device such as a charge coupled device (CCD). The PSD receiver generates a single analog output and has a faster response time than the output pixelized array device because less post-processing is required. It is also typically smaller so that the overall sensor size will be smaller. Pixelized array devices, however, are useful when the surface of the target is irregular or transparent.

Triangulation is also a simple conventional technique used in a wide range of distance measurements, from 1 cm to 100 m. There are several versions of triangulation for distance measurements. The simple sensing system for triangulation can be constructed by combining only a light-emitting diode as a light source and a linear photo-detector (PD) array. Image-detector and light-beam scanning can also be used for obtaining distant images.

Important specifications for this type of sensor are the working range and the standoff distance. The standoff distance is the distance from the sensor face to the center of the working range. Both diffuse and specular designs are available for this type of sensor. A diffuse design is useful for targets with surfaces which scatter light such as anodized aluminum. A specular design is useful for targets with surfaces which reflect light well such as mirrors. In addition, the target color and transparency should also be considered when investigating this type of sensor because these properties affect the absorption of light by the target. Optical triangulation sensors are high resolution and typically offer ranges up to one half meter. Higher ranges can be achieved, albeit with significantly increased cost.

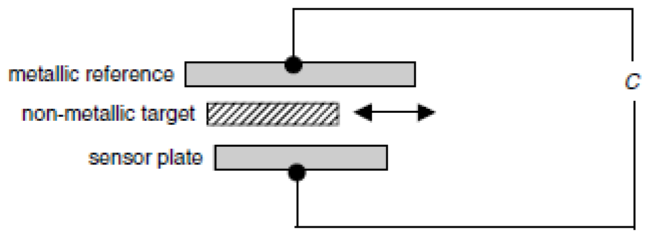


Figure 5.62: Dielectric variation in capacitive sensor measurement

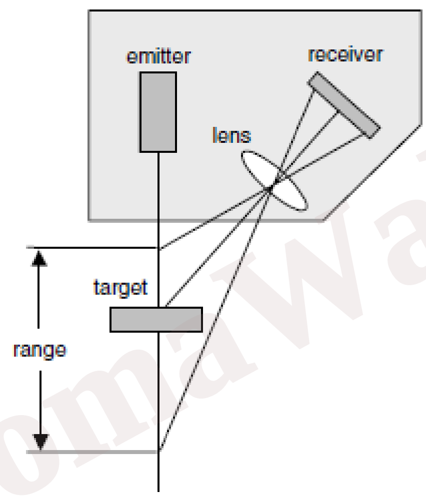


Figure 5.63: Optical triangulation displacement sensor

Triangulation Ranging

Triangulation ranging is based upon an important premise of plane trigonometry, which states that given the length of a side and two angles of a triangle, it is possible to determine the length of the other sides and the remaining angle. The basic Law of Sines can be rearranged as shown below to represent the length of side B as a function of side A and the angles θ and $\emptyset$.

In ranging applications, length B would be the desired distance to the object of interest at point P_3 (Fig. 5.64) for known sensor separation baseline A. Triangulation ranging systems are classified as either passive (use only the ambient light of the scene) or active (use an energy source to illuminate the target). Passive stereoscopic ranging systems position directional detectors (video cameras, solid-state imaging arrays, or position sensitive detectors) at positions corresponding to locations P_1 and P_2 (Fig. 5.65). Both imaging sensors are arranged to view the same object point, P3, forming an imaginary triangle. The measurement of angles θ and $\emptyset$ in conjunction with the known orientation and lateral separation of the cameras allows the calculation of range to the object of interest.

Active triangulation systems, on the other hand, position a controlled light source (such as a laser) at either point P_1 or P_2, directed at the observed point P_3. A directional imaging sensor is placed at the remaining triangle vertex and is also aimed at P_3. Illumination from the source will be reflected by the target, with a portion of the returned energy falling on the detector. The lateral position of the spot as seen by the detector provides a quantitative measure of the unknown angle $\emptyset$, permitting range determination by the Law of Sines.

The performance characteristics of triangulation systems are to some extent dependent on whether the system is active or passive. Passive triangulation systems using conventional video cameras require special ambient lighting conditions that must be artificially provided if the environment is too dark. Furthermore, these systems suffer from a correspondence problem resulting from the difficulty in matching points viewed by one image sensor with those viewed by the other. On the other hand, active triangulation techniques employing only a single detector do not require special ambient lighting, nor do they suffer from the correspondence problem. Active systems, however, can encounter instances of no recorded strike because of specular reflectance or surface absorption of the light.

Limiting factors common to all triangulation sensors include reduced accuracy with increasing range, angular measurement errors, and a missing parts (also known as shadowing) problem. Missing parts refers to the scenario where particular portions of a scene can be observed by only one viewing location (P_1 or P_2). This situation arises because of the offset distance between P_1 and P_2, causing partial occlusion of the target (i.e., a point of interest is seen in one view but otherwise occluded or not present in the other). The design of triangulation systems must include a tradeoff analysis of the offset: as this baseline measurement increases, the range accuracy increases, but problems due to directional occlusion worsen.

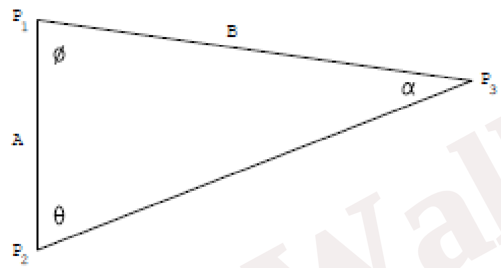


Figure 5.64: Triangulation ranging systems determine range B to target point P_3 by measuring angles $f\ and\ q$ at P_1 and P_2

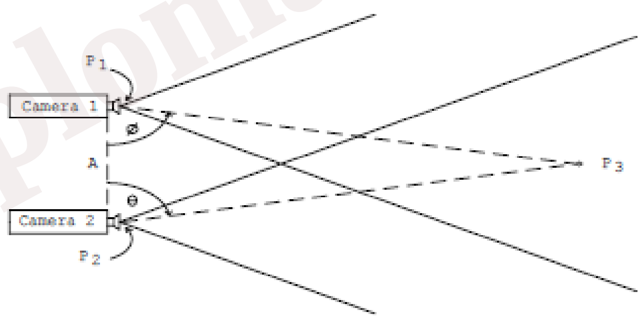


Figure 5.65: passive stereoscopic ranging system configuration

ii. Optical Time-Of-Flight

Optical time-of-flight sensors detect the position of objects by measuring the time it takes for light to travel to the object and back. As in the case of optical triangulation sensors, time-of-flight sensors also contain an emitter and a receiver. The emitter is a laser or an LED while the receiver is a photodiode. The emitter generates one or more pulses of light toward a target. Some of the light is reflected off the target and captured by the photodiode. The photodiode generates a pulse when it receives the light, and the time difference between the emitted pulse and the received pulse is determined by the sensor electronics. The distance to the target is then calculated based on the speed of light and the time difference.

Most time-of-flight sensors have measurement ranges of several meters. However, laser based time off light sensors can have a range of several miles if a series of pulses from the emitter is used. The accuracy of these sensors are not as high as optical triangulation sensors but the range is typically greater.

Ranging by Time-of-Flight (TOF)

Time-of-flight (TOF) is illustrated in Fig. 5.67 and 5.68. A gated wave (a burst of a few cycles) is emitted, bounced back from the target, and detected at the receiver located near the emitter. The emitter and receiver may physically be both one sensor. The receiver may also be mounted on the target. The TOF is the time elapsed from the beginning of the burst to the beginning of the return signal. The distance is defined as $d = c$. TOF/2 when emitter and receiver are at the same location, or $d = c$. TOF when the receiver is attached to the target. The accuracy is usually 1/4 of the wavelength when detecting the return signal, as its magnitude reaches a threshold limit. Gain is automatically increased with distance to maintain accuracy. Accuracy may be improved by detecting the maximum amplitude, as shown in Fig.5.69. This makes detecting the time of arrival of the wave less dependent on the amplitude of the signal. Ultrasonic, RF, or optical energy sources are typically employed; the relevant parameters involved in range calculation, therefore, are the speed of sound in air (roughly 0.305 m/ms), and the speed of light (0.305 m/ns). Potential error sources for TOF systems include the following:

- Variations in the speed of propagation, particularly in the case of acoustical systems
- Uncertainties in determining the exact time of arrival of the reflected pulse (Figueroa & Lamancusa,1992)
- Inaccuracies in the timing circuitry used to measure the round-trip time of flight
- Interaction of the incident wave with the target surface

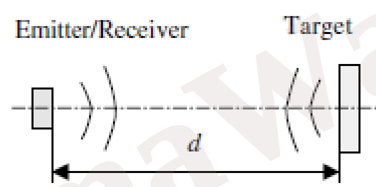


Figure 5.67: A wave is emitted and bounced from a target object.

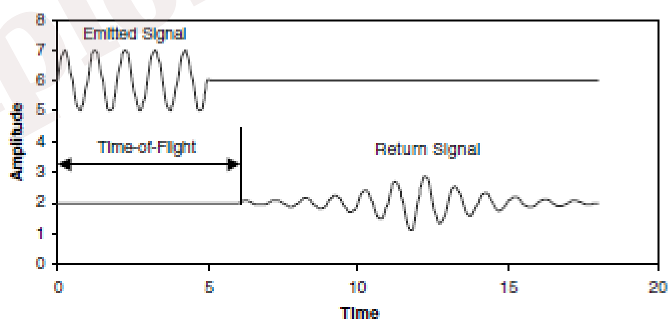


Figure 5.68: Definition of Time-of-Flight

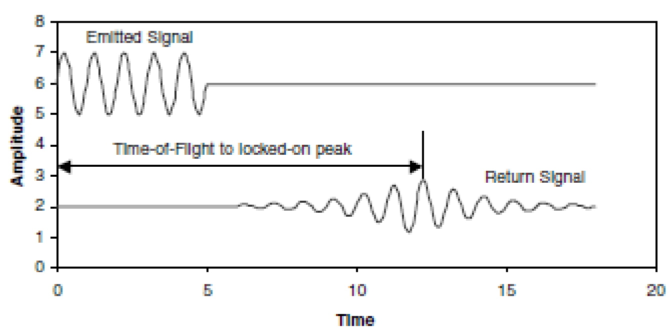


Figure 5.69: TOF to the maximum amplitude of the received signal for improved accuracy

Propagation Speed —For most applications, changes in the propagation speed of electromagnetic energy are for the most part inconsequential and can basically be ignored, with the exception of satellite-based position-location systems. This is not the case, however, for acoustically based systems, where the speed of sound is markedly influenced by temperature changes, and to a lesser extent by humidity. (The speed of sound is actually proportional to the square root of temperature in degrees Rankine; an ambient temperature shift of just 30°can cause a 1-ft error at a measured distance of 35 ft.)

Detection Uncertainties —So-called *time-walk errors* are caused by the wide dynamic range in returned signal strength as a result of (1) varying reflectivity of target surfaces, and (2) signal attenuation to the fourth power of distance due to spherical divergence. These differences in returned signal intensity influence the rise time of the detected pulse, and in the case of fixed-threshold detection will cause the less reflective targets to appear further away (Lang et al., 1989). For this reason, *constant fraction timing discriminators* are typically employed to establish the detector threshold at some specified fraction of the peak value of the received pulse (Vuylsteke et al., 1990; Figueroa & Doussis, 1993).

Timing Considerations —The relatively slow speed of sound in air makes TOF ranging a strong contender for low-cost acoustically based systems. Conversely, the propagation speed of electromagnetic energy can place severe requirements on associated control and measurement circuitry in optical or RF implementations. As a result, TOF sensors based on the speed of light require sub-nanosecond timing circuitry to measure distances with a resolution of about a foot (Koenigsburg, 1982). More specifically, a desired resolution of 1 mm requires a timing accuracy of 3 ps (Vuylsteke et al., 1990). This capability is somewhat expensive to realize and may not be cost effective for certain applications, particularly at close range where high accuracies are required.

Surface Interaction —When light, sound, or radio waves strike an object, any detected echo represents only a small portion of the original signal. The remaining energy reflects in scattered directions and can be absorbed by or pass through the target, depending on surface characteristics and the angle of incidence of the beam. Instances where no return signal is received at all can occur because of specular reflection at the object surface, especially in the ultrasonic region of the energy spectrum. If the transmission source approach angle meets or exceeds a certain critical value, the reflected energy will be deflected outside the sensing envelope of the receiver. Scattered signals can reflect from secondary objects as well, returning to the detector at various times to generate false signals that can yield questionable or otherwise noisy data. To compensate, repetitive measurements are usually averaged to bring the signal-to-noise ratio within acceptable levels, but at the expense of additional time required to determine a single range value.

iii. Laser Range Radar

Laser Range Radar (Laser-based TOF ranging systems), also known as *laser radar* or *lidar*, first appeared in work performed at the Jet Propulsion Laboratory, Pasadena, CA, in the 1970s (Lewis & Johnson, 1977). Laser energy is emitted in a rapid sequence of short bursts aimed directly at the object being ranged. The TOF of a given pulse reflecting off the object is used to calculate distance to the target based on the speed of light. Accuracies for early sensors of this type could approach a few centimeters over the range of 1–5 m (NASA, 1977; Depkovich & Wolfe, 1984).

Schwartz Electro-Optics, Inc. (SEO), Orlando, FL, produces a number of laser TOF range finding systems employing an innovative *time-to-amplitude-conversion* scheme to overcome the sub-nanosecond timing requirements necessitated by the speed of light. As the laser fires, a precision film capacitor begins discharging from a known set point at a constant rate, with the amount of discharge being proportional to the round-trip time-of-flight (Gustavson & Davis, 1992). An analog-to-digital conversion is performed on the sampled capacitor voltage; at the precise instant a return signal is detected, whereupon the resulting digital representation is converted to range and time-walk corrected using a look-up table.

The *LRF-X* series rangefinder shown in Fig. 5.70 features a compact size, high-speed processing, and an ability to acquire range information from most surfaces (i.e., minimum 10% Lambertian reflectivity) out to a maximum of 100 m. The basic system uses a pulsed InGaAs laser diode in conjunction with an avalanche photodiode detector and is available with both analog and digital (RS-232) outputs.

RIEGL Laser Measurement Systems, Horn, Austria, offers a number of commercial products (i.e., laser binoculars, surveying systems, “speed guns,” level sensors, profile measurement systems, and tracking laser scanners) employing short-pulse TOF laser ranging. Typical applications include lidar altimeters, vehicle speed measurement for law enforcement, collision avoidance for cranes and vehicles, and level sensing in silos.

The RIEGL *LD90-3 series* laser rangefinder (Fig. 5.71) employs a near-infrared laser diode source and a photodiode detector to perform TOF ranging out to 500 m with diffuse surfaces, and to over 1000 m in the case of cooperative targets. Round-trip propagation time is precisely measured by a quartz-stabilized clock and converted to measured distance by an internal microprocessor, using one of two available algorithms. The *clutter suppression* algorithm incorporates a combination of range measurement averaging and noise rejection techniques to filter out backscatter from airborne particulates, and is, therefore, useful when operating under conditions of poor visibility (Riegel, 1994). The *standard measurement* algorithm, on the other hand, provides rapid range measurements without regard for noise suppression, and can subsequently deliver a higher update rate under more favorable environmental conditions.

Figure 5.70: The LRF-200 series range finder

Figure 5.71: The class 1 (eye-safe) LD90-3 series TOF laser rangefinder is a self-contained unit available in several versions with maximum ranges of 150-1500 m under average atmospheric conditions.

iv. Laser Interferometry distance meter

Laser interferometers are capable of measuring incremental linear motions with resolution on the order of nanometers. In an interferometer, collimated laser light passes through a beam-splitter, sending the light energy on two different paths. One path is directly reflected to the detector, such as an optical sensing array, giving a flight path of fixed length. The other path reflects back to the detector from a retroreflector (mirror) attached to the target to be measured. The two beams constructively or destructively interfere with each other at the detector, creating a pattern of light and dark fringes. The interference pattern can be interpreted to find the phase relationship between the two beams, which depends on the relative lengths of the two paths, and therefore the distance to the moving target. As the target moves, the pattern repeats when the length of the variable path changes by the wavelength of the laser. Thus the laser interferometer is inherently an incremental measuring device. Laser interferometers are easily the most expensive sensors discussed in this chapter. They also have the highest resolution. Laser interferometers are very sensitive to mechanical misalignment and vibrations.

Laser interferometers provide the most accurate position feedback for servosystems. They offer very high resolution (to 5.72 nm), noncontact measurement, a high update rate, and intrinsic accuracies of up to 0.02 ppm. They can be used in servosystems either as passive position readouts or as active feedback sensors in a position servo loop. The laser beam path can be precisely aligned to coincide with the load or a specific point being measured, eliminating or greatly reducing Abbe error. A single-axis system based on the Michaelson interferometer is illustrated in Figure 5.72. It consists of a helium–neon laser, a polarizing beam splitter with a stationary retroreflector, a moving retroreflector that can be mounted on the object whose position is to be measured, and a photodetector, typically a photodiode.

Light from the laser is directed toward the polarizing beam splitter, which contains a partially reflecting mirror. Part of the laser beam goes straight through the polarizing beam splitter, and part of the laser beam is reflected. The part that goes straight through the beam splitter reaches the moving reflectometer, which reflects it back to the beam splitter that passes it on to the photodetector. The part of the beam that is reflected by the beam splitter reaches the stationary retroreflector, a fixed distance away. The retroreflector reflects it back to the beam splitter before it is also reflected into the photodetector.

As a result, the two reflected laser beams strike the photodetector, which converts the combination of the two light beams into an electrical signal. Because of the way laser light beams interact, the output of the detector depends on a difference in the distances traveled by the two laser beams. Because both light beams travel the same distance from the laser to the beam splitter and from the beam splitter to the photodetector, these distances are not involved in position measurement. The laser interferometer measurement depends only on the difference in distance between the round trip laser beam travel from the beam splitter to the moving retroreflector and the fixed round trip distance of laser beam travel from the beam splitter to the stationary retroreflector.

If these two distances are exactly the same, the two light beams will recombine in phase at the photodetector, which will produce a high electrical output. This event can be viewed on a video display as a bright light fringe. However, if the difference between the distances is as short as one-quarter of the laser's wavelength, the light beams will combine out-of-phase, interfering with each other so that there will be no electrical output from the photodetector and no video output on the display, a condition called a dark fringe.

As the moving retroreflector mounted on the load moves farther away from the beam splitter, the laser beam path length will increase and a pattern of light and dark fringes will repeat uniformly. This will result in electrical signals that can be counted and converted to a distance measurement to provide an accurate position of the load. The spacing between the light and dark fringes and the resulting electrical pulse rate is determined by the wavelength of the light from the laser. For example, the wavelength of the light beam emitted by a helium–neon (He–Ne) laser, widely used in laser interferometers, is 0.63 μm, or about 0.000025 in. Thus the accuracy of load position measurement depends primarily on the known stabilized wavelength of the laser beam. However, that accuracy can be degraded by changes in humidity and temperature as well as airborne contaminants such as smoke or dust in the air between the beam splitter and the moving retroreflector.

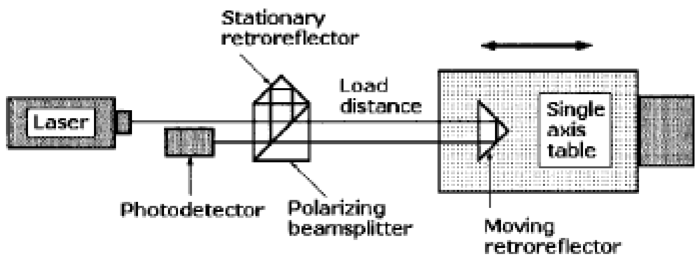


Figure 5.72: Diagram of a laser interferometer for position feedback that combines high resolution

v. Lasser Doppler Velocimetric

Laser Doppler velocimetry (LDV) , also known as laser Doppler anemometry (LDA), is the technique of using the Doppler shift in a laser beam to measure the velocity in transparent or semi-transparent fluid flows, or the linear or vibratory motion of opaque, reflecting, surfaces.

With the development of the helium-neon laser (He-Ne) at the Bell Telephone Laboratories in 1962, the optics community had available a source of continuous wave electromagnetic radiation highly concentrated at a wavelength of 632.8 nanometers (nm), in the red portion of the visible spectrum. It was soon shown fluid flow measurement could be made from the Doppler effect on a He-Ne beam scattered by very small polystyrene spheres entrained in the fluid. At the Research Laboratories of Brown Engineering Company (later Teledyne Brown Engineering), this phenomenon was used in developing the first laser Doppler flowmeter using heterodyne signal processing.

The instrument was soon called the Laser Doppler Velocimeter (LDV) and the technique Laser Doppler Velocimetry, also abbreviated LDV. Another application name is laser Doppler anemometry (LDA). Early LDV applications ranged from measuring and mapping the exhaust from rocket engines with speeds up to 1000 m/s to determining flow in a near-surface blood artery. A variety of similar instruments were developed for solid-surface monitoring, with applications ranging from measuring product speeds in production lines of paper and steel mills, to measuring vibration frequency and amplitude of surfaces.

Its simplest and most presently used form, LDV crosses two beams of collimated, monochromatic, and coherent laser light in the flow of the fluid being measured. The two beams are usually obtained by splitting a single beam, thus ensuring coherence between the two. Lasers with wavelengths in the visible spectrum (390-750 nm) are commonly used; these are typically He-Ne, Argon ion, or laser diode), allowing the beam path to be observed. A transmitting optics focuses the beams to intersect at their waists (the focal point of a laser beam), where they interfere and generate a set of straight fringes. As particles (either naturally occurring or induced) entrained in the fluid pass through the fringes, they reflect light that is then collected by a receiving optics and focused on a photodetector (typically an avalanche photodiode).

The reflected light fluctuates in intensity, the frequency of which is equivalent to the Doppler shift between the incident and scattered light, and is thus proportional to the component of particle velocity which lies in the plane of two laser beams. If the sensor is aligned to the flow such that the fringes are perpendicular to the flow direction, the electrical signal from the photodetector will then be proportional to the full particle velocity. By combining three devices (e.g.; He-Ne, Argon ion, and laser diode) with different wavelengths, all three flow velocity components can be simultaneously measured.

Another form of LDV, particularly used in early device developments, has a completely different approach akin to an interferometer. The sensor also splits the laser beam into two parts; one (the measurement beam) is focused into the flow and the second (the reference beam) passes outside the flow. A receiving optics provides a path that intersects the measurement beam, forming a small volume. Particles passing through this volume will scatter light from the measurement beam with a Doppler shift; a portion of this light is collected by the receiving optics and transferred to the photodetector. The reference beam is also sent to the photodetector where optical heterodyne detection produces an electrical signal proportional to the Doppler shift, by which the particle velocity component perpendicular to the plane of the beams can be determined. Similar arrangements using optical heterodyning are also used in laser Doppler sensors for measuring the linear velocity of solids and for measuring vibrations of surfaces; the latter sensor is usually called a laser Doppler vibrometer, also abbreviated LDV.

Laser Doppler velocimetry is often chosen over other forms of flow measurement because the equipment can be outside of the flow being measured and therefore has no effect on the flow. Some typical applications include the following:

Laser Doppler velocimetry is effective in measuring surface vibrations via reflection of the laser light from the vibrating surface. The technology, adapted to include a scanning capability (to provide measurement of the vibration over an array of points, as in the Polytec MSA-500 and Aries Laser Vibrometer, VELA), has been used to measure vibration generation and propagation for ultrasonic motors[11] and acoustic and ultrasonic microfluidics. Remarkably, it is possible to measure the deformation of capillary waves as well using a laser Doppler vibrometer.

Laser Doppler velocimeter (LDV) can be configured to measure any desired component velocity, perpendicular or parallel to the direction of the optical axis. An LDV system has been constructed with a semiconductor laser and optical fibers and couplers to conduct the optical power. Frequency modulation of the semiconductor laser (or, alternatively, an external fiber-optic frequency modulator) is used to introduce an offset frequency. Some commercial laser Doppler velocimeters are available with optical-fiber leads and small sensing heads. However, these commercial systems still use bulk optical components such as acoustooptic modulators or rotating gratings to introduce the offset frequency.

With an LDV system, the velocity can be measured with high precision in a short period of time. This means that the method can be applied for real-time measurements to monitor and control the velocity of objects as well as measure their vibration. Because the laser light can be focused to a very small spot, the velocity of very small objects can be measured, or if scanning techniques are applied, high spatial resolution can be achieved. This method is used for various applications in manufacturing, medicine, and research. The demands on system performance with respect to sensitivity, measuring range, and temporal resolution are different for each of these applications.

In manufacturing processes, for example, LDV systems are used to control continuous roll milling of metal (Fig. 5.73), to control the rolling speed of paper and films, and to monitor fluid velocity and turbulence in mixing processes. Another industrial application is vibration analysis. With a noncontact vibrometer, vibration of machines, machine tools, and other structures can be analyzed without disturbing the vibrational behavior of the structure. Interestingly, the LDV system proved useful in the measurement of arterial blood velocity (Fig. 6.28), thereby providing valuable medical information. Another application in medical research is the study of motion of the tympanic membrane in the ear.

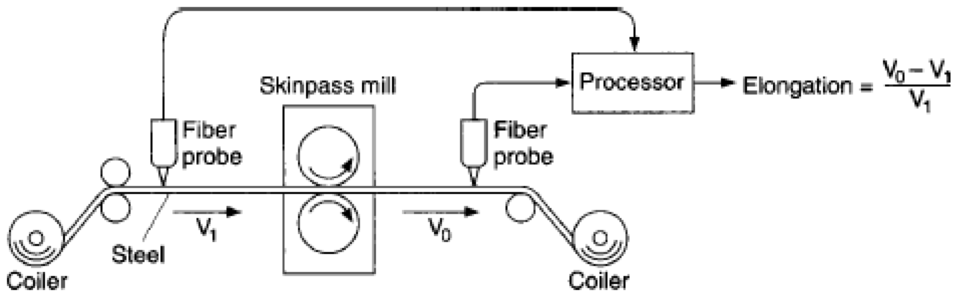


Figure 5.73: Fiber-optic Doppler velocimeter at a rolling mill controls pressure by measuring inputs speeds

END OF CHAPTER 5

1. Describe meaning of sensor.

2. Explain the functions of a sensor.

3. List three types of internal sensor and three types of external sensor.

4. How does Range sensor work?

5. Explain about analog displacement sensor and digital sensor.

6. Name two types of inputs that would be analog input values (versus a digital value).

7. What is the resolution of an absolute optical encoder that has six tracks? nine tracks? Twelve tracks?

8. What are the following: Shaft encoder, Incremental encoder and Absolute encoder.

9. Define the following term for sensor
   i. Range
   ii. Sensitivity
   iii. Error
   iv. Dynamic response
   v. Resolution

10. In general, how do sensor pilot devices operate?

11. What is the main feature of a proximity sensor?

12. List the main components of an inductive proximity sensor.

13. Explain the term *hysteresis* as it applies to a proximity sensor.

14. How a two-wire sensor is connected relative to the load it controls?

15. In what way is the sensing field of a capacitive proximity sensor different from that of the inductive proximity sensor?

16. For what type of target would a capacitive proximity sensor be selected over an inductive type?

17. Outline the principle of operation of a photoelectric sensor.

18. Name the three most common scan techniques for photoelectric sensors.

19. What are the advantages of fiber optic sensing systems?

20. Outline the principle of operation of a Laser Range radar sensor.

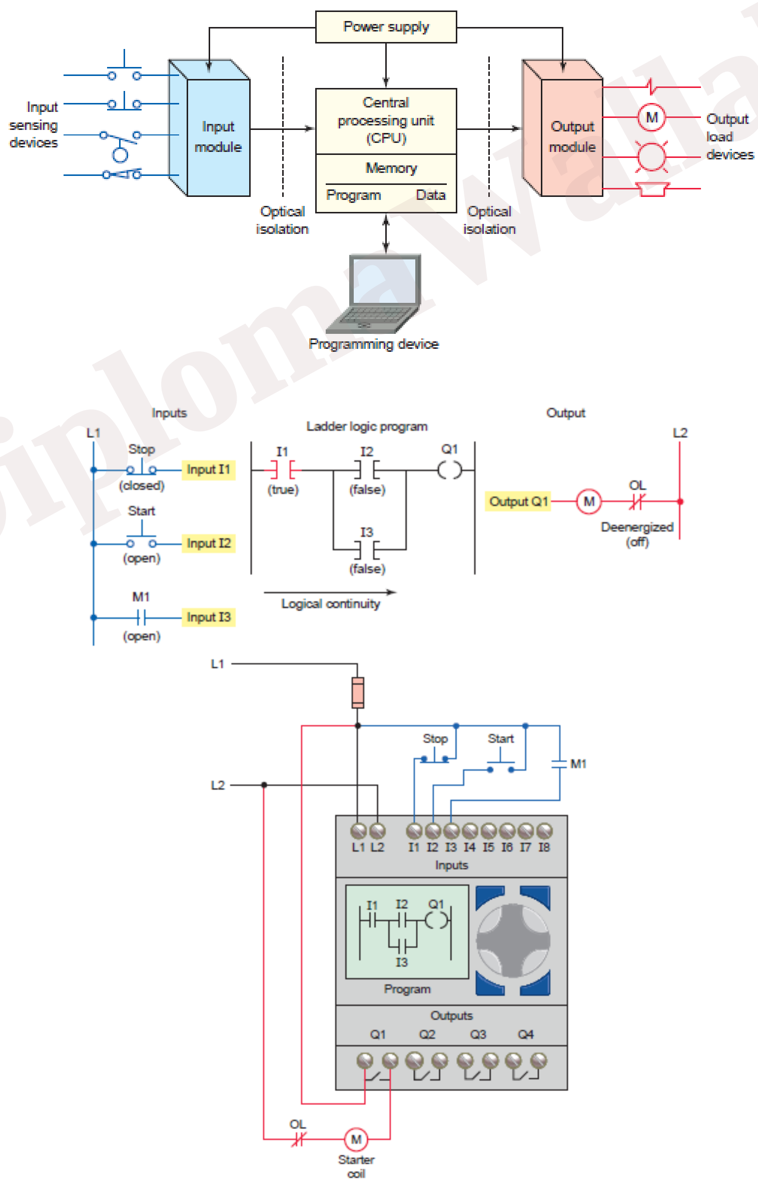
21. Compare the way in which a tachometer and magnetic pickup are used in speed measurement.

22. Named 4 types of sensor for used in angular position measurement.
23. Explain how linear variable differential transformer (LVDT) operates?
24. Explain about the principles of optical sensor and named three examples of optical sensor?
25. Explain about the principles of encoder scheme.
26. Explain how velocity sensors work and named three examples of velocity sensor.
27. Explain how the strain gages sensor works.
28. What is a strain gauge?
29. What are the types of strain gauges?
30. Give other classification of strain gauges
31. What is a Resistance Strain Gauge?
32. What is a semiconductor strain gauge?
33. Give Gauge Factor for few materials.
34. What are the points to be considered for selecting sensor?
35. What are the different types of strain gauges? Name four resistance materials used in wire and foil gauges.
36. Mention desirable characteristics of strain gauges.
37. Name two light sensitive sensor
38. What is Tachogenerator?
39. Mention major applications of L.V.D.T.s?
40. Enlist advantages' of LVDT.
41. Mention one disadvantage of LVDT.
42. Explain the main functions for micromechanical angular velocity.
43. Explain the concept of contact sensor and named three types of contact sensor.
44. Explain about the Lesser Doppler Velocimeter.
45. Give comparison between Active and Passive transducers.

CHAPTER 6
DESIGN AN EXAMPLE FOR INDUSTRIAL AUTOMATION SYSTEM

Upon completion of this course, students should be able to:-

- Discuss the Automation Design and process specifications
- Describe the Encoder Selection
- Develop the Control Structure: Programmable Logic Controller used for Industrial Automation



6.1 Automation Design and process specifications

6.1.1 System Specifications

An *automated assembly system* performs a sequence of automated assembly operations to combine multiple components into a single entity. The single entity can be a final product or a subassembly in a larger product. In many cases, the assembled entity consists of a base part to which other components are attached. The components are joined one at a time (usually), so the assembly is completed progressively. A typical automated assembly system consists of the following subsystems; (1) one or more workstations at which the assembly steps are accomplished, (2) parts feeding devices that deliver the individual components to the workstations, and (3) a work handling system for the assembled entity. In assembly systems with one workstation, the work handling system moves the base part into and out of the station. In systems with multiple stations, the handling system transfers the partially assembled base part between stations. Control functions required in automated assembly machines are the same as in the automated processing lines: (I) sequence control, (2) safety monitoring, and (3) quality control. The issue of memory control versus instantaneous control is especially relevant in multi-station automated assembly systems.

Automated assembly systems can be classified according to physical configuration. The principal configurations, illustrated in Table 6.1, are: (a) in-line assembly machine, (b) dial-type assembly machine, (c) carousel assembly system, and (d) single station assembly machine. The *in-line assembly machine,* Table 6.1 (a), consists of a series of automatic workstations located along an in-line transfer system. It is the assembly version of the machining transfer line. Synchronous and asynchronous transfer systems are the common means of transporting base pans from station-to-station with the in-line configuration.

In the typical application of the *dial-type machine,* Table 6.1(b), base parts art: loaded onto fixtures or nests attached to the circular dial. Components ate added and/or joined to the base part at the various workstations located around the periphery of the dial. The dial indexing machine operates with a synchronous or intermittent motion, in which the cycle consists of the service time plus indexing time. Dial-type assembly machines are sometimes designed to use a continuous rather than intermittent motion. This is common in beverage bottling and canning plants, but it is not common in mechanical and electronics assembly.

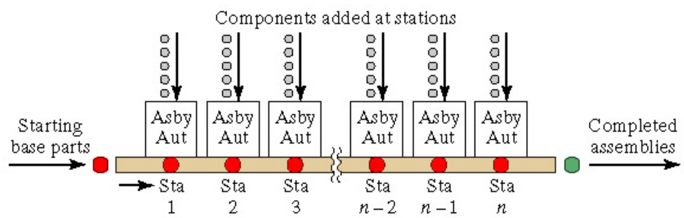
The operation of dial-type and in-line assembly systems is similar to the operation of their counterparts for processing operations, except that assembly operations are performed. For synchronous transfer of work between stations, the ideal cycle time equals the operation time at the slowest station plus the transfer time between stations. The production rate, at 100% uptime, is the reciprocal of the ideal cycle time. Because of part jams at the workstations and other malfunctions, the system operates at less than 100W, uptime. As seen in Table 6.1(c), the *carousel assembly system* represents a hybrid between the circular work flow of the dial assembly machine and the straight work flow of the online system. The carousel configuration can be operated with continuous, synchronous, or asynchronous transfer mechanisms to move the work around the carousel. Carousels with asynchronous transfer of work are often used in partially automated assembly systems, in the *single station assembly machine.* Table 6.1(d), assembly operations are performed on a base part at a single location. The typical operating cycle involves the placement of the base part at a stationary position in the workstation, followed by the addition of components to the base and finally the removal of the completed assembly from the station.

An important application of single station assembly is the component insertion machine, widely used in the electronics industry to populate components onto printed circuit boards. For mechanical assemblies, the single station cell is sometimes selected as the configuration for robotic assembly applications. Parts are fed into the single station, and the robot adds them to the base part and performs the fastening operations.

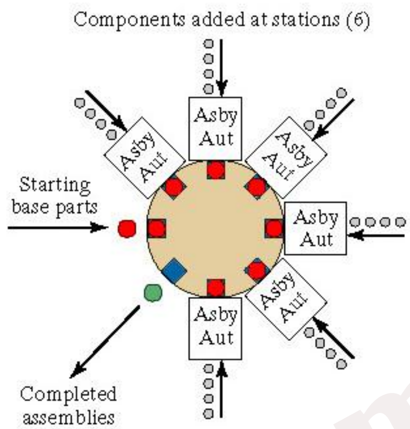
Compared with the other three system types, the single station system is inherently slower, since all of the assembly tasks are performed and only one assembled unit is completed each cycle.

Table 6.1: System configurations for Automated Assembly Systems

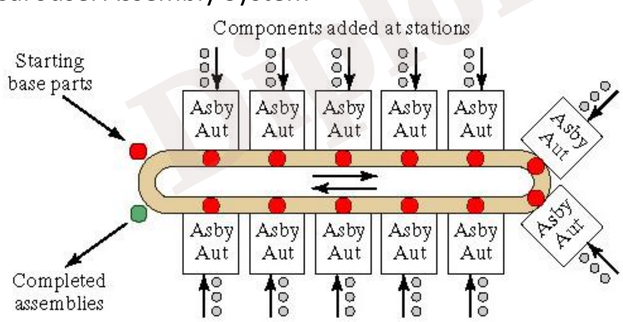
<table>
<tr><th>Configuration</th><th>Description</th></tr>
<tr><td>In-line



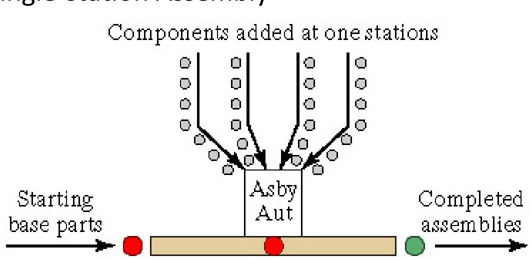
</td><td>A series of automatic workstations located along an in-line transfer system—the assembly version of the machining transfer line. Synchronous and asynchronous transfer systems may be used to transport parts from workstation to workstation.</td></tr>
<tr><td>Dial-type



</td><td>Base parts loaded onto fixtures or nests around the periphery of the circular dial, and—as the dial table turns—components are assembled sequentially onto the base part. Synchronous transfer system in operation, as all nests move at the same time, sometimes through continuous motion, but more often intermittently.</td></tr>
<tr><td>Carousel Assembly System



</td><td>Represents a hybrid between the circular work flow of the dial-type assembly machine, and the straight work flow of the in-line system. Carousels can be operated with continuous, synchronous, or asynchronous transfer mechanisms.</td></tr>
<tr><td>Single-station Assembly



</td><td>Consists of one workstation where components are assembled, successively, onto a base part that has entered the system. Once all the components have been assembled onto the base part, the base part leaves the system. Inherently slower than the other three system configurations, as only one base part is processed at a time.</td></tr>
</table>

6.1.2 Mechanical Description of the Automation

In each of the configurations described above, a workstation accomplishes one or both of the following tasks: (1) a part is delivered to the assembly workhead and added to the existing base part in front of the workhead (in the case of the first station in the system, the base part is often deposited into the work carrier), and/or (2) a fastening or joining operation is performed at the station in which parts added at the workstation or at previous workstations are permanently attached to the existing base part. In the case of a single station assembly system, these tasks are carried out multiple times at the single station. For task (1), a means of delivering the parts to the assembly workhead must be designed. The parts delivery system typically consists of the following hardware:

1. *HOPPER.* This is the container into which the components are loaded at the workstation. A separate hopper is used for each component type. The components are usually loaded into the hopper in bulk. This means that the parts are initially randomly oriented in the hopper.

2. Parts feeder. This is a mechanism that removes the components from the hopper one at a time for delivery to the assembly workhead. The hopper and parts feeder are often combined into one operating mechanism. A vibratory bowl feeder, pictured in Figure 6.1 is a very common example of the hopper-feeder combination.

3. *Selector and/or orientor*. These elements of the delivery system establish the proper orientation of the components for the assembly workhead. A selector is a device that's act as a filter, permitting only parts in the correct orientation to pass through. In correctly oriented parts are rejected back into hopper. An orientor is a device that allows properly oriented parts pass through, but it reorients parts that are not properly oriented initially. Several selector and orientor schemes are illustrated in Figure 6.2.

4. Feed track. The preceding elements of the delivery system are usually separated from the assembly workhead by a certain distance. *A feed track* is used to move the components from the hopper and parts feeder to the location of the assembly workhead, maintaining proper orientation of the parts during the transfer. There are two general categories of feed tracks: gravity and powered. Gravity feed tracks are the most common. In this type, the hopper and parts feeder are located at an elevation above that of the workhead. The force of gravity is used to deliver the components to the workhead. The powered feed track uses vibratory action, air pressure, or other means to force the parts to travel along the feed track toward the assembly workhead.

5. Escapement and *placement device.* The purpose of the *escapement device* is to remove components from the feed track at time intervals that are consistent with the cycle time of the assembly workhead. The *placement device* physically places the component in the correct location at the workstation for the assembly operation. These elements are sometimes combined into a single operating mechanism. In other cases, they are two separate devices. Several types of escapement and placement devices are pictured in Table 6.3.

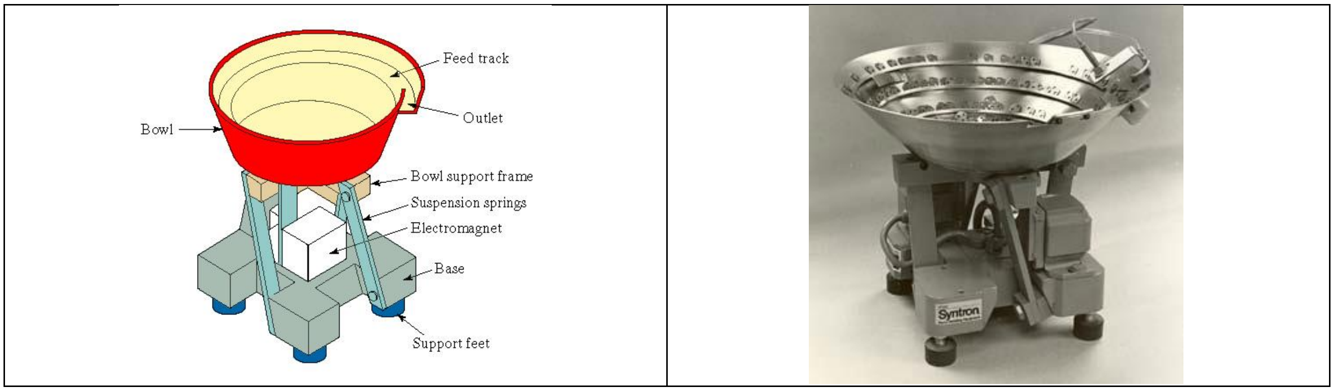


Figure 6.1: Hopper and parts feeder

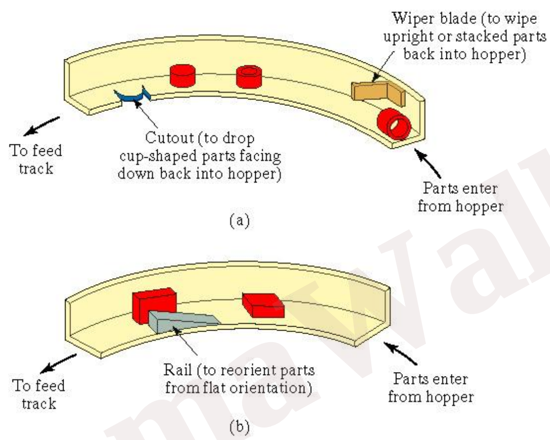


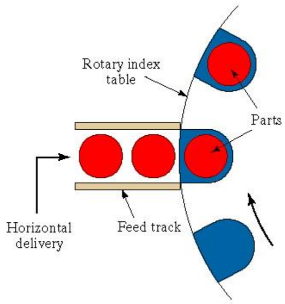
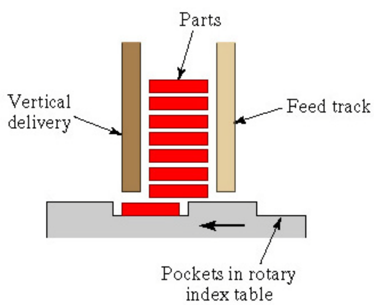
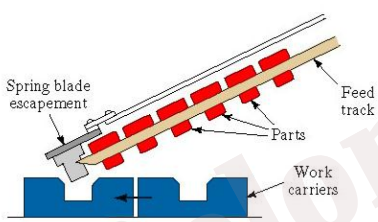
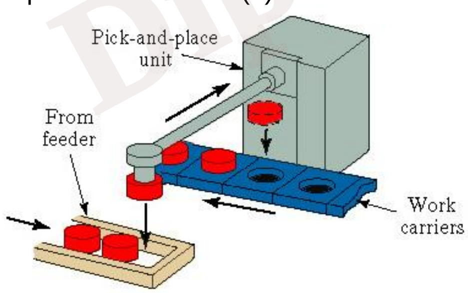
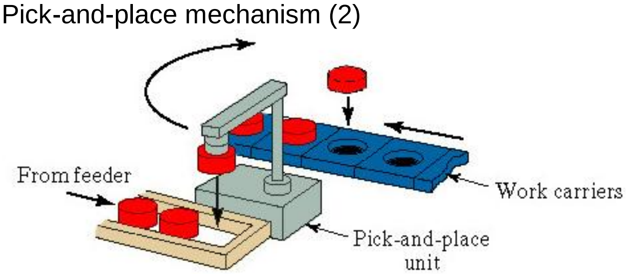
Figure 6.2: (a) Selector and (b) orientor devices used upon the feedtrack

The hardware elements of the parts delivery system are illustrated schematically in Figure 19.5.A parts selector is illustrated in the diagram. Improperly oriented parts are fed back into the hopper. In the case of a parts orientor, improperly oriented parts are reoriented and proceed to the feed track. A more detailed description of the various elements of the delivery system is provided in.

One of the recent developments in the technology of parts feeding and delivery systems is the programmable parts feeder. A *programmable parts feeder* is capable of feeding components of varying geometries with only a few minutes required to make the adjustments (change the program) for the differences. The flexibility of this type of feeder permits it to be used in batch production or when product design changes occur. Most parts feeders are designed as fixed automated systems for high production assembly of stable product designs.

Automated assembly systems are used to produce a wide variety of products and subassemblies. It should be noted that certain assembly processes are more suitable for automation than are others. For example, threaded fasteners (e.g., screws, bolts, and nuts), although common in manual assembly, are a challenging assembly method to automate. This issue, along with some guidelines for designing products for automated assembly, is discussed in the following section.

Table 6.3: Escapement and placement devices

Horizontal placement device 	Device used on dial-type assembly machines: parts move via horizontal delivery into vacant nests on the dial, as they appear, from the feed track; meanwhile the circular motion of the dial table means that the nests are revolved away from the feed track, permitting the next component in the feed track to move into the next vacant nest, and so forth.
Vertical placement device 	Device used on dial-type assembly machines: here, the parts feeder is arranged vertically above the dial table, so that when the table turns, to reveal an empty nest, the component can fall by gravity from the feed track into the empty nest. Successive parts fall by gravity to take up their position at the mouth of the feed track in turn.
Escapement device 	This device is actuated by the top of the carrier contacting the lower surface of the rivet-shaped part, causing its upper surface to press against the spring blade, which releases the part so that it falls into the work carrier nest. The work carriers are moved horizontally to cause the release of the part, and—after the first part has escaped—the work carrier and released part move off, to be replaced by the next work carrier, and so forth.
Pick-and-place mechanism (1) 	This mechanism uses a pick-and-place unit with a horizontal arm that may be extended and retracted as necessary, so that parts may be removed from the feed track, and placed into work carriers.
Pick-and-place mechanism (2) 	This mechanism uses a pick-and-place unit with a revolving arm, so that parts may be removed from the feed track, and placed into work carriers.

6.1.3 Motion Sequence

One of the obstacles to automated assembly is that many of the traditional assembly method evolved when humans were the only available means of assembling a product. Many of the mechanical fasteners commonly used in industry today require the special anatomical and sensory capabilities of human beings. Consider for example, the use of a bolt, lock washer and nut to fasten two sheet metal parts on a partially assembled cabinet. This kind of operation is commonly accomplished manually at either a single assembly station or on an assembly line. The cabinet is positioned at the workstation with the two sheet metal parts to be fastened at an awkward location for the operator to reach. The operator picks up the bolt, lock-washer and nut, somehow manipulating them into position on opposite sides of the two parts, and places the lock-washer and then the nut onto the bolt. As luck would have it, the threads of the nut initially bind on the bolt threads, and so the operator must unscrew slightly and restart the process, using a well-developed sense of touch to ensure that the threads are matching. Once the bolt and nut have been tightened with fingers, the operator reaches for the appropriate screwdriver (there are various bolt sizes with different heads) to tighten the fastener

This kind of manual operation has been used commonly and successfully in industry for many years to assemble products. The hardware required is inexpensive, the sheet metal is readily perforated to provide the matching clearance holes, and the method lends itself to field service, what is becoming very expensive is the manual labor at the assembly workstation required to accomplish the initial fastening. The high cost of manual labor has resulted in a reexamination of assembly technology with a view toward automation. However, automating the assembly operation just described is very difficult. First, the positions of the holes through which the bolt must he inserted are different for each fastener, and some of the positions may be difficult for the operator to reach. Second, the holes between the two sheet metal parts may not match up perfectly, requiring the operator to reposition the two parts for a better fit. Third, the operator must juggle three separate hardware items (bolt, lock-wash and nut) to perform the fastening operation. And the part to be fastened may also have to be included in the juggling act. Fourth, a sense of touch is required to make sure that the nut is started properly onto the bolt thread. Each of these four problems makes automation of the operation difficult. All four problems together make it nearly impossible, As a consequence, attempts at assembly automation have led to an examined, none of the methods specified by the designer to fasten together the various components of a product.

The first and most general lesson, which is obvious from this last example, is that the methods traditionally used for manual assembly are not necessarily the best methods for automated assembly. Humans are the most dexterous and intelligent machines, able to move to different positions in the workstation, adapt to unexpected problems and new situations during the work cycle, manipulate and coordinate multiple objects simultaneously and make use of a wide range of senses in performing work. For assembly automation to be achieved, fastening procedures must be devised and specified during product design that does not require all of these human capabilities. The following are some recommendations and principles that can be applied in product design to facilitate automated assembly:

- *Reduce the amount of assembly required.* This principle can be realized during design by combining functions within the, same part that were previously accomplished by separate components in the product. The use of plastic molded parts to substitute for sheet metal parts may be a way to actualize this principle. A more complex geometry molded into a plastic part might replace several metal parts. Although the plastic part may seem to be more costly, the savings in assembly time will justify the substitution in many cases.

- *Use of modular design.* In automated assembly, increasing the number of separate assembly steps accomplished by a single automated system results in a decrease in system reliability. To reduce this effect Riley, suggests that the design of the product be modular with perhaps each module requiring a maximum of 12 or so parts to be assembled on a single assembly system. Also, the subassembly should be designed around a base part to which other components are added.

- *Reduce the number of fasteners required.* Instead of using separate screws, nuts and similar fasteners, design tile fastening mechanism into the component design using snap fits and similar features.

- *Reduce the need for multiple components to be handled at once.* The preferred practice in automated assembly machine design is to separate the operations at different stations rather than to simultaneously handle and fasten multiple components at the same workstation. For the case of the single station assembly system, this principle must be interpreted to mean that the handling of multiple components must be minimized in each assembly work clement.

- *Limit the required direction of access.* This principle simply means that the number of directions in which new components are added to the existing subassembly should be minimized. If all of the components can be added vertically from above, this is the ideal situation. Obviously the design of the subassembly determines this.

- *High quality required in components.* High performance of the automated assembly system requires consistently good quality of the components added at each workstation. Poor quality components cause jams in the feeding and assembly mechanisms which cause downtime in an automated system.

- *Hopperability. This* is a term that Riley uses to identity the case with which a given component can be fed and oriented reliably for delivery from the parts hopper to the assembly workhead. One of the major costs in the development of an automated assembly system is the engineering time to devise the means of feeding the components in the correct orientation for the assembly operation. The product designer is responsible for providing the orientation features and other geometric aspects of the components that determine the ease of feeding and orienting the parts.

6.1.4 Motor and Drive Mechanism Selection

The selection process often highlights difficulties in three areas. Firstly, as we have discovered in the preceding chapters, there is a good deal of overlap between the major types of motor and drive. This makes it impossible to lay down a set of hard and fast rules to guide the user straight to the best solution for a particular application. Secondly, users tend to underestimate the importance of starting with a comprehensive specification of what they really want, and they seldom realize how much weight attaches to such things as the steady-state torque–speed curve, the inertia of the load, the pattern of operation (continuous or intermittent) and the question of whether or not the drive needs to be capable of regeneration. And thirdly, they may be unaware of the existence of standards and legislation, and hence can be baffled by questions from any potential supplier. The whole business of selection is so broad that it really warrants a book to itself, but the cursory treatment here should at least help the user to specify the drive rating and arrive at a shortlist of possibilities.

i. Power range for motor and drives

The diagrams (Figures 6.3 and 6.4) give a broad indication of the power range for the most common types of motor and drive. Because the power scales are logarithmic it would be easy to miss the exceptionally wide power range of some types of motor: induction and d.c. motors, for example, extend from watts to megawatts, an astonishing range that few other inventions can match. The width of the bands is intended to give some idea of relative importance, while the shading reflects the fact that there is no sharp cut-off at the extremities of the range. We should also bear in mind that we are talking here about the continuously rated maximum power at the normal base speed, and as we have seen most motors will be able to exceed this for short periods, and also to run faster than base speed, provided that reduced torque is acceptable.

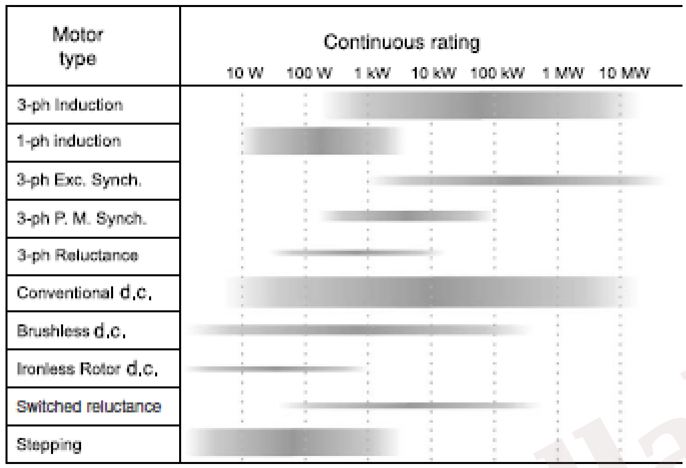


Figure 6.3: Continues power rating for various types of motor

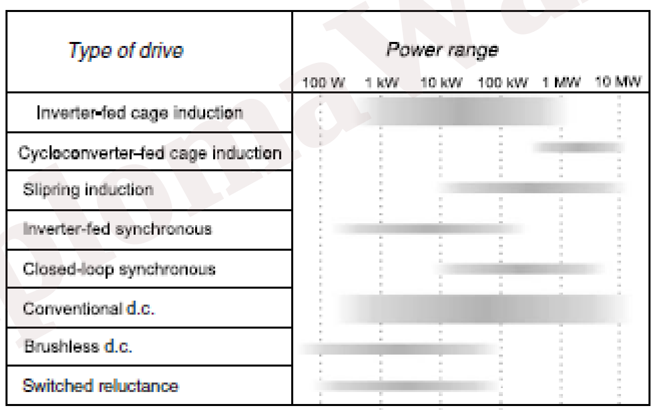


Figure 6.4: Power range for various types of drive

ii. Maximum speed and speed range

As a general rule, for a given power the higher the base speed the smaller the motor. In practice, there are only a few applications where motors with base speeds below a few hundred rev/ min are attractive, and it is usually best to obtain low speeds by means of the appropriate mechanical speed reduction.

Speeds over 10 000 rev/min are also unusual except in small universal motors and special-purpose inverter-fed motors. The majority of medium-size motors have base speeds between 1500 and 3000 rev/min. Base speeds in this range are attractive as far as motor design is concerned, because good power/weight ratios are obtained, and are also satisfactory as far as any mechanical transmission is concerned. In controlled-speed applications, the range over which the steady state speed must be controlled, and the accuracy of the speed holding, are significant factors in the selection process. In general, the wider the speed range, the more expensive the drive: a range of 10:1 would be unexceptional, whereas 100:1 would be demanding. Figures for accuracy of speed holding can sometimes cause confusion, as they are usually given as a percentage of the base speed. Hence with a drive claiming a decent speed holding accuracy of 0.2% and a base speed of 2000 rev/min, the user can expect the actual speed to be between 1996 and 2004 rev/min

when the speed reference is 2000 rev/min. But if the speed reference is set for 100 rev/min, the actual speed can be anywhere between 96 and 104 rev/min, and still be within the specification. For constant torque loads, which require operation at all speeds, the inverter-fed induction motor, the d.c. drive and any of the self-synchronous drives are possibilities, but only the d.c. drive would automatically come with a force-ventilated motor capable of continuous operation with full torque at low speeds.
Fan-type loads (see below) with a wide operating speed range are a somewhat easier proposition because the torque is low at low speeds. In the medium and low power ranges the inverter-fed induction motor (using a standard motor) is satisfactory, and will probably be cheaper than the d.c. drive. For restricted speed ranges (say from base speed down to 75%) and particularly with fan-type loads where precision speed control is unnecessary, the simple voltage-controlled induction motor is likely to be the cheapest solution.

iii. Load requirements –Torque – Speed Characteristics

The most important things we need to know about the load are the steady-state torque–speed characteristic, and the effective inertia as seen by the motor. In addition, we clearly need to know what performance is required. At one extreme, for example, in a steel-rolling mill, it may be necessary for the speed to be set at any value over a wide range, and for the mill to react very quickly when a new target speed is demanded. Having reached the set speed, it may be essential that it is held very precisely even when subjected to sudden load changes. At the other extreme, for example, a large ventilating fan, the range of set speed may be quite limited (perhaps from 80% to 100%); it may not be important to hold the set speed very precisely; and the time taken to change speeds, or to run-up from rest, are unlikely to be critical.

At full speed both of these examples may demand the same power, and at first sight might therefore be satisfied by the same drive system. But the ventilating fan is obviously an easier proposition, and it would be overkill to use the same system for both. The rolling mill would call for a regenerative d.c. or a.c. drive with tacho or encoder feedback, while the fan could quite happily manage with a cheaper open-loop inverter fed induction motor drive, or even perhaps a simple voltage-controlled induction motor. Although loads can vary enormously, it is customary to classify them into two major categories, referred to as 'constant-torque' or 'fan or pump' types. We will use the example of a constant-torque load to illustrate in detail what needs to be done to arrive at a specification for the torque–speed curve. An extensive treatment is warranted because this is often the stage at which users come unstuck.

iv. Constant-torque load

A constant torque load implies that the torque required to keep the load running is the same at all speeds. A good example is a drum-type hoist, where the torque required varies with the load on the hook, but not with the speed of hoisting. An example is shown in Figure 11.3. The drum diameter is 0.5 m, so if the maximum load (including the cable) is say 1000 kg, the tension in the cable (mg) will be 9810 N, and the torque applied by the load at the drum will be given by force _ radius ¼ 9810 _ 0:25 _ 2500 Nm. When the speed is constant (i.e. the load is not accelerating), the torque provided by the motor at the drum must be equal and opposite to that exerted at the drum by the load. (The word 'opposite' in the last sentence is often omitted, it being understood that steady-state motor and load torque must necessarily act in opposition.)

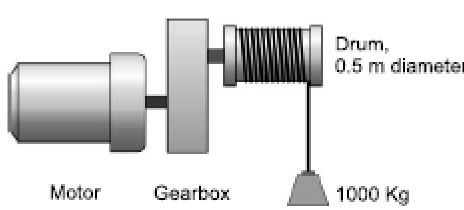


Figure 6.5: Motor-driven hoist – a constant-torque load

Suppose that the hoisting speed is to be controllable at any value up to a maximum of 0.5 m/s, and that we want this to correspond with a maximum motor speed of around 1500 rev/min, which is a reasonable speed for a wide range of motors. A hoisting speed of 0.5 m/s corresponds to a drum speed of 19 rev/min, so a suitable gear ratio would be say 80:1, giving a maximum motor speed of 1520 rev/min.

The load torque, as seen at the motor side of the gearbox, will be reduced by a factor of 80, from 2500 to 31 Nm at the motor. We must also allow for friction in the gearbox, equivalent to perhaps 20% of the full-load torque, so the maximum motor torque required for hoisting will be 37 Nm, and this torque must be available at all speeds up to the maximum of 1520 rev/min.

We can now draw the steady-state torque–speed curve of the load as seen by the motor, as shown in Figure 6.4. The steady-state motor power is obtained from the product of torque (Nm) and angular velocity (rad/s). The maximum continuous motor power for hoisting is therefore given by

$$P_{max} = 37\ x\ 1520\ x\ \frac{2\pi}{60} = 5.9kW$$

At this stage it is always a good idea to check that we would obtain roughly the same answer for the power by considering the work done per second at the load. The force (*F*) on the load is 9810 N, the velocity (v) is 0.5 m/s so the power (*Fv*) is 4.9 kW. This is 20% less than we obtained above, because here we have ignored the power lost in the gearbox. So far we have established that we need a motor capable of continuously delivering 5.9 kW at 1520 rev/min in order to lift the heaviest load at the maximum required speed. However, we have not yet addressed the question of how the load is accelerated from rest and brought up to the maximum speed. During the acceleration phase the motor must produce a torque greater than the load torque, or else the load will descend as soon as the brake is lifted. The greater the difference between the motor torque and the load torque, the higher the acceleration. Suppose we want the heaviest load to reach full speed from rest in say 1 s, and suppose we decide that the acceleration is to be constant. We can calculate the required accelerating torque from the equation of motion, i.e.

$$Torque\ (Nm) = Inertia\ (kgm^2) x\ Angular\ acceleration\ (\frac{rad}{s^2})$$

We usually find it best to work in terms of the variables as seen by the motor, and therefore we first need to find the effective total inertia as seen at the motor shaft, and then calculate the motor acceleration, and finally use equation to obtain the accelerating torque. The effective inertia consists of the inertia of the motor itself, the referred inertia of the drum and gearbox, and the referred inertia of the load on the hook. The term 'referred inertia' means the apparent inertia, viewed from the motor side of the gearbox. If the gearbox has a ratio of n:1 (where n is greater than 1), an inertia of J on the low-speed side appears to be an inertia of $J=n^2$ at the high-speed side.

Inertia matching

There are some applications where the inertia dominates the torque requirement, and the question of selecting the right gearbox ratio has to be addressed. In this context the term 'inertia matching' often causes confusion, so it worth explaining what it means. Suppose we have a motor with a given torque, and we want to drive an inertial load via a gearbox. As discussed previously, the gear ratio determines the effective inertia as 'seen' by the motor: a high step-down ratio (i.e. load speed much less than motor speed) leads to a very low referred inertia, and vice-versa. If the specification calls for the acceleration of the load to be maximized, it turns out that the optimum gear ratio is that which causes the referred inertia of the load to be equal to the inertia of the motor.

Applications in which load acceleration is important include all types of positioning drives, e.g. in machine tools and phototypesetting. (There is another electrical parallel here – to get the most power into a load from a source with internal resistance R, the load resistance must be made equal to R.) It is important to note, however, that inertia matching only maximizes the acceleration of the load. Frequently it turns out that some other aspect of the specification (e.g. the maximum required load speed) cannot be met if the gearing is chosen to satisfy the inertia matching criterion, and it then becomes necessary to accept reduced acceleration of the load in favor of higher speed.

Fan and pump loads

Fans and pumps have steady-state torque–speed characteristics which generally have the shapes shown in Figure 6.6. These characteristics are often approximately represented by assuming that the torque required is proportional to the square or the cube of the speed, giving rise to the terms 'square-law' or 'cube-law' load. We should note, however, that the approximation is seldom valid at low speeds because most real fans or pumps have a significant static friction or breakaway torque (as shown in Figure 6.6), which must be overcome when starting. When we consider the power–speed relationships the striking difference between the constant-torque and fan-type load is underlined. If the motor is rated for continuous operation at the full speed, it will be very lightly loaded (typically around 20%) at half speed, whereas with the constant torque load the power rating will be 50% at half speed

Fan-type loads which require speed control can therefore be handled by drives which can only allow reduced power at such low speeds, such as the inverter-fed cage induction motor without additional cooling, or the voltage-controlled cage motor. If we assume that the rate of acceleration required is modest, the motor will require a torque–speed characteristic, which is just a little greater than the load torque at all speeds. This defines the operating region in the torque–speed plane, from which the drive can be selected. Many fans do not require speed control of course, and are well served by mains-frequency induction motors. The run-up behavior being contrasted with that of a constant-torque load is shown in Figure 6.6.

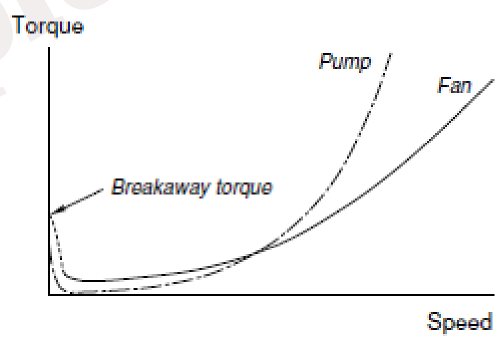


Figure 6.6: Torque-speed characteristics for fan- and pump-type loads

GENERAL APPLICATION CONSIDERATIONS

Regenerative operation and braking

All motors are inherently capable of regenerative operation, but in drives the basic power converter as used for the 'bottom of the range' version will not normally be capable of continuous regenerative operation. The cost of providing for fully regenerative operation is usually considerable and users should always ask the question 'do I really need it?'

In most cases it is not the recovery of energy for its own sake, which is of prime concern, but rather the need to achieve a specified dynamic performance. Where rapid reversal is called for, for example, kinetic energy

has to be removed quickly, and, as discussed in the previous section, this implies that the energy is either returned to the supply (regenerative operation) or dissipated (usually in a braking resistor).
An important point to bear in mind is that a non-regenerative drive will have an asymmetrical transient speed response, so that when a higher speed is demanded, the extra kinetic energy can be provided quickly, but if a lower speed is demanded, the drive can do no better than reduce the torque to zero and allow the speed to coast down.

Duty cycle and rating

This is a complex matter, which in essence reflects the fact that whereas all motors are governed by a thermal (temperature rise) limitation, there are different patterns of operation which can lead to the same ultimate temperature rise.

Broadly speaking the procedure is to choose the motor on the basis of the r.m.s. of the power cycle, on the assumption that the losses (and therefore the temperature rise) vary with the square of the load. This is a reasonable approximation for most motors, especially if the variation in power is due to variations in load torque at an essentially constant speed, as is often the case, and the thermal time-constant of the motor is long compared with the period of the loading cycle. (The thermal time-constant has the same significance as it does in relation to any first order linear system, e.g. an R/C circuit. If the motor is started from ambient temperature and run at a constant load, it takes typically four or five time-constants to reach its steady operating temperature.) Thermal time-constants vary from more than an hour for the largest motors (e.g. in a steel mill) through tens of minutes for medium power machines down to minutes for fractional horsepower motors and seconds for small stepping motors.

Enclosures and cooling

There is clearly a world of difference between the harsh environment faced by a winch motor on the deck of an ocean-going ship, and the comparative comfort enjoyed by a motor driving the drum of an office photocopier. The former must be protected against the ingress of rain and seawater, while the latter can rely on a dry and largely dust-free atmosphere.

Classifying the extremely diverse range of environments poses a potential problem, but fortunately this is one area where international standards have been agreed and are widely used. The International Electrotechnical Committee (IEC) standards for motor enclosures are now almost universal and take the form of a classification number prefixed by the letters IP, and followed by two digits. The first digit indicates the protection level against ingress of solid particles ranging from 1 (solid bodies greater than 50-mm diameter) to 5 (dust), while the second relates to the level of protection against ingress of water ranging from 1 (dripping water) through 5 (jets of water) to 8 (submersible). A zero in either the first or second digit indicates no protection.

Methods of motor cooling have also been classified and the more common arrangements are indicated by the letters IC followed by two digits, the first of which indicates the cooling arrangement (e.g. 4 indicates cooling through the surface of the frame of the motor) while the second shows how the cooling circuit power is provided (e.g. 1 indicates motor-driven fan).

Dimensional standards

Standardization is improving in this area, though it remains far from universal. Such matters as shaft diameter, center height, mounting arrangements, terminal box position and overall dimensions are fairly closely defined for the mainstream motors (induction, d.c.) over a wide size range, but standardization is relatively poor at the low-power end because so many motors are tailor-made for specific applications.

Supply interaction and harmonics

Most converter-fed drives cause distortion of the mains voltage which can upset other sensitive equipment, particularly in the immediate vicinity of the installation. There are some drives that are equipped with 'front-end' conditioning (whereby the current drawn from the mains is forced to approximate closely to a sinewave at unity power-factor), but this increases the cost of the power-electronics and is limited to small and medium-power drives. With more and larger drives being installed the problem of mains distortion is increasing, and supply authorities therefore react by imposing increasingly stringent statutory limits governing what is allowable.

6.2 Encoder Selection

A number of static and dynamic factors must be considered in selecting a suitable sensor to measure then desired physical parameter. Following is a list of typical factors:

- *Range* —Difference between the maximum and minimum value of the sensed parameter
- *Resolution* —The smallest change the sensor can differentiate
- *Accuracy* —Difference between the measured value and the true value
- *Precision* —Ability to reproduce repeatedly with a given accuracy
- *Sensitivity* —Ratio of change in output to a unit change of the input
- *Zero offset* —A nonzero value output for no input
- *Linearity* —Percentage of deviation from the best-fit linear calibration curve
- *Zero Drift* —The departure of output from zero value over a period of time for no input
- *Response time* —The time lag between the input and output
- *Bandwidth*—Frequency at which the output magnitude drops by 3 dB
- *Resonance* —The frequency at which the output magnitude peak occurs
- *Operating temperature* —The range in which the sensor performs as specified
- *Deadband* —The range of input for which there is no output
- *Signal-to-noise ratio* —Ratio between the magnitudes of the signal and the noise at the output

Choosing a sensor that satisfies all the above to the desired specification is difficult, at best. For example, finding a position sensor with micrometer resolution over a range of a meter eliminates most of the sensors. Many times the lack of a cost-effective sensor necessitates redesigning the mechatronic system. It is, therefore, advisable to take a system level approach when selecting a sensor and avoid choosing it in isolation. Once the above-referred functional factors are satisfied, a short list of sensors can be generated. The final selection will then depend upon the size, extent of signal conditioning, reliability, robustness, maintainability, and cost.

FACTORS AFFECTING THE SELECTION OF POSITION SENSORS

In selecting a position sensor, several key factors should be considered:

- *Cost.* Both initial purchase price and life-cycle cost must be considered.
- *Sensing distance.* Photoelectric sensors are often the best selection when sensing distances are longer than 25 mm. Photoelectric sensors can have sensing ranges as long as 300,000 mm for outdoor or extremely dirty applications, down to 25 mm for extremely small parts or for ignoring background. Inductive proximity sensors and limit switches, on the other hand, have short sensing distances. The inductive proximity sensors are limited by the distance of the electromagnetic field—less than 25 mm for most models—and limit switches can sense only as far as the lever operator reaches.
- *Type of material.* Inductive proximity sensors can sense only ferrous and nonferrous materials, whereas photoelectric and limit switches can detect the presence of any solid material. Photoelectric sensors, however, may require a polarizer if the target's surface is shiny.
- *Speed.* Electronic devices using DC power are the fastest—as fast as 2000 cycles per second for inductive proximity models. The fastest-acting limit switches can sense and reset in 4 m/s or about 300 times per second.
- *Environment.* Proximity sensors can best handle dirty, gritty environments, but they can be fooled by metal chips and other metallic debris. Photoelectric sensors will also be fooled or left inoperable if they are fogged or blinded by debris.
- *Types of voltages, connections, and requirements of the device are housing.* All three types can accommodate varying requirements, but the proper selection must be made in light of the power supplies, wiring schemes, and environments.
- *Third-party certification.* The Underwriters Laboratories (UL), National Electrical Manufacturers Association (NEMA), International Electrotechnical Commission (IEC), Factory Mutual, Canadian Standards Association (CSA), and other organizations impose requirements for safety, often based on the type of application. The certification will ensure the device has been tested and approved for certain uses.
- *Intangibles.* These can include the availability of application support and service, the supplier's reputation, local availability, and quality testing statements from the manufacturer.

6.3 Develop the Control Structure: Programmable Logic Controller used for

As the global marketplace demands higher quality goods and lower costs, factory floor automation has been changing from separate machines with simple hardware-based controls, if any, to an integrated manufacturing enterprise with linked and sophisticated control and data systems. For many organizations the transformation has been gradual, starting with the introduction of programmable logic controllers and personal computers to machines and processes. However, for others the change has been rapid and is still accelerating. This chapter discusses the current state of control and data systems that make up manufacturing automation.

The appropriate level of control and automation depends on the process to be automated. Before this can be accomplished, questions about the physical process and product requirements must be answered.
1. What types of process and product feedback are required to control the process (e.g., line speed, force, pressure, temperature, length, thickness, moisture, color)?
2. How is the process run (continuous, batch, sequential operations)?
3. What is the current level of automation (none, relay logic, programmable controller, etc.)?
4. What is the process operation schedule (single shift or 24-h operation)?
5. What cost opportunities are available from reduction of waste, improvement of quality, reduction of downtime?

The last question, which is financial, is typically the most important. When applied correctly process control and automation will rapidly pay for itself.

Controllers

There are many different distinctions in the area of industrial automation controllers. The most widely used controllers are motion controllers, programmable logic controllers (PLC), distributed control systems (DCS), and PC-based control. Each controller type has special features that make it the controller of choice for different automation projects. Many of the distinctions of the various controller types are beginning to be less noticeable with each successive product generation, as options expand and pitfalls are addressed.

PLC: Programmable Logic Controller

The programmable logic controller (PLC) has been part of manufacturing automation forever two decades, replacing the hard-wired relay logic controllers (Figure 6.7). For smaller-scale, event-driven processes and machines with limited I/O points, stand-alone PLCs are the controller of choice. PLCs are rugged, relatively fast, and low cost with excellent sequential control performance.
In the last 10 years the functionality of the PLC and systems using PLCs has been growing rapidly, with integration of networking, peripherals, and expanded programming options. The distinction between PLC systems and other more complex controllers (e.g., DCS) is diminishing, as PLCs are moving up in function and connectivity. Advanced PLCs and DCSs overlap each other's controller areas. Indeed, the networked PLC-SCADA systems are virtually equivalent to the larger DCS systems. Programming standards and networks have also removed the limitations of PLCs. In addition to the traditional ladder logic, four other standardized programming languages are available (discussed in software).

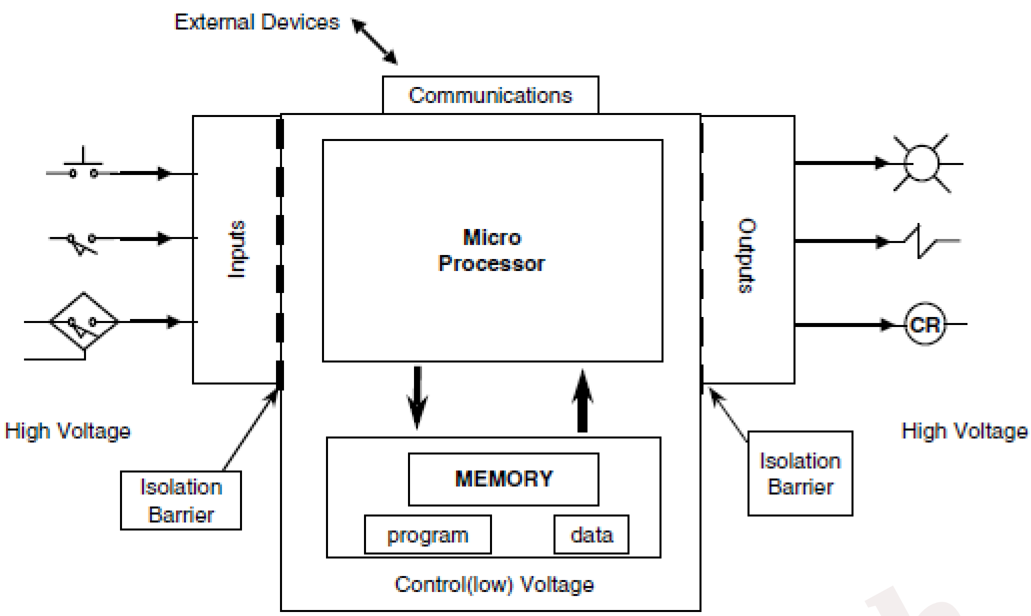


Figure 6.7: PLC-I/O structure

Programming

Software is the center of great debate and change. Ladder logic, rung ladder logic, or ladder logic diagrams (LLD) have long been the main mode of PLC programming as the diagrams graphically resemble the relay logic they originally replaced. Other styles of programming have evolved to augment computational abilities and ease program development. In addition to LLD, four other major programming paradigms are in use in PLCS and DCS systems. These are structured text (ST), function block diagrams (FBD), sequential flowcharts (SFC), and instruction lists (IL). Special purpose motion controller switch augmented processing logic use proprietary text-based languages. While powerful, these unique control languages are not transportable to other controllers and make replacement difficult.

As particular controller vendors often customize their programming functions, it has become more important to have a programming standard for the five prevalent process control languages. One major reason is so one does not need to learn different instruction sets for different PLC manufacturers. The International Standard IEC 61131 is a complete collection of standards on programmable controllers and their associated peripherals.

i. Ladder Logic Diagrams

The ladder logic diagram (LLD) is the most common programming language used in PLC applications. LLD is a graphical language resembling wiring diagrams, as seen in Figure 6.8. Each instruction set is a rung on the ladder. Each rung is executed in sequential order for every control cycle. A rung is a logic statement, reading from left to right (Figure 6.8). Rungs can have more than one branch, as seen in the 4th rung of Figure 6.8.

As ladder logic was the first replacement for relay logic, it is most easily suited to Boolean I/O variables. However, over many years of use and refinement, the current LLD instruction set is quite large, reducing the need for other languages. Many special instructions are available for motion control, array handling, diagnostics, serial port I/O, and character variable manipulation. LLD is used in continuous and batch processes. Both off-line and on-line editing of individual rungs is possible, allowing changes to a running system (on the next cycle).User interfaces allow off-line programming and reviewing of LLD code at either PLC display panels or HMIs, or at remotely via PC software.

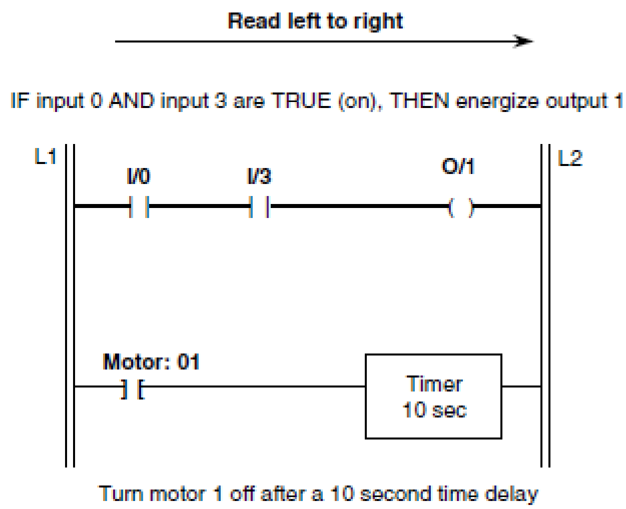


Figure 6.8: Reading ladder logic diagrams

ii. Structured Text

Structured text (ST) is a high-level programming language similar to BASIC, Fortran, Pascal, or C. Most engineers are familiar with structured programming languages. Structured text has standard program flow control command statements such as If/Then, Case, Do/While, Do/Until, and For/Next constructs. An example of a structured text is given in Figure 6.9.Most LLD, SFC, and FBD instructions are supported in ST. While ST provides great flexibility in programming and editing, it is somewhat more difficult to read and follow logic flow, as compared with the more graphical languages. Thus, it is less maintainable over time.


```
;************* force control loop ****************
WHILE (M362>Q5 AND M362<Q4)            ;begin loop
          P6 = (M156*Q2)                ;DESIRED FORCE (N * force bits/N)
          P15 = M1001                   ;GET WEIGHT
          P8 = P15−P14                  ;TARED FORCE
          P7 = P8−P6                    ;FORCE ERROR
          P9 = P9+P7                    ;INTEGRAL ERROR

          IF (P9>Q8)
                    P9 = Q8             ;LIMIT INTEGRATION
          ENDIF
          IF (P9 < (−Q8))
                    P9 = −Q8
          ENDIF
                                        ; UPDATE GAINS
          Q6 = M153*P20                 ;40 nom PROPORTIONAL GAIN
          Q7 = M154*P20                 ;80 nom INTEGRAL GAIN
          X0Y0Z ((P7*Q6+P9*Q7))         ;COMMAND MOVE

ENDWHILE                                ;end of loop
```

Figure 6.9: Structure text example

iii. Function Block Diagram

Function block diagram (FBD) is a highly visual language and is easy to understand because it resembles circuit diagrams. The graphical free-form programming environment is easy for manipulating process variables and control. A typical function block diagram is given in Figure 6.9. The foundation of the FBD program is a set of instruction blocks with predefined structures of inputs and outputs from each block. Different blocks are placed and connections are drawn to pass parameters or variables between blocks. Blocks are positioned and organized, based on the specific application, to improve readability.

Although simple instructions can take as much program space as complex instructions, the architecture simplifies program creation and modification. It is ideal for analog control algorithms, as it graphically represents control loops and other signal conditioning devices. This language is commonly found in distributed control systems (DCS) with continuous or batch process control.

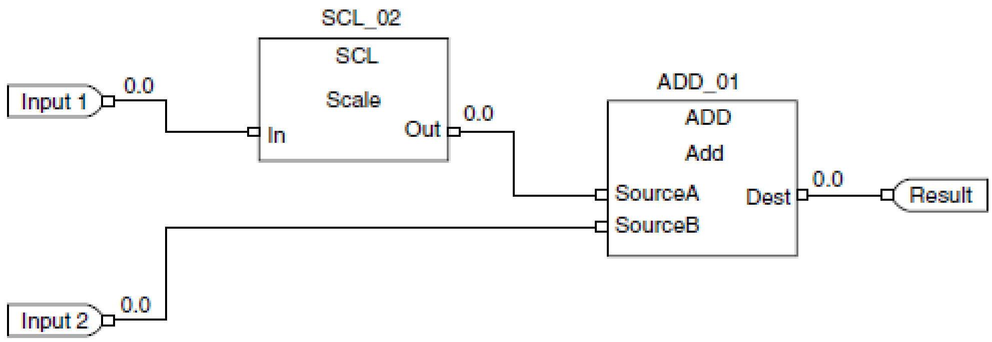


Figure 6.9: Function block diagram example

iv. Sequential Flow Chart

Sequential flow chart (SFC) is another highly visual programming language yielding applications that are easy to create and read. SFC is a graphical flowchart-based programming environment. Logical steps (blocks) are placed on the visual layout in an organized manner. Connections are drawn to determine execution flow of the program. An example is provided in Figure 6.10. This figure shows two branches from the start block of the program. Note: Blocks are descriptively labeled for case in following program logic. I/O and other functions are embedded within each block. Multiple branches allow for transitional and simultaneous execution flow. Position and organization of blocks are used to increase readability.

Floating or linked text boxes provide application documentation (comments). Embedded structured text in action blocks directly improves readability and maintenance, while reducing the number of subroutine calls. The flow chart structure is ideal for sequencing of machine states (e.g., Idle, Run, Normal, Stop). High level program/subroutine flow management provides a more flexible approach to developing process sequences. SFC is ideal for machines with repetitive operations and batch processing.

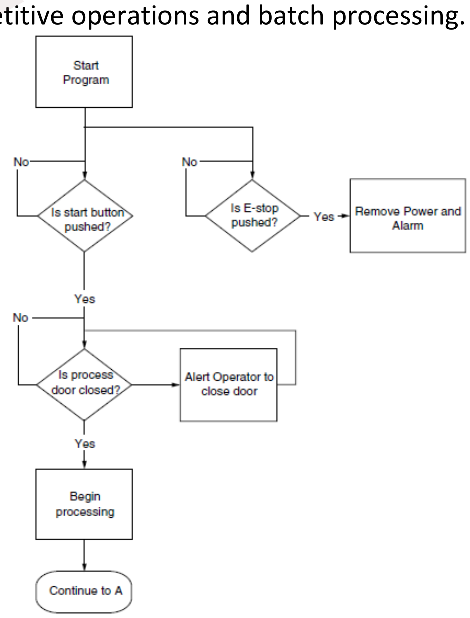


Figure 6.10: Sequence flowchart example

v. IL: Instruction List

Instruction list (IL) is the most basic-level programming language for PLCs. All languages can be converted to IL, although it is most often used with LLD. IL resembles assembly language, using logical operator codes, operands, and an instruction stack. This language is difficult to follow and is not typically selected. The major application of IL is found in hand-held program for non-networked PLCs.

EXAMPLE 6.1:

Consider the fluid storage lank illustrated in Figure 6.11. When the start button X1 is depressed, this energizes the control relay C1.ln tum, this energizes solenoid S1, which opens a valve allowing fluid to flow into the tank. When the tank becomes full, the float switch FS closes, which opens relay C1, causing the solenoid S1 to be de-energized, thus turning off the in-flow. Switch FS also activates timer T1.which provides a 120-sec delay for a certain chemical reaction to occur in the tank. At the end of the delay time, the timer energizes a second relay C2, which controls two devices: (1) It energizes solenoid S2, which opens a valve to allow the fluid to flow out of the tank; and (2) it initiates timer T2, which waits 90 sec to allow the contents of the tank to be drained. At the end of the 90 sec, the timer breaks the current and de-energizes solenoid S2, thus dosing the outflow valve. Depressing the start button X1 resets the timers and opens their respective contacts. Construct the ladder logic diagram for the system.

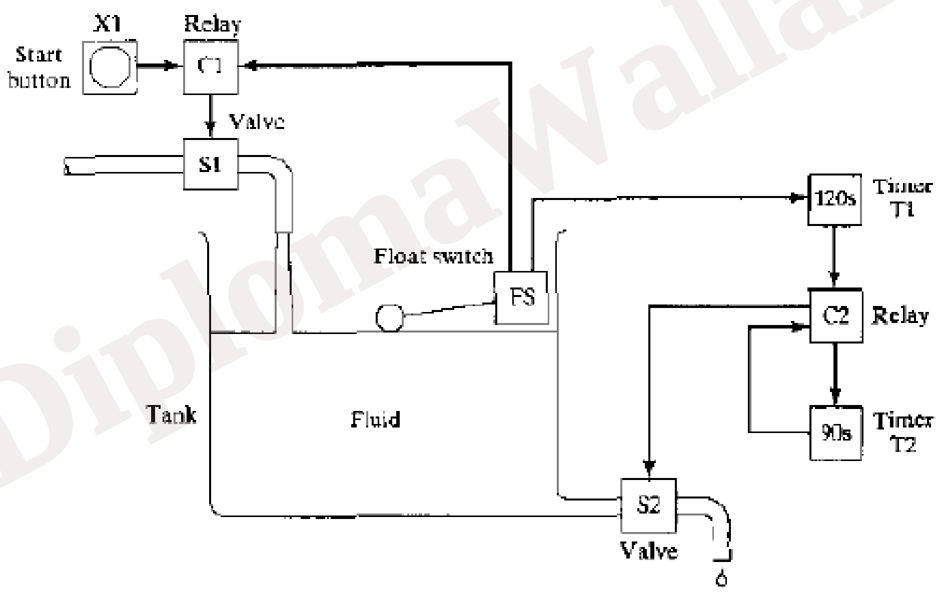


Figure 6.11: Fluid filling operation

Solution:

The ladder logic diagram is an excellent way to represent the combinatorial logic control problems in which the output variables are based directly on the values of the inputs. As indicated by Example 6.1, it can also be used to display sequential control (timer) problems, although the diagram is somewhat more difficult to interpret and analyze for this purpose. The ladder diagram is the principal technique for setting up the control programs in PLCs.

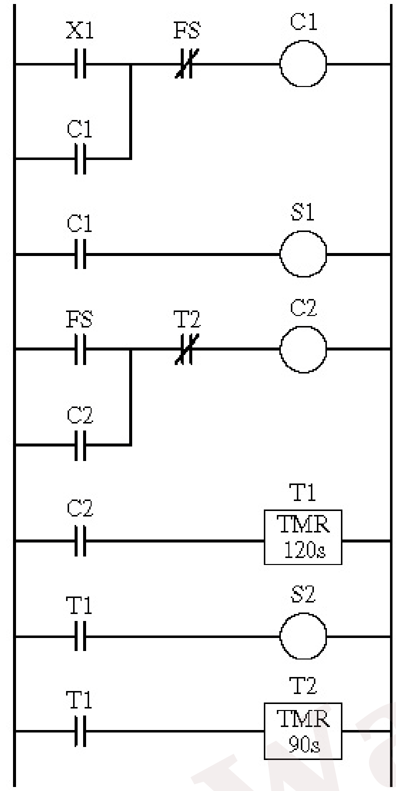


Figure 6.12: A ladder logic diagram for Example 6.1

EXAMPLE 6.2:

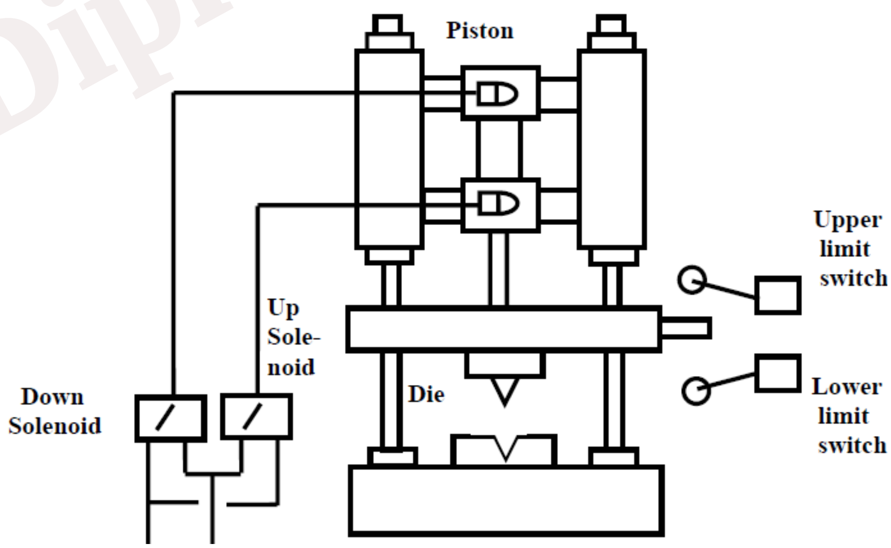


Figure 6.13: An Industrial Logic Control Example

The die stamping process is shown in figure below. This process consists of a metal stamping die fixed to the end of a piston. The piston is extended to stamp a work piece and retracted to allow the work piece to be removed. The process has 2 actuators: an up solenoid and a down solenoid, which respectively control the hydraulics for the extension and retraction of the stamping piston and die. The process also has 2 sensors: an upper limit switch that indicates when the piston is fully retracted and a lower limit switch that indicates when the piston is fully extended. Lastly, the process has a master switch which is used to start the process and to shut it down.

The control computer for the process has 3 inputs (2 from the limit sensors and 1 from the master switch) and controls 2 outputs (1 to each actuator solenoid). The desired control algorithm for the process is simply as follows. When the master switch is turned on the die-stamping piston is to reciprocate between the extended and retracted positions, stamping parts that have been placed in the machine. When the master switch is switched off, the piston is to return to a shutdown configuration with the actuators off and the piston fully retracted.

Linguistic description of the industrial stamping process

The die stamping process is shown in Figure 6.13. This process consists of a metal stamping die fixed to the end of a piston. The piston is extended to stamp a work piece and retracted to allow the work piece to be removed. The process has 2 actuators: an up solenoid and a down solenoid, which respectively control the electro-hydraulic direction control valves for the extension and retraction of the stamping piston and die. The process also has 2 sensors: an upper limit switch that indicates when the piston is fully retracted and a lower limit switch that indicates when the piston is fully extended. Lastly, the process has a master switch which is used to start the process and to shut it down.

The control computer for the process has 3 inputs (2 from the limit sensors and 1 from the master switch) and controls 2 outputs (1 to each actuator solenoid). The desired control algorithm for the process is simply as follows. When the master switch is turned on, the die-stamping piston is to reciprocate between the extended and retracted positions, stamping parts that have been placed in the extended piston machine. When the master switch is switched off, the piston is to return to a shutdown configuration with the actuators off and the piston fully retracted. At first, let us consider an Relay Ladder Logic program that has been written directly from the linguistic description and assess it for suitability of operations.

The first version of sequence control program for the industrial stamping process

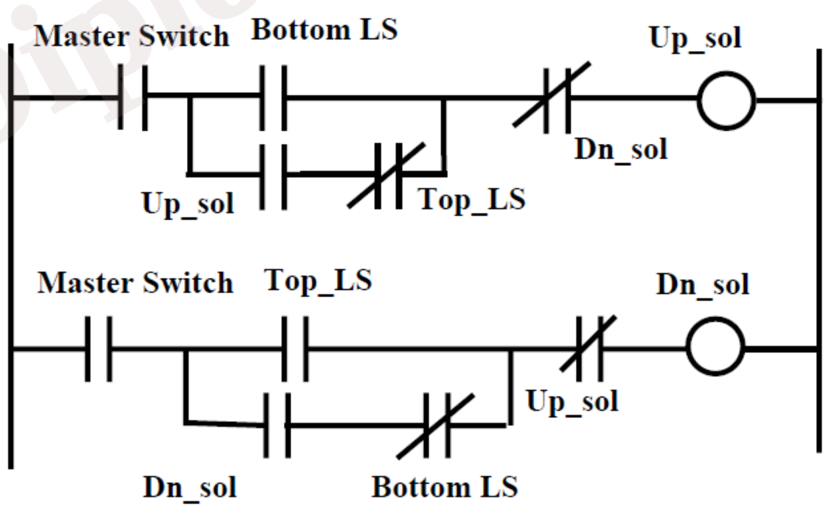


Figure 6.14: An RLL program for the industrial stamping process

A hastily constructed RLL Relay Ladder diagram program for the above process may look like the one given in Figure 6.14. The above program logic indicates that the Up solenoid output becomes activated when the Master Switch is on and the bottom Limit Switch is on. Also there is interlock provided, so that when the Down solenoid is on, the Up solenoid cannot be on. Further, once the Up solenoid is on, the output is latched by an auxiliary contact, so that it remains on till the bottom LS is made on, when it turns off. A similar logic has been implemented for the activation of the Down solenoid.

However, on closer examination, several problems may be discovered with the above program. Some of these are discussed below.

- For example, there is no provision for a Master stop switch to stop the press from stopping in an emergency, except by turning the Master Switch off. This would indeed stop the process, however, if the press stops midway, both bottom and top limit switches would be off. Now the process would not start, even if the Master switch is turned on again. Therefore, either a manual jogging control needs to be provided, so that the operator can return the piston to the up position by manually operating the hydraulics, or special auto mode logic should be designed to perform this.
- As a second example, note that this process does not have a part detect sensor. This implies that the moment the Master switch is on the press would start going up and down at its own travel speed, regardless of whether a part has been placed for pressing or not. Apart from wastage of energy, this could be safety hazard for an operator who has to place the part on the machine between the interval of a cycle of operation.

The above discussion clearly indicates the need for a systematic approach towards the development of RLL programs for industrial logic control problems. This is all the more true since industrial process control is critical application domain where control errors can lead to loss of production or operator safety. Therefore, in this chapter, we discuss a systematic approach towards the design of RLL programs.

Process Control Inputs

- *Part sensor*: A position switch that detects when a part has been placed. In cases where proper positioning of the part can take time. One may also consider using manual switch to be operated by the operator once he is satisfied that the part is properly placed and ready to be stamped. Here an automated part detect sensor has been assumed.
- *Auto PB*: A push button that indicates that machine is ready to stamp parts one after the other in the 'automatic mode'
- *Stop PB*: A push button that the operator can use to stop the machine any time during the time that the piston is moving down. This is needed to avoid stamping a part if any last second error is discovered by the operator regarding, say, the placement of the part.
- *Reset PB*: In case the piston has been stopped due to some error condition, it is desired that the operator explicitly presses this push button to indicate that the error has been taken care of, and the machine is ready to return to stamping in the auto mode.
- *Bottom LS*: This sensor indicates when the piston has reached the bottom position.
- *Top LS*: This sensor indicates when the piston has reached the top position.

Process Control Outputs

- *Up Solenoid*: Control output that drives the Up solenoid of the electro-hydraulic direction control valve which in turn drives the piston up.
- *Down Solenoid*: Control output that drives the Up solenoid of the electro-hydraulic direction control valve which in turn drives the piston up.
- *Auto Mode Indicator*: An indicator lamp that indicates that the machine is in 'Auto' mode.
- *Part Hold*: A gripping actuator that holds the part firmly to avoid movements during stamping

Formal process modeling

Once the requirements have been ascertained, formal process modeling can be undertaken. In this step the informal linguistic descriptions have to be rigorously checked for ambiguity, inconsistency or incompleteness. This is best achieved by converting linguistic descriptions into formal process models. Initially one may use intermediate forms like list of operations, flowchart etc. Eventually and before developing the control programs, these are to be converted into mathematically unambiguous and consistent description using a formal modeling framework such as a Finite State Machine (FSM). It is the experience of practical engineers that modeling paradigms that can be represented pictorially are particularly suited to human beings.

For formal modeling, a process often can be viewed as a Discrete Event System (DES). Much formalism for creating timed or untimed models of DESs exist (e.g. Petri Nets). A detailed description of these is beyond the scope of this lesson. An interested reader is referred to literature on real-time systems for a more detailed discussion on these. In this lesson, it is shown how the process dynamics can be modeled as a Finite State Machine. The following facts which are very important to modeling are mentioned.

A. An FSM is a simple formalism for DES in which, at any time, the system exists in any one discrete-state of a finite set of such states.

B. A state is basically an assignment of values to the set of variables of the system. For a discrete event system, the process variables are assumed to take only a finite set of values. For example, the limit switches can only take two values each, namely, either ON or OFF.

C. Further, the set of the process variables have to be chosen in such a manner that, the future behaviour of the process would be determined solely based on the values of the chosen set of variables at the present time. For example, for the stamping press example, the set of process variables would include the values for the Top and Bottom limit switches. However, based only on these the behaviour of the process cannot be determined. This is because from these it cannot be determined whether the piston is moving up or moving down. Therefore one would have to add the state of the motion as a state variable. The set of values that this variable can take are: 'going up', 'going down' and 'stationary'. In this case, one would also have to add another variable, namely the value of the part detect sensor output to be able to distinguish between the behavioral difference between the case when it is ON and when it is OFF, when the piston is at the top position.

D. Some of the state variables may be measured physically with sensors. Others may not be.

E. The choice of state variables can be subjective and different designers might pick others. The choice also depends on the nature of control actions that one would like to take. Thus, the choice of states is specific to the machine and its operation.

F. During its life cycle, the process moves from state to state over time. Thus it spends most of its time in the states. Occasionally however, it makes a transition from one state to another. The occurrence of a transition depends entirely on the occurrence of discrete events. Such events are names given to conditions involving states, some of which may change due to external factors, such as operator inputs, or due to internal factors, such passage of time. On occurrence of such an event, mechanisms causing state transitions are triggered. State transitions or events are generally considered instantaneous and thus, the system spends time only in the various states. State variables are modified by the occurrence of transitions. In fact, it is this change in the values of the state variables, which is taken to be a transition from one state to another.

G. All possible combinations of state variables may not be valid state assignments for a system. In other words, the system can have only some of all the possible combinations of state variables. These are said to be the combinations that are 'reachable' by the system.

H. One of the states is generally taken to be an 'initial state'. The system, when it starts its life cycle, that is, at the time from which its behaviour is described by the State Diagram, is supposed to be at the initial state.

I. At each state, a set of outputs are exercised. This is described by an output table, where the values for each output variable at each of the states is shown. The output table for the stamping press is shown in Figure 6.16

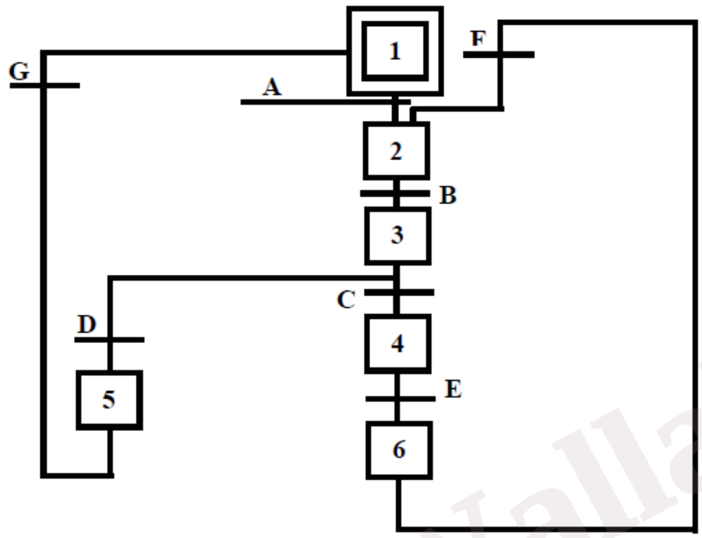


Figure 6.15: The state diagram for the industrial stamping press

State No. / O/P	1	2	3	4	5	6
Auto Indicator	0	1	1	1	0	1
Part Hold	0	0	1	1	0	0
Up Sol	0	0	0	1	0	0
Down Sol	0	0	1	0	0	0

Figure 6.16: The output table for the industrial stamping machine

From the State Diagram in Fig. 6.15, the following may be noted.

A. The states are marked numerically using the numerals 1-6. At each state, one has a unique set of values for the process state variables. For example, in state 3, both the limit switches are OFF, the part detect switch in ON, the motion state variable of the piston is 'going down'. State 1 is marked as the initial state with a double square.

B. The transitions are marked alphabetically using the letters A-G. Each transition has an associated condition under which it occurs. For example, the condition for transition B may be simply stated as "when it in state 2 and part detect goes ON'. Note that the condition described by the phrase 'in state 2' can be further explained in terms of the values of the state variable corresponding to state 2.

C. The outputs exercised at each state are described by the Output Table in Fig. 6.16.

Design of RLL Program

Based on the formal model, the sequence control program can be developed systematically. In fact, one of the main advantages of formal process modeling is that the process of development of the control program becomes mechanical. Thus it can be done quickly and with a much reduced chance of error. In the case of FSM models one has to write the RLL program such that over the scan cycles it executes the state machine itself. The outputs of the state machine go to the process, and since the state machine is nothing but a behavioral model for the process, the process also executes the transitions of the machine, as desired. The realization of a state machine by an RLL program involves computation of the process transitions, in terms of the inputs and the internal state variables of the program followed by computation of the new states and finally, the outputs corresponding to the states. This method is demonstrated here for RLL programming using the above example of the industrial stamping process.

The ladder logic begins with a section to initialize the states and transitions to a single value, corresponding to the initial state. Some PLCs programming languages provide special instructions for such initialization. In this case, however, it is assumed that all auxiliary variables representing the states are set to zero initially. Logic is provided such that in the first scan the auxiliary state variable corresponding to the initial state would be set to 1.

The next section of the ladder logic considers the transitions and then checks for transition conditions. Each transition condition contains an auxiliary NO contact corresponding to the source state from which it is defined. For example, note that the logic for the rung corresponding to transition A contains a contact corresponding to state 1, which is the source state for transition A. Further, it contains the other logical terms corresponding to the input state variables as well as timer outputs, if applicable. In the case of transition A, the external condition is simply the pressing of the Auto PB. Note that, in any scan cycle, at most one transition can be enabled.

The next block of rungs constitutes the state logic. If the transition logic for any transition is satisfied, the following state logic which is the destination state for the enabled transition is to be turned on and the state logic which is the source state for the enabled transition is to be turned off. Therefore each of the rungs corresponding to a state contains one auxiliary NO contact corresponding to the transition for which that state is a destination state. Similarly, each of the rungs corresponding to a state contains one auxiliary NC contact corresponding to the transition for which that state is a source state in series with the above NO contact. If there is more than one transition for which the state is a destination, all the auxiliary NO contacts corresponding to these transitions should be put in parallel. Similarly, if there is more than one transition, for which the present state is a source, all the auxiliary NC contacts corresponding to these transitions are to be put in series. This occurs in the same scan cycle in which the transition logic is turned on, since the state logic rungs follow those corresponding to transition logic. Note that at the end of every scan cycle, there is only one state logic that is enabled.

Now that the state logic has changed, in the next scan cycle the transition that was enabled, turns off and the system stays in that state, till the next transition logic gets enabled. So that the state logic remains turned on, even if a transition for which this state is a source, turns off, an NO auxiliary contact corresponding to the state that latches the state logic is to be provided in parallel with the parallel block of all the auxiliary NO contacts for each transition for which the present state is a destination.

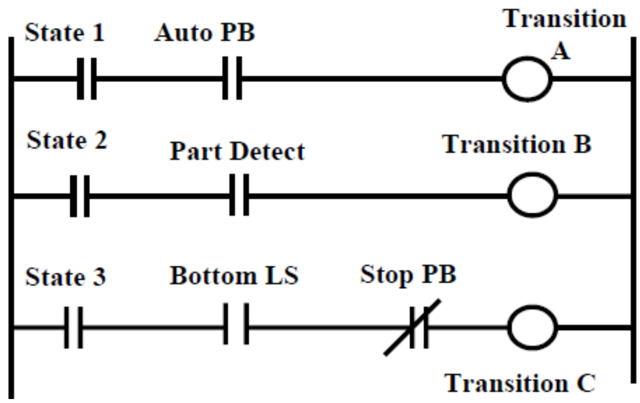


Figure 6.17: State transition logic

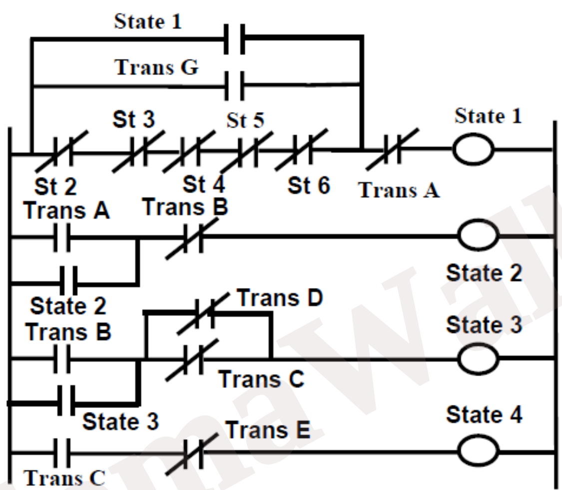


Figure 6.18: State logic

This is followed by ladder logic to turn on outputs as requires by the steps. This section of ladder logic corresponds to the actions for each step. The rung for each output therefore contains one NO auxiliary contact corresponding to the state in which it is enabled. If an output is enabled at more than one state, the auxiliary NO contacts corresponding to those states would be connected in parallel. Similarly, if Manual switch or PB contacts are required, they also have to be put in parallel with the contacts for the states.

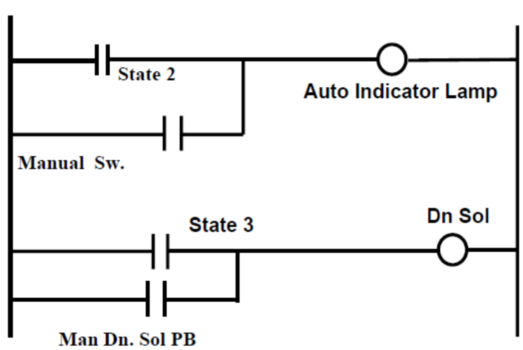


Figure 6.19: Output logic

Some of the rungs corresponding to state, transition and output logics of the industrial stamping process are shown in Fig. 6.170-6.19. These are mostly self-explanatory. The reader is advised to check the correctness of these rungs.

END OF CHAPTER 6

1. What are automated assembly systems? What system configurations can automated assembly systems take?

2. List the hardware components used for parts delivery at workstations.

3. What would generally be seen as typical automated assembly processes?

4. How do the high level sensor and the low level sensor in parts delivery at workstations function?

5. What are the recommendations and principles that can be applied in product design to facilitate automated assembly?

6. Criteria to be consider for motor selection.

7. Factor to be consider for sensor encoder selection.

8. Give an example of where a PLC could be used.

9. A PLC can effectively replace a number of components. Give examples and discuss some good and bad applications of PLCs.

10. What is PLC?

11. List the PLC components.

12. Give five types of PLC language.

13. Develop a simple ladder logic program that will turn on an output X if inputs A and B, or input C is on.

DESIGN PROJECTS

You are required to design an example for industrial automation system. Your design should include

a. Physical systems modeling (drawing, mechanism and so on)
b. Sensors and actuators
c. Logic control system (PLC)

Choose any of those projects.

Single Unit Work Cell Equipment
Adhesive Dispensing Cells
Assembly Assist Stations
Automated Riveters
Automatic Drilling / Tapping
Automatic Packaging
Automatic/Semi-Automatic and Manual Screw Driving Systems
Automatic Staplers
Bar Code Reading ILVS Tracking Systems
Blow Molding Secondary Finishing Machines
Clip Installation Machines
Color Inspection
Conveyers
Crimping Machines
De-Flash Trim Presses & Tooling
De-Gating Equipment
Function Testers
Hole Punch Presses
Hot-Air / Cold-Upset
Hot Melt Foam Dispensing
Hot Plate Welding
Inspection Machines
Modular Systems
Parts Handling and Feeding Systems
Part Marking & ID Systems
Pick & Place
Press Insertion and Punch Machines
Pin Insertion Machines
Plastics Secondary Joining
Pressure Decay Leak Systems
Router Cells
Robotic Work Cells
Rotary Assembly Systems
Special Machines and Fixtures
Tube Benders
Ultrasonic Welding Equipment
Vision Inspection and Vision Guided Systems

Automatic Drilling Machine
Bulk feeds a variety of metal pins used to produce chain links and automatically performs a variety of machining operations and a final inspection. Built-in and self- contained flood coolant and chip control create a true custom machining center.

Automobile Hood Latch
Design a hood latch for an automobile. The latch must be able to hold the hood securely closed during operation of the vehicle. But it should be easy to open for servicing the contents of the engine compartment. Theft-proofing is an important design goal. Attachment of the latch to the frame of the car and to the hood should be defined. Mass production should be a requirement.

Hydraulic Lift
Design a hydraulic lift to be used for car repair. Obtain pertinent dimensions from representative cars for initial height, extended height, design of the pads that contact the car, and so on. The lift will raise the entire car.

Car Jack
Design a floor jack for a car to lift either the entire front end or the entire rear end. The jack may be powered by hand, using mechanical or hydraulic actuation. Or it may be powered by pneumatic pressure or electrical power.

Portable Crane
Design a portable crane to be used in homes, small industries, warehouses, and garages. It should have a capacity of at least 1000 lb (4.45 kN). Typical uses would be to remove an engine from a car, lift machine components, or load trucks.

Can Crusher
Design a machine to crush soft-drink or beer cans. The crusher would be used in homes or restaurants as an aid to recycling efforts. It could be operated either by hand or electrically. It should crush the cans to approximately 20% of their original volume.

Transfer Device
Design an automatic transfer device for a production line. The parts to be handled are steel castings with the following characteristics:

Weight: 187 N
Size: Cylindrical; 6.75-in diameter and 10.0 in high. Exterior surface is free of projections or holes and has a reasonably smooth, as-cast finish.
Transfer rate: Continuous flow, 2.00 s between parts.
Parts enter at a 24.0-in elevation on a roller conveyor. They must be elevated to 48.0 inches in a space of 60.0 in horizontally. They leave on a separate conveyor

Drum Dumper
Design a drum dumper. The machine is to rai.se a 55-gal drum of bulk material from floor level to a height of 60.0 in and dumps the contents of the drum into a hopper.

Paper Feeder
Design a paper feed device for a copier. The paper must be fed at a rate of 120 sheets per min.

Gravel Conveyor

Design a conveyor to elevate gravel into a track. The top edge of the track bed is 8.0 ft (2.44 m) off the ground. The bed is 6.5 ft wide, 12.0 ft long, and 4.0 ft deep (1.98 m X 3.66 m X 1.22 m). It is desired to fill the track in 5.0 min or less.

Construction Lift
Design a construction lift. The lift will raise building materials from ground level to any height up to 40.0 ft (12.2 m). The lift will be at the top of a rigid scaffold that is not a part of the design project. It will raise a load of up to 500 lb (2.22 kN) at the rate of 1.0 ft/s (0.30 m/s). The load will be on a pallet, 3.0 ft by 4.0 ft (0.91 m X 1.22 m). At the top of the lift, means must be provided to bring the load onto a platform that supports the lift.

Packaging Machine
Design a packaging machine. Toothpaste tubes are to be taken from a continuous belt and inserted into cartons. Any standard tube size may be chosen. The device may include the means to close the cartons after the tube is in place.

Carton Packer
Design a machine to insert 24 cartons of toothpaste into a shipping case.

Robot Gripper
Design a gripper for a robot to grasp a spare-tire assembly from a rack and insert it into the trunk of an automobile on an assembly line. Obtain dimensions from a particular car.

Weld Positioner
Design a weld positioner. A heavy frame is made of welded steel plate in the shape. The welding unit will be robot-guided, but it is essential that the weld line be horizontal as the weld proceeds. Design the device to hold the frame securely and move it to present the part to the robot. The plate has a thickness of 3/8 in (9.53 mm).

Garage Door Opener
Design a garage door opener.

Lift Device Using Acme Screws
An electric motor drives the worm at a speed of 1750 rpm. The two Acme screws rotate and lift the yoke, which in tum lifts the hatch. See Example Problem 17-1 for additional details. Complete the entire unit, including the worm gear set, the chain drive, the Acme screws, the bearings, and their mountings. The hatch is 60 in (1524 mm) in diameter at its top surface. The screws should be nominally 30 in (762 mm) long. The total motion of the yoke will be 24 in, to be completed in 15.0 s or less.

Brake for a Drive Shaft
Design a brake. A rotating load (as sketched in Figure 22-21) is to be stopped from 775 rpm in 0.50 s or less. Use any type of brake, and complete the design details, including the actuating means: springs, air pressure, manual lever, and so on.

Brake for a Winch
Design a complete brake for the application shown in Figure 22-23 and described in Problem 11 in Chapter 22.

Indexing Drive

Design an indexing drive for an automatic assembly system. The items to be moved are mounted on a square steel fixture plate, 6.0 in (152 mm) on a side and 0.50 in (12.7 mm) thick. The total weight of each assembly is 10.0 lb (44.5 N). The center of each fixture (intersection of its diagonals) is to move 12.0 in (305 mm) with each index. The index is to be completed in 1.0 s or less, and the fixture must be held stationary at each stafion for a minimum of 2.0 s. Four assembly stations are required. The arrangement may be linear, rotary, or any other, provided that the fixtures move in a horizontal plane.

Child's Ferris Wheel

Design a child's Ferris wheel. It should be capable of holding one to four children weighing up to 80 lb (356 N) each. The rotational speed should be 1 rev in 6.0 s. It should be driven by an electric motor.

Merry-Go-Round

Design an amusement ride for small children, age 6 or younger, that they can sit in safely while enjoying an interesting motion pattern. At lea.st two children at a time can ride the machine. The ride will be marketed to shopping centers and department stores to amuse customers' children.

Transfer Device

Design a device to move automotive camshafts between processing stations. Each movement is to be 9.0 in (229 mm). The camshaft is to be supported on two unfinished bearing surfaces having a diameter of 3.80 in (96.5 mm) and an axial length of 0.75 in (19.0 mm). The spread between the bearing surfaces is 15.75 in (400.0 mm). Each camshaft weighs 16.3 lb (72.5 N). One motion cycle is to be completed each 2.50 s. Design the complete mechanism, including the drive from an electric motor.

Chain Conveyor

Design a powered, straight chain conveyor to move eight pallets along an assembly line.
The pallets are 18 in long and 12 in wide. The maximum weight of each pallet and the product it carries is 125 lb. At the end of the conveyor, an external downward force of 500 lb is applied to the product which must be carried through the pallet to the conveyor structure. You may design the configuration of the sides and bottom of the pallet.